APPENDIX 10

VOCAL FOLD MODELS: THEORY

A10.1 INTRODUCTION

The larynx provides phonation for speech. When there is sufficient airflow through the glottis, sufficient air pressure drop across the glottis, and appropriate configurational and physiologic conditions for the laryngeal tissues, the vocal folds of the larynx vibrate (Ishizaka and Flanagan, 1972; Titze and Talkin, 1979a and 1979b; van den Berg and Tan, 1959). The human phonatory mechanism produces a vocal tract excitation of quasi-periodic pulses of air that are generated by the glottis as it valves the airflow from the trachea (Carhart, 1940).

The glottis is the three-dimensional airspace between the vocal folds. The "naturalness" of synthetic speech is closely related to the shape of the glottal flow, which mainly depends on the geometry of the glottis. Traditionally, the glottis is viewed and described primarily from the superior view because most visual imaging techniques are limited to this aspect. Radiographic imaging can provide an additional third dimension, but the detail is poor because of the rapid dynamic changes of the glottis during phonation. Ultrasonic imaging is also limited in its ability to resolve the details of the time-varying boundary between the tissue and air (Titze, 1989).

The vibratory pattern of the vocal folds is a major factor relating laryngeal function to sound production (Childers and Lee, 1991; Childers and Wu, 1990). Due to the relative inaccessibility of the larynx, many direct examinations of laryngeal function are precluded. One method that can integrate the various subsystems of the phonatory mechanism is a computer model of the vocal folds. Computer simulation models can be constructed with adequate relationships among the system parts to not only predict various relationships, but also some hypotheses can be advanced concerning phonatory mechanisms (Scherer, 1981). An understanding of phonatory mechanisms can benefit speech synthesis and analysis, linguistics, and clinical practice including the detection, diagnosis, and treatment of vocal disorders.

The problem of determining the internal structure and movement from speech or other measured data is central to speech analysis. One aspect of acoustic phonetics, for example, deals with inferring articulatory shapes and movement from the speech waveform (or its spectrum). Usually, the success of this inverse problem depends on the ability to accurately model the underlying biophysical processes. Constructing simple predictive models of phonatory acoustics, tissue mechanics, and glottal aerodynamics is difficult. Unlike the acoustic theory of wave propagation in nonuniform ducts, which for years has been the foundation of studies in resonance and articulation (Fant, 1960), the myoelastic–aerodynamic theory of phonation has only recently become quantitative. It has been recognized that an adequate mathematical description of the interaction between fluid flow and tissue movement in the larynx is considerably more complex than acoustic wave propagation in ducts.

Imaging and modeling the vocal folds is difficult, yet holds promise both to increase our scientific understanding of the speech process and to improve our potential for advancing the clinical procedures for treating disorders of the larynx. The vocal fold models toolbox in Chapter 9 provides two vocal fold vibratory models that yield a three-dimensional vibratory image of the vocal folds. This appendix provides the background and theory for the toolbox. Since a knowledge of the structure of the vocal folds as a mechanical vibrator is important for modeling, we first introduce the basic anatomy of the larynx, followed by the accepted theory of vocal fold vibration, and an overview of some existing vocal fold vibratory models.

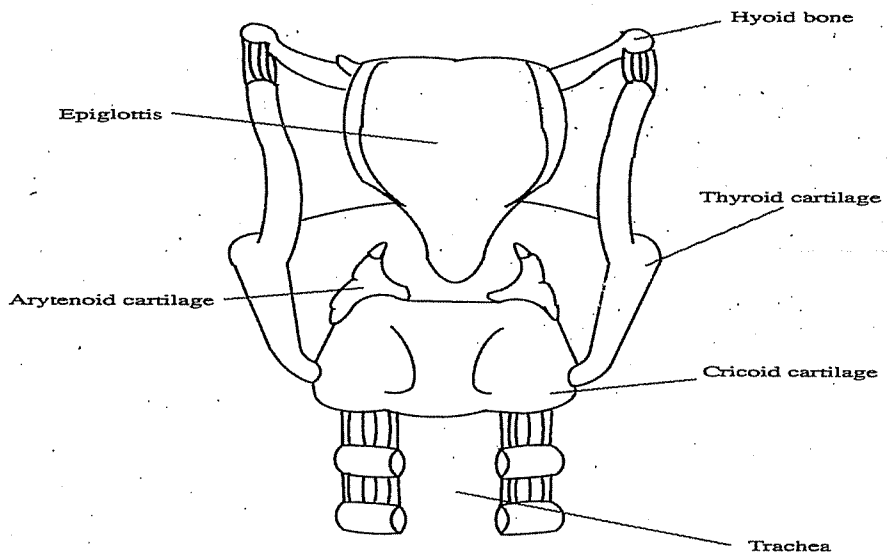


FIGURE A10.1 Posterior view of the larynx.

A10.2 BASIC ANATOMY OF THE LARYNX

A10.2.1 Skeletal Structure of the Larynx and Its Function

The larynx, shown as a sketch in Figure A10.1, is composed of nine cartilages, three paired (arytenoid, corniculate, and cuneiform) and three unpaired (thyroid, cricoid, and epiglottis), and one hyoid bone, as depicted in Figure A10.2 (Schneiderman, 1984; Shearer, 1979).

The hyoid bone can be considered as support for the tongue as well as part of the laryngeal system. The horseshoe-shaped hyoid serves as an attachment for many of the extrinsic laryngeal

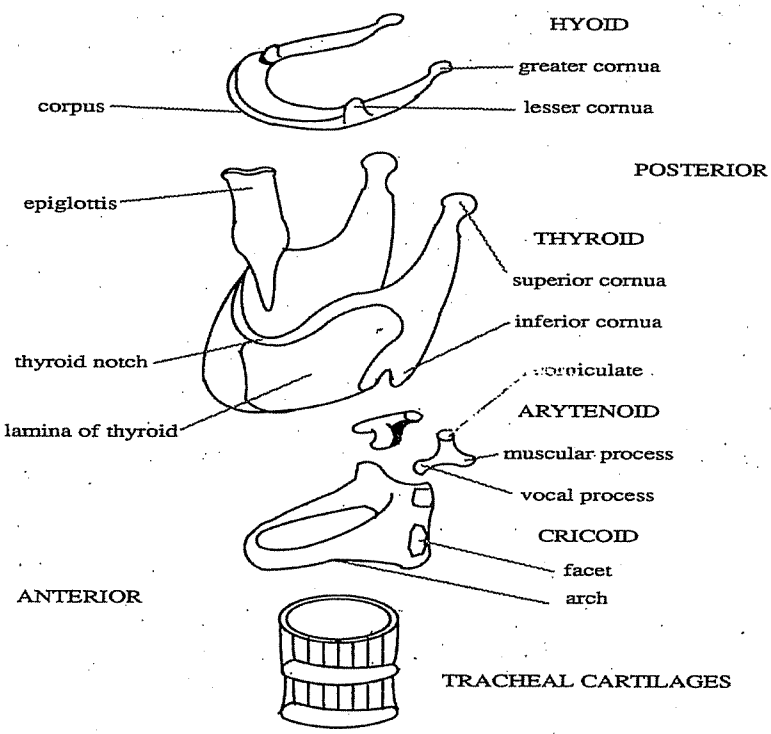


FIGURE A10.2 Structure of the larynx.

muscles. The thyroid is the largest cartilage of the larynx. Its superior cornua attaches indirectly to the corresponding major cornua of the hyoid, and its inferior cornua attaches to the posterior aspect of the cricoid arch. The cricoid cartilage is shaped like a signet ring with the anterior arch, a narrow convex ring, and posteriorly the lamina form the "signet." Two arytenoid cartilages rest on the superior border of the cricoid lamina. Each arytenoid approximates a pyramidal shape with a lateral projection that is called the muscular process and an anterior projection that is called the vocal process. Its superior aspect, or apex, curves slightly backward. The cricoarytenoid joint is a saddle joint that permits a rocking motion and a limited amount of gliding action (Zemlin, 1964 and 1988). The corniculate cartilages (cartilages of Santorini) are two pyramidal shaped nodules located on the apex of each arytenoid for protection of the arytenoid. The cuneiform cartilages (cartilages of Wrisberg) are rod shaped elastic cartilages found in the posterior portions of the aryepiglottic folds that give support to the membrane. The epiglottis is a single leaf-like structure bound by ligaments to the base of the tongue, walls of the pharynx, and thyroid cartilage. This structure acts to close off the laryngeal airway and deflect bulbi of food posteriorly into the esophagus during swallowing.

A10.2.2 Articulations and the Larynx

The cricothyroid and cricoarytenoid cartilages (the articulations are known by the same name) in the larynx are important for understanding the speech process. The cricothyroid articulations are located between the inferior horn of the thyroid and the lateral wall of the cricoid at the point where the arch and laminae meet (see Figures A10.1 and A10.2). Two types of movements are possible for this articulation. One is a gliding movement in a ventrodorsal (anterior to posterior) direction of either cartilage on the other. The second is a tipping (tilting) or rocking motion of either cartilage on the other around an axis that is parallel to a line joining the two eardrums when the head is facing forward (Figure A10.3). The combination of tipping and gliding is the anatomic basis of a stretching or tightening movement of the vocal folds.

Figure A10.3(*a*) shows a gliding movement of the thyroid on the cricoid. The solid line indicates a normal position; the dashed lines show the extremes of the glides. In this case, the thyroid moves on the cricoid as shown by the arrow. Figure A10.3(*b*) shows a tipping (tilting) movement of

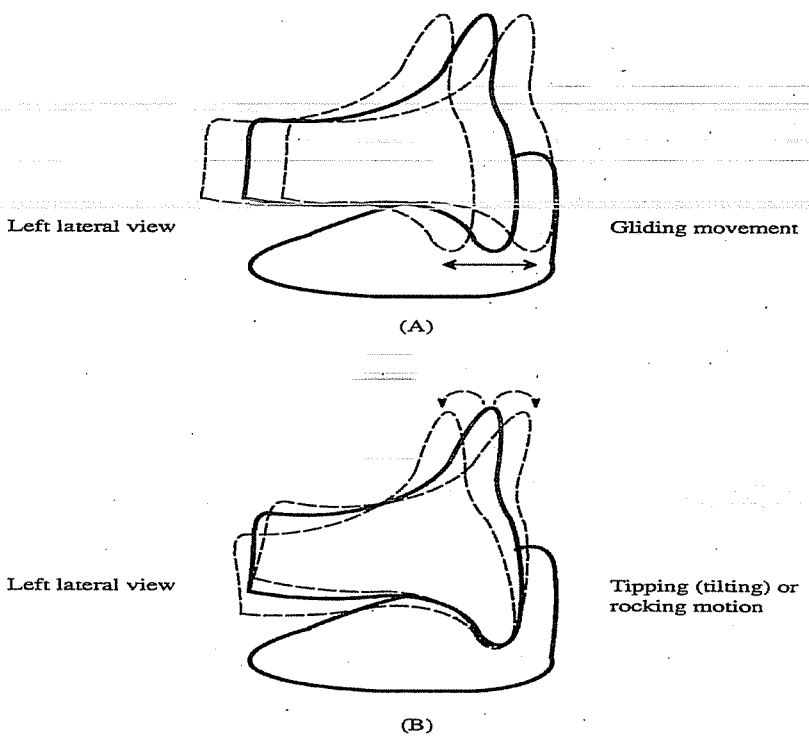


FIGURE A10.3 Movements of the thyroid on the cricoid: cricothyroid articulations.

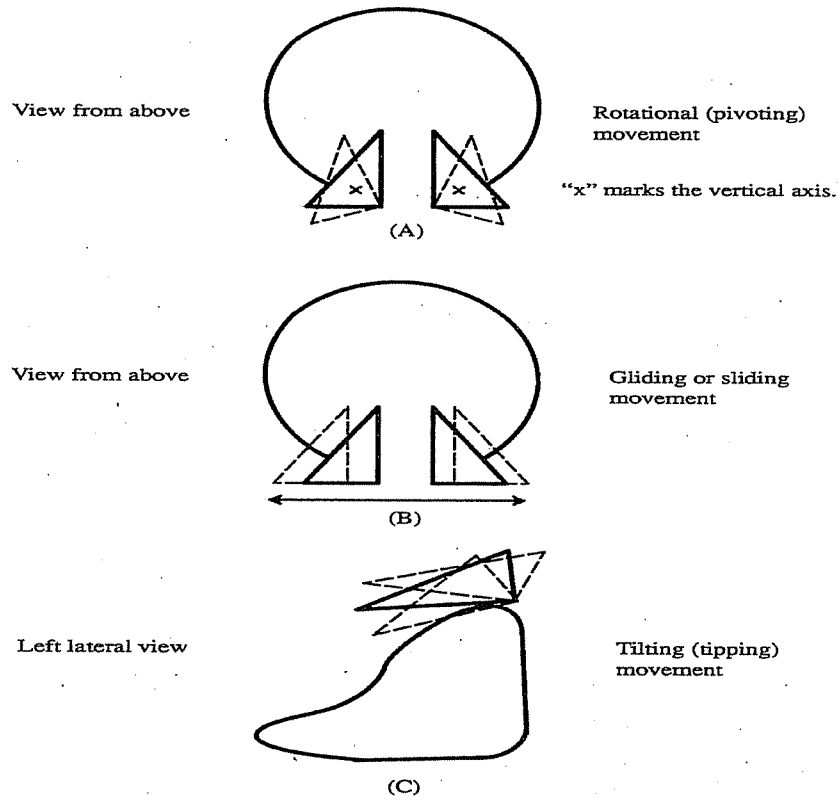


FIGURE A10.4 Misconceived movements of the arytenoid (cricoarytenoid articulations).

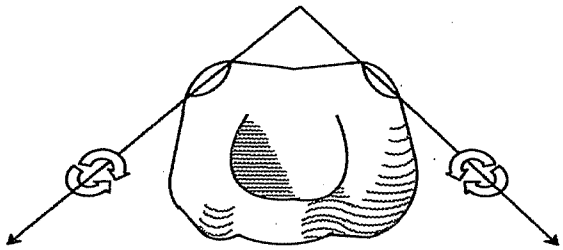
the thyroid on the cricoid. The solid line indicates a normal position; the dashed lines show tipped positions. Note that the inferior horn does not glide on the cricoid.

The cricoarytenoid articulations are very complex. The two arytenoids are perched on the high back part (signet portion) of the cricoid (see Figure A10.1 and Figure A10.2). The mechanics of the cricoarytenoid joint control abduction and adduction of the vocal folds, and thereby facilitate respiration, protect the airway, and permit phonation and other functions of the larynx. The obscure position of this joint and the complex structure have led to at least three misconceptions of arytenoid motion as shown in the Figure A10.4 (von Leden and Moore, 1961). The first misconception is that there is a rotational movement around an axis that parallels the spinal column. The second misconception is that there is a gliding or sliding movement of the arytenoids on the superior border of the lamina of the cricoid. This gliding movement is considered to be back and forth along an axis that is parallel with a line drawn between the eardrums. The third misconception is that there is a tilting (tipping) movement around a horizontal axis that parallels a line drawn between the eardrums (van Riper and Irwin, 1958).

A monocular inspection of the larynx, for instance, leads the observer to the conclusion that the motions of the cricoarytenoid joint may be based on a vertical axis of rotation or on a linear lateral glide (von Leden and Moore, 1961). Based on anatomic, cinematographic, and mathematical studies, von Leden and Moore (1961) revised the three misconceived movements between each arytenoid and the cricoid as shown in Figure A10.5:

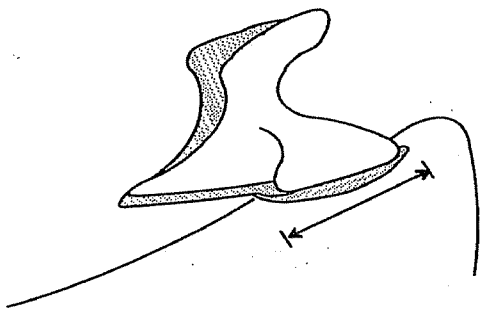
There is a rocking or rotating movement around the axis of the joint (the principal axis of rotation). The principal axis of rotation extends in a dorsomediocranial and ventrolaterocaudal direction, Figure A10.5(*a*). This rocking motion represents an example of internal rotation (similar to the rotation of the earth around its polar axis). The distance of the vocal process from the axis of rotation provides the leverage for the massive movement of the vocal folds during the opening and closing of the glottis. Incidentally, the same rocking movement lowers the vocal process in adduction and slightly shortens the vocal folds.

There is a linear gliding movement parallel to the principal axis of rotation. The direction of linear motion occurs along the longitudinal dimension of the cricoid facet; that is, dorsomediocranially



A rocking or rotating movement
around the principal axis of the joint

(A)



A linear glide movement
parallel to the principal axis of the joint

(B)

FIGURE A10.5 The two principal (correct) types of motion at the cricoidarytenoid joint.

and ventrolaterocaudally, Figure A10.5(*b*). This excursion is limited to approximately 2 mm, the extension of the cricoid facet beyond the longitudinal diameter of the arytenoid facet. In isolation, this motion tends to shorten or lengthen the vocal cords during vocal adjustments, with a small amount of lateral displacement.

A third, although very limited, type of rotating motion is around the secondary axis of rotation, which pivots outside the cricoarytenoid joint, near the attachment of the posterior cricoarytenoid ligaments into the cricoid lamina. This movement represents an illustration of external rotation (comparable to the motion of the earth around the sun). Zemlin (1964 and 1988) feels that the third type of movement is a very restricted and controversial rotary motion. Because of the nature of the joint, this motion is negligible and quite probably does not occur in the normal larynx. However, it is sometimes recognized and has been largely confirmed by means of a mathematical analysis by von Leden and Moore (1961).

Since the vocal folds are attached at the anterior to the thyroid and at the posterior to the arytenoids, and both the thyroid and the two arytenoids are attached to the thyroid with complex articulations, an almost infinite variety of positions of the vocal folds is possible. It is this complexity of movement that makes it so difficult to analyze the possible movements of the vocal folds. Yet it is this same complexity that enables us to make an incredible variety of sounds.

A10.2.3 Laryngeal Membranes

Extrinsic membranes and ligaments connect the laryngeal cartilages, hyoid bone, and trachea. The thyrohyoid membrane runs between the thyroid cartilage and the hyoid bone. The hyoepiglottic membrane runs from the epiglottis to the hyoid, and the cricotrachial ligament connects the cricoid to the first tracheal ring. The intrinsic membrane covers the median surface of the larynx, connecting the laryngeal cartilages together.

A10.2.4 Muscles of the Larynx

The laryngeal musculature is divided into extrinsic and intrinsic groups. Extrinsic muscles are those that support the larynx, and are also called strap muscles. Their function is to move the larynx as a whole. They have at least one attachment to a structure outside the larynx. The intrinsic muscles control phonation. They have both relatively fixed and movable muscle attachments within the larynx, and may be grouped according to the function of opening or closing the glottis.

The intrinsic muscles are described next (Moore, 1971; Schneiderman, 1984; Shearer, 1979). The posterior cricoarytenoid muscle is the abductor of the glottis, which is situated on the posterior surface of the cricoid cartilage (Figure A10.6). It rotates the arytenoid so that the muscular process is drawn backward and the vocal process outward. Thus, the vocal folds are moved laterally and the glottis is widened. The lateral fibers of this muscle help slide the arytenoids laterally.

There are four adductor muscles (oblique arytenoid, transverse arytenoid, lateral cricoarytenoid, and thyroarytenoid) that close the glottis. The oblique arytenoid (paired) originate on the posterior surface of the muscular process of one arytenoid (Figure A10.7). Together the two oblique arytenoids serve as a weak sphincter for the superior aperture of the larynx. The transverse arytenoid (unpaired) covers the entire posterior surfaces of the arytenoids, extending from the base to the summit (Figure A10.6). It approximates the arytenoid cartilages by sliding toward one another. This muscle closes the posterior part of the glottis. Complete glottal closure is often assisted by the lateral cricoarytenoid muscle.

The lateral cricoarytenoid (paired) are located interior to the thyroid cartilage in the lateral wall of larynx (Figure A10.8). They rotate the arytenoid so that the muscular process is drawn forward with the vocal process medialward. The thyroarytenoid (paired) form the substance of the vocal folds and include the thyrovocalis muscle and thyromuscularis fibers (Figure A10.9). They may influence the vibration of the vocal folds by drawing the arytenoids closer to the thyroid. The thyroarytenoids make the folds more flaccid. By varying the tension in the walls to which the vocal folds attach, the thyroarytenoids affect the mode of vibration of the folds. The cricothyroid muscle lies on the external surface of the larynx arising along the lower border and outer surface of the cricoid arch (Figure A10.10). This muscle increases the distance between the angle of the thyroid and the arytenoids, thus increasing the tension of the vocal folds.

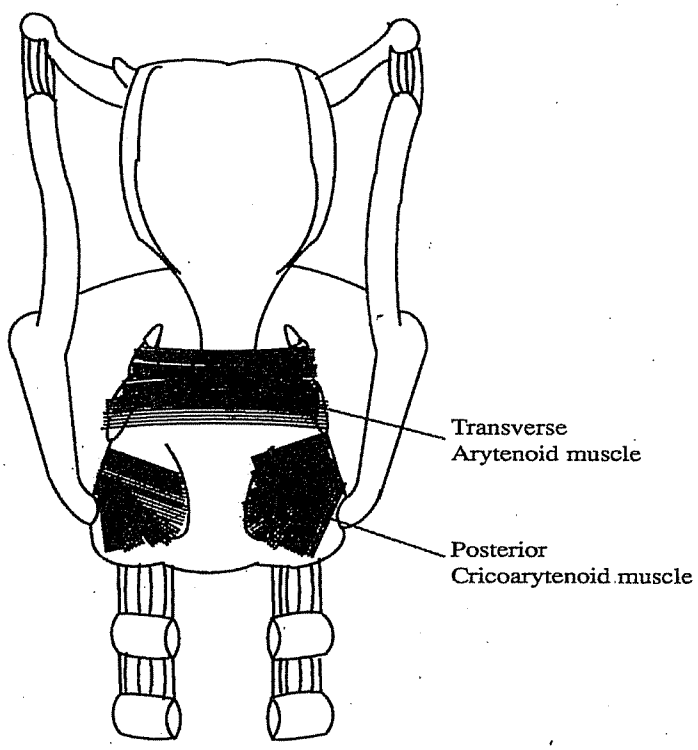


FIGURE A10.6 A posterior view of the intrinsic muscles of the larynx.

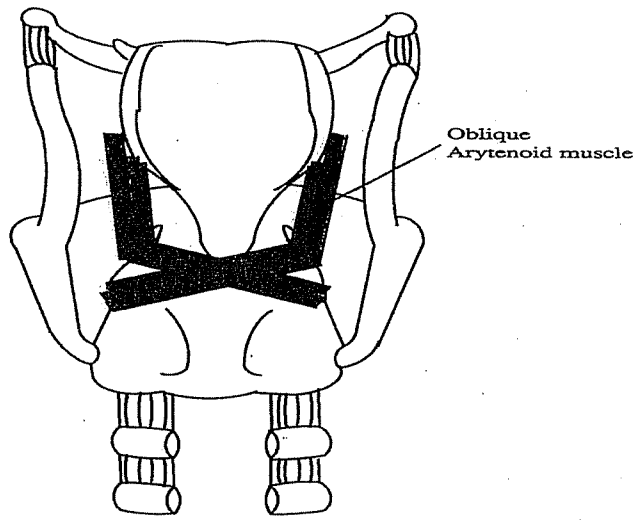


FIGURE A10.7 A posterior view of the intrinsic muscles of the larynx.

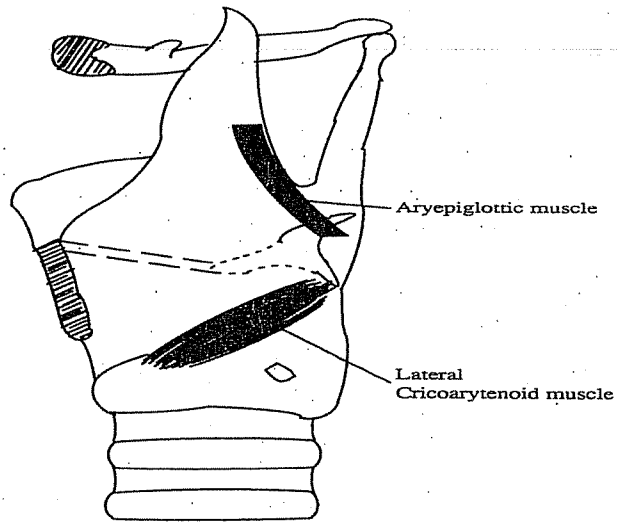


FIGURE A10.8 Intrinsic muscles of the larynx as viewed from the left with the left side of the thyroid cartilage removed.

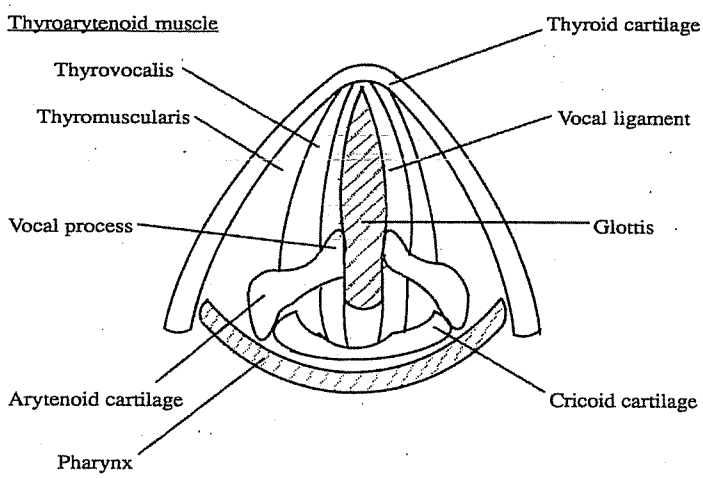


FIGURE A10.9 Intrinsic muscles of the larynx. A superior view of transverse section at the level of the vocal folds.

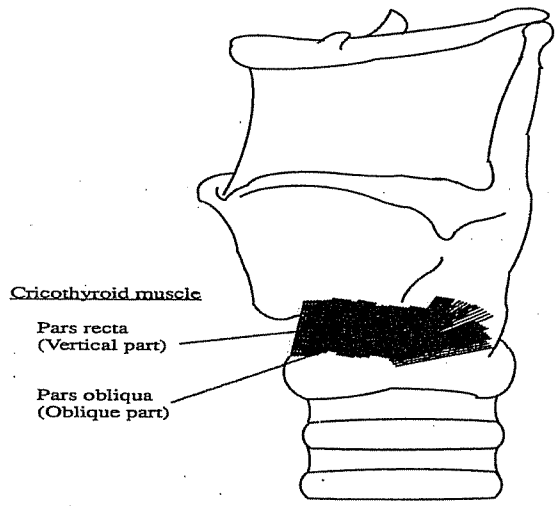


FIGURE A10.10 A left lateral view of the intrinsic muscles of the larynx.

A10.2.5 Laryngeal Cavity

The laryngeal cavity is the upward continuation of the cavity of the trachea from the cricoid cartilage to the superior entrance bounded by the glossoepiglottic folds (Figure A10.11). The laryngeal cavity has three divisions. The superior subdivision or vestibule, extends from the superior aperture of the larynx to the ventricular or false folds. The middle subdivision, or ventricle, extends from the ventricular folds to the true vocal folds. The inferior division extends from the vocal folds to the trachea (Boone, 1971).

The larynx is a valving system. The ventricular folds (paired) run horizontally along the lateral wall of the laryngeal cavity. Each soft and somewhat flaccid fold contains the lower part of the quadralateral membrane with its ventricular ligament and mucous glands, which lubricate the vocal folds. The ventricular folds are more widely separated than the vocal folds, and form a valve that serves to keep air inside the lungs when excess intralung pressure is needed. The ventricle of Morgagni (which is a cave-like cavity having its opening between the ventricular fold and the vocal fold on the same side) aids the valving action of the ventricular folds. Laterally it has an upward extension. When the ventricular folds are approximated and the intrathoracic pressure is increased, the pressure inside the ventricle of Morgagni is increased, and, thus, the ventricular folds are pressed harder against each other, aiding the folds to resist even greater intrathoracic pressure.

The vocal folds or true vocal folds (paired) are parallel to and inferior to the ventricular folds. They extend from the posterior surface of the thyroid angle to the vocal processes of the arytenoids (Figure A10.9). Each band is attached along the lateral wall of the larynx and is composed of a

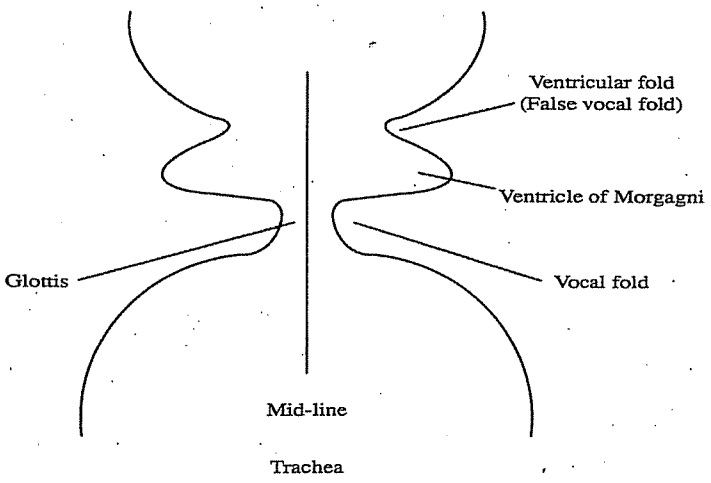


FIGURE A10.11 Schematic representation of the laryngeal cavity. A frontal view.

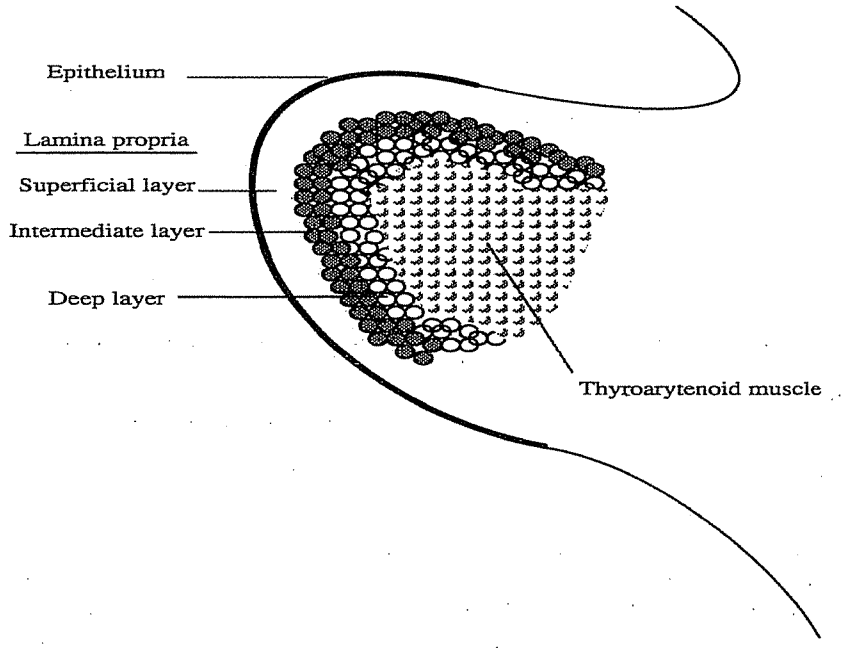


FIGURE A10.12 Schematic representation of a coronal section through the right vocal fold showing tissue layers of the vocal fold.

medial ligament (vocal ligament) and two lateral muscle groups (the thyromuscularis and thyrovocalis muscles). Its mucous membrane is thin and pale in color. Its median edge (vocal ligament) is pearly white. The vocal folds form a valve that prevent the entrance of air or other substances into the trachea and lungs.

The glottis is the opening between the vocal folds. The intermembranous portion of the glottis is the anterior section bounded by the vocal folds. The intercartilagious portion is the posterior section bounded laterally by the medial surfaces of the arytenoid cartilages and posteriorly by the transverse arytenoid muscle (Figure A10.9). The width of the glottis is determined by movements of the arytenoid cartilages. When the arytenoids slide toward each other or rotate so that their vocal processes are approximated, the glottis is narrowed. The opening is widest during inhalation and narrowest during phonation. Tilting the arytenoids, though it changes slightly the length of the glottis, is important, and affects the tension of the vocal folds. The length of the glottis may be altered by the rotating movements allowed by the cricothyroid joint.

A10.2.6 Vocal Fold Tissue Layers

Figure A10.12 shows a drawing of the layered structure of the right vocal fold in coronal section (Mackenzie, Lavar, and Hiller, 1983; Titze, 1994). The outermost layer is thin, 0.05 to 0.10 mm thick, made up of stratified squamous (layered and scalelike) epithelium (Hirano, 1977). The epithelium encapsulates softer, fluidlike tissue, somewhat like a balloon filled with water. The lamina propria, a layered system of nonmuscular tissues, is between the epithelium and muscle, and can conveniently be divided into three layers: superficial, intermediate, and deep. The superficial layer is approximately 0.5 mm thick in the middle of the vocal fold (Hirano, Kurita, and Nakashima, 1981) and consists primarily of loosely organized elastin fibers surrounded by interstitial fluids. Elastin fibers are made of a special type of protein structure that allows for ample elongation (like a rubber band). The intermediate layer is also made up primarily of elastin fibers, but they are more uniformly oriented in the anterior–posterior (longitudinal) direction. There are also some collagen fibers. The intermediate and deep layers of the lamina propria together are about 1 to 2 mm thick (Hirano, Kurita, and Nakashima, 1981). The deep layer is made up primarily of collagen fibers. These fibers have a protein structure that limits elongation. Like a cotton thread, they are nearly inextensible. The fibers in the deep layer also run parallel in the anterior–posterior direction. The thyroarytenoid muscle, approximately 7 to 8 mm thick, lies laterally to the lamina propria. This is the major portion of the vocal fold. In total, the epithelium, the three layers of the lamina propria, and the muscle constitute a five-layer scheme.

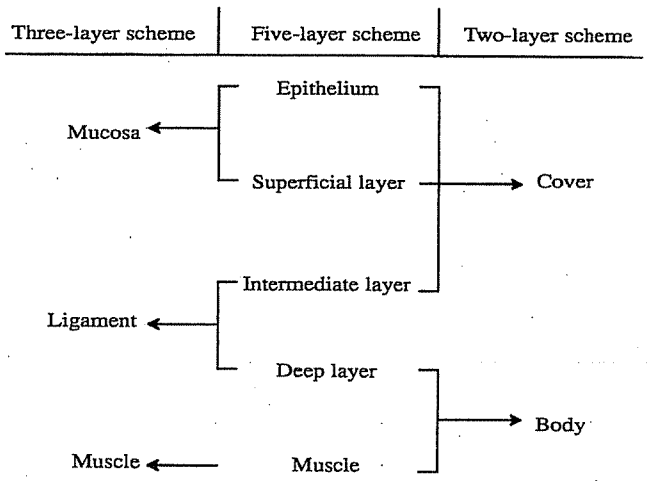


FIGURE A10.13 Three schemes used to label the layered structure of the vocal folds.

Different labeling schemes have been used to group the vocal fold soft-tissue layers, depending on the physiology to be described. In a three-layer scheme, the mucosa consists of the epithelium and the superficial layer of the lamina propria, the ligament consists of the intermediate and deep layers of the lamina propria, and muscle refers to the thyroarytenoid muscle (Hirano, 1974, 1975, 1977; Hirano and Sato, 1993). In a two-layer scheme, the body is equivalent to the deep layer of the lamina propria and the muscle, and the term "cover" is used to describe the combination of epithelium, superficial, and intermediate layers of the lamina propria (Hirano and Kakita, 1985). The three schemes are summarized in Figure A10.13.

A10.3 THEORY OF VOCAL FOLD VIBRATION

A10.3.1 Myoelastic–Aerodynamic Theory

The myoelastic–aerodynamic theory states that phonation occurs when the vocal folds are approximated by muscle contraction on the arytenoid cartilages. Air from the lungs increases in speed as it flows through the narrowed glottis. The increased air flow through the glottis results in a drop in pressure along the margin of the vocal folds. When tissue pressure inherent within the folds exceeds the pressure at the glottal margin, the folds are "sucked" closer together. This effect is the Bernoulli principle. This principle states that as a gas or liquid increases in velocity across a plane, there is a pressure drop along the plane (i.e., less pressure is exerted perpendicular to the flow). Applying this to the action of the vocal folds, the continuing effect of narrowing the glottis and increasing the air flow velocity eventually closes the glottis. At this point, the subglottal air pressure increases until it exceeds the tissue pressure holding the folds together, and the folds are moved apart. The entire cycle then repeats itself.

A10.3.2 Description of Vocal Fold Vibration

We can divide a single vibratory cycle into three distinct phases to describe the vocal fold vibration (Figure A10.14). The first is an opening phase, during which the vocal folds pull apart, increasing the area of the glottal opening. Second is a closing phase, during which the vocal folds come together, reducing the glottal area. Finally, there is a closed phase, during which the vocal folds are maximally closed. Note that in some vibratory modes as in a breathy voice, a distinct closed phase may not exist and the area of the glottal opening shows an almost sinusoidal variation with time.

Based on observations using excised larynges (Baer, 1981a, 1981b), ultra-high speed photography (Hildebrand, 1976; Moore, 1975; Timcké, von Lenden, and Moore, 1958), ultrasonography (Hamlet, 1981), and x-ray stroboscopy (Saito et al., 1981), the movements of the vocal folds during these three phases in normal chest (modal) voice can be described as follows.

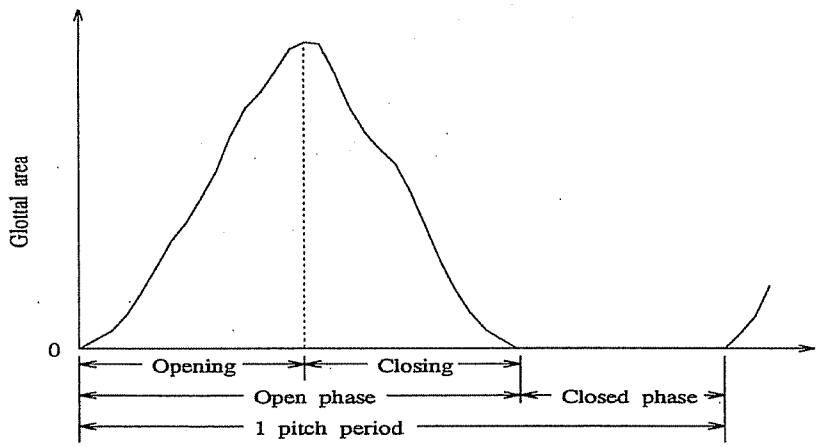


FIGURE A10.14 Divisions of the glottal area cycle.

During the opening phase, the vocal folds first separate inferiorly and the opening moves upward with a wave-like motion in the mucous membrane. Occasionally, the opening first appears on the superior surface as a small "chink" that opens up in a "zipper" like fashion (Baken, 1987; Childers et al., 1986; Childers et al., 1990). The closing phase begins with contact between the lower edges of the glottis. The closure then proceeds along the length of the lower edge and is followed by the mucosal layers coming together. The closed phase is not necessarily associated with an increasing amount of contact between the vocal folds. It is often observed (Baer, 1981a) that as the vocal folds come into contact in a vertical plane, they may be pulling apart at the same time in a different vertical plane.

A schematic representation of vocal fold vibration from Stevens and Klatt (1974) is shown in Figure A10.15. The sketches numbered 1 to 7 in part *a* represent schematized sections through the

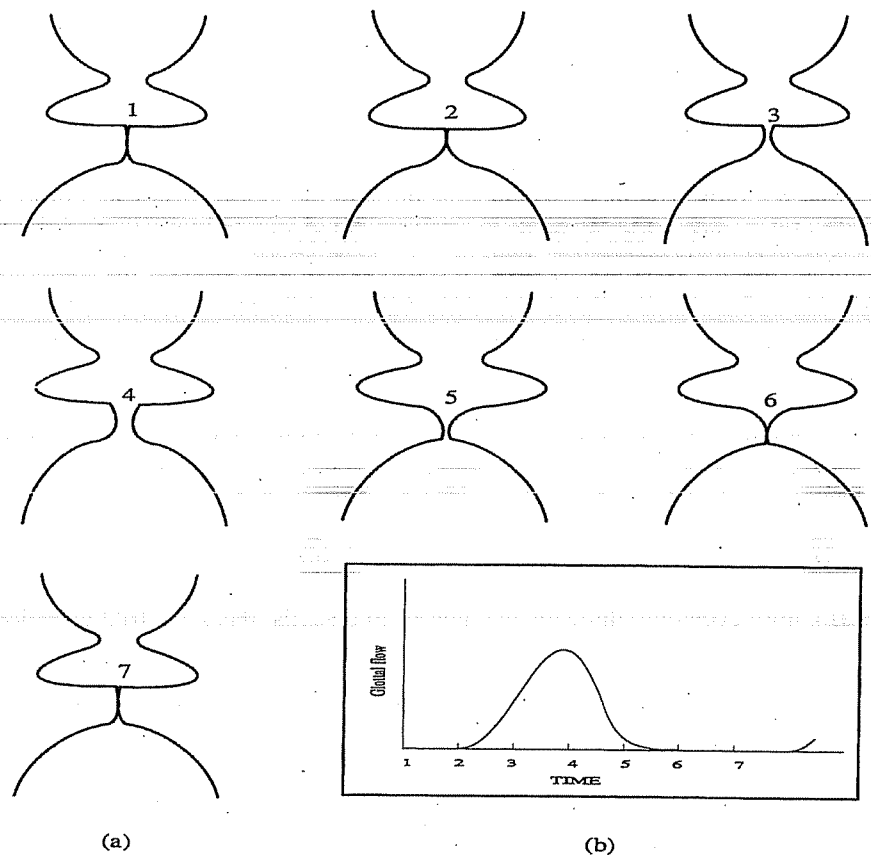


FIGURE A10.15 A schematic representation of vocal fold vibration.

larynx at various instants of time during a cycle of vocal fold vibration. A sketch of the waveform of the volume velocity through the glottis is given in *b*. The points indicated on the sketch correspond to the various sections in part *a* of the figure. These sketches are not based on actual measurements, but are derived from reports of stroboscopic tomography of the larynx, and observations from high speed and stroboscopic pictures of the vibrating larynx. See Childers et al. (1986) for other more detailed descriptions. The vocal fold model toolbox in Chapter 9 shows this motion.

A10.4 REVIEW OF VOCAL FOLD VIBRATORY MODELS

The dynamics of vocal folds have been extensively studied for several decades and a number of models of the vocal folds have been developed. These models include the one-mass model (Flanagan and Landgraf, 1968), the two-mass model (Flanagan and Ishizaka, 1978; Ishizaka and Flanagan, 1972, 1977), the multiple-mass model (Titze, 1973, 1974), the continuum model (Titze and Strong, 1975; Titze and Talkin, 1979a), the four-parameter model (Titze, 1984, 1989), and the body-cover model (three-mass model) (Story and Titze, 1995). One can classify the above models into four main categories as follows. Mechanical models include the one-mass model, two-mass model, and multiple-mass model; and the continuum model, ribbon model with four parameters, and the body-cover model with three-masses.

A10.4.1 Mechanical Models

We can classify the one-mass, two-mass, and multiple-mass models as mechanical models because these models represent the glottal source as lumped mechanical oscillators. In a lumped mechanical model, the subglottal system is represented by an air reservoir with pressure (P_s) that provides an air flow with the volume velocity (U_g). The vocal folds are mechanically modeled by an oscillatory system of masses, viscous damping, and springs (Titze and Strong, 1975; Titze and Talkin, 1979a).

A10.4.1.1 One-Mass Model In the one-mass model, the vocal fold vibrations are modeled with a single mass-spring oscillator driven by airflow from the lungs as shown in Figure A10.16. The one-mass model has the following characteristics. It is simple with a low computational burden. There is source-tract interaction. The phase-difference between the motion of fold edges is disregarded and the glottal area and volume velocity can be simulated.

A10.4.1.2 Two-Mass Model In the two-mass model, the vocal folds are divided in depth into an upper and a lower mass due to the anatomic and functional division between the mucosa and the vocalis. Each part consists of a simple mechanical oscillator having mass, spring, and damping (m, s, r) as in Figure A10.17. The springs represent the elastic properties of the folds. The damping represents dissipative forces such as viscosity and friction. There is also an interaction between the two masses, represented by a coupling stiffness S_c. The coupling stiffness represents the

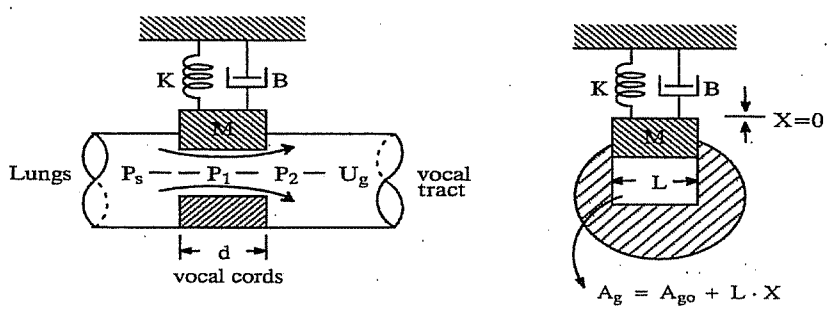


FIGURE A10.16 One-mass model of the vocal folds.

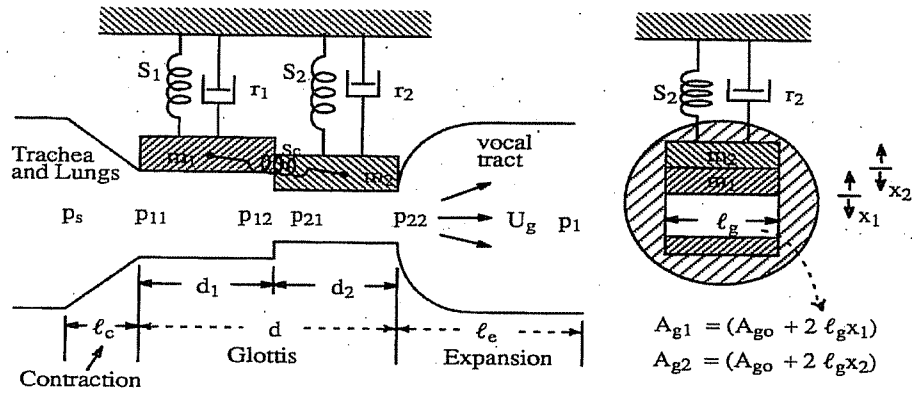


FIGURE A10.17 Two-mass model of the vocal folds.

fact that as one of the masses is displaced relative to the other, there is a force tending to restore the masses to their equilibrium position relative to one another. The two-mass model has the following characteristics: realistic simulation of glottal properties, phase-difference between the motion of fold edge is considered [the mucosal surface wave is not considered in Ishizaka and Flanagan (1972), but it is in the modified model of Koizumi et al. (1987)]. Natural speech can be produced with a reasonable computational burden.

A10.4.1.3 Multiple-Mass Model Although the two-mass model is a milestone in quantifying vocal fold vibration, it models only the vocal folds as a minimal mechanical structure capable of responding to aerodynamic forces and sustaining oscillation. It is not capable of exhibiting various longitudinal vibratory modes observed in human phonation. Titze (1973), in an attempt to enlarge the horizontal degrees of freedom, proposed a 16-mass model composed of two rows of eight masses each (Figure A10.18). The top-row of masses represents primarily the mucosa and the bottom row represents primarily the vocal ligament and the vocalis muscle. The forces T_m and T_v represent the longitudinal tensions as determined by the balance of forces between the cricothyroid and thyroarytenoid muscles. Specifically, the spring constants for the upper and lower rows increase nonlinearly with elongation of the vocal folds. The 16-mass model has the following characteristics. It is complex, with a high computational burden. The mucosal surface wave can be simulated. It has the ability to regulate the model by parameters that have direct physiologic correlates. There is increased naturalness of the utterances, and the phonation is in at least two distinct registers.

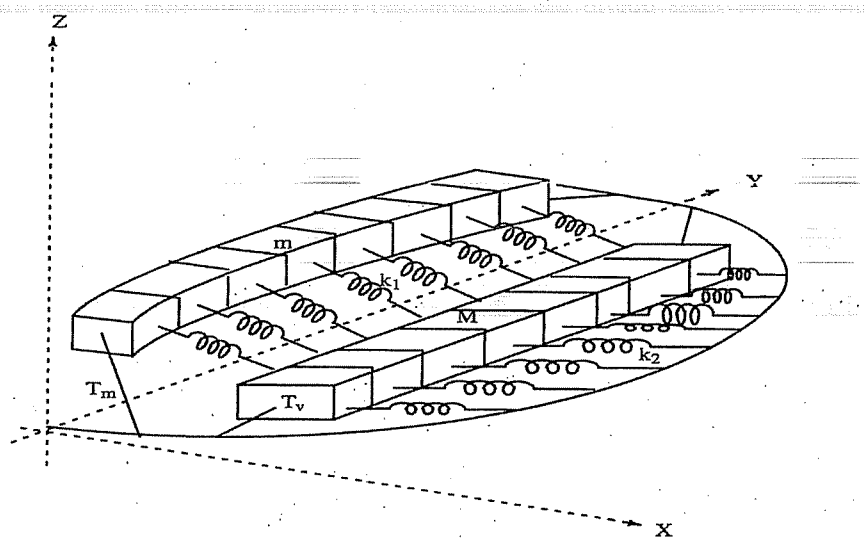
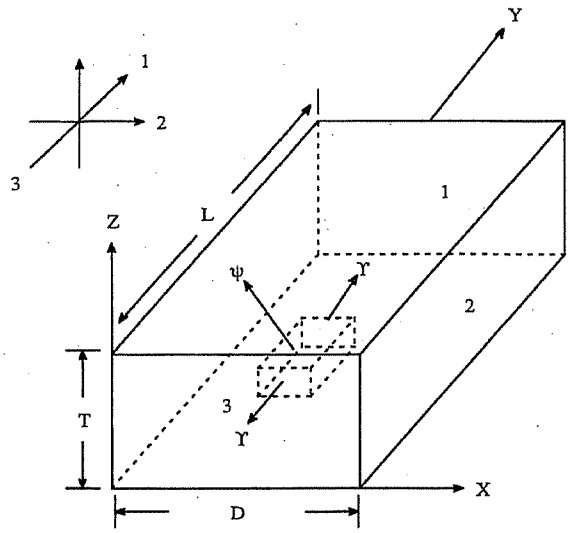


FIGURE A10.18 A 16-mass model of the vocal folds.



The origin of the coordinate system is centered at the vocal processes.

Within the rectangular parallelepiped representing the vocal fold.

Surfaces 1, 2, and 3 are fixed, and others are free.

ψ : Displacement vector of the differential element

Υ : Longitudinal stress of the differential element

FIGURE A10.19 A continuum model of the vocal folds.

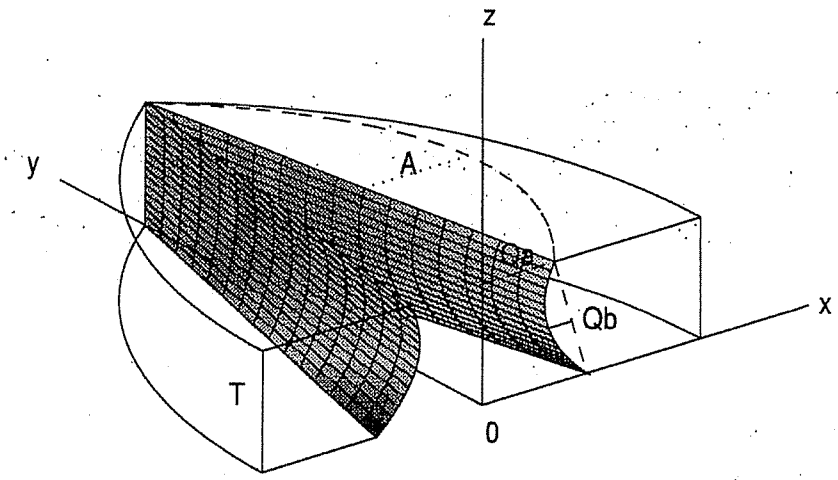
A10.4.2 Continuum Model

One could extend the 16-mass model by increasing the number of masses and degrees of freedom. Titze and Strong (1975) went a step further and represented the vocal folds not as a coupled set of discrete masses but as a continuous deformable medium (Figure A10.19). The incompressibility of the vocal folds dictates a coupling between the horizontal and vertical motion. An important consequence of the incompressibility of the vocal folds is that the most easily excited vibratory mode appears to involve vertical phase differences, since this mode tends to preserve the volume of the vocal folds. The Titze and Strong (1975) study also showed that the layered structure of the vocal folds is ideally adapted to support vocal fold vibration. The longitudinal fibrous structure is more loose in the vertical direction than in the longitudinal direction. This allows vertical phase differences to occur (Chan, 1989).

The continuum model is highly informative concerning the relationship between the vocal fold structure and the vocal fold vibratory modes. However, the shape of the vocal folds in this model is restricted to a rectangular form. The tissue properties are uniform in the plane normal to the longitudinal direction for ease of manipulation (Titze, 1976). In addition, the model lacks a complete representation of the interaction between the aerodynamic air flow and the elastic vocal fold tissue because the normal modes of vocal fold vibration are derived based on an eigen-value analysis of the fold tissue.

A10.4.3 Ribbon Model

Vocal fold vibration occurs mainly in a thin layer of the nonmuscular tissue at the vocal fold surface. It is estimated that the effective depth of vibration into the vocal fold is on the order of 1 mm. Hence, one can think of the vibrating portion as a stretched ribbon that is fixed at the horizontal endpoints ($Y = 0$ at the posterior arytenoid part, $Y = L$ at the anterior thyroid part) but is free to bend and flex in the vertical dimension between those endpoints. The motion of the ribbon, therefore, can be described by a wave equation with appropriate boundary conditions, and its eigenfunction will give the approximate vibration patterns of vocal folds. Based on this concept, a kinematic four-parameter model (Figure A10.20) for the three-dimensional glottis was presented by Titze (1984, 1989). Titze's model can



A : unit amplitude of vibration
Q_s : shape quotient
Q_a : abduction quotient
Q_b : bulging quotient
Displacement function : d(y,z,t)

$$d(y,z,t) = 2*[\, h_o(y,z) + h_1(y,z,t)\,]$$

$$h_o(y,z) = [Q_a+(Q_s - 4*Q_b*z/T)*(1 - z/T)](1 - y/L)$$

$$h_1(y,z,t) = \sin(\pi\, y/L) * \sin(\omega t - \omega z/c)$$

FIGURE A10.20 Ribbon model.

provide the glottal flow, glottal area, and vocal fold contact area waveforms. The static glottis is controlled by the abduction quotient (Q_a), the shape quotient (Q_s), and the bulging quotient (Q_b). The phase quotient (Q_p) and fundament frequency (F_o) control the dynamic glottis. The displacement function, $h_1(y, z, t)$, is used to calculate the glottal area and is a sinusoidal function.

A10.4.4 Body-Cover Model (Three-Mass Model)

The body-cover concept (Hirano, 1974) is generally used to describe the layered structure of vocal folds (Figures A10.12 and A10.13). It suggests that the vocal folds can be divided into two tissue layers with different mechanical properties. The body layer consists of muscle fibers and some tightly connected collagen fibers of the vocal ligament. The cover layer consists of pliable, noncontractile tissue that acts as a flexible sheath around the body layer. The cover typically is loosely connected to the body during vibration. The motion of the cover layer is usually observed as a surface wave that propagates from the bottom of the vocal folds to the top, thus experiencing movement in both the lateral and vertical directions. Self-sustained vocal fold oscillation is highly dependent on this surface-wave behavior (typically referred to as the vertical phase difference) and is the primary mechanism for transferring energy from the glottal flow to the tissue to fuel the vibration. The body layer is primarily involved in lateral motion. Based on his findings, Hirano (1974) suggests that the vocal folds should be treated as a double structured vibrator with stiffness parameters that should be based on the relative actions of the thyroarytenoid and cricothyroid muscles. Thus, the resultant vibration of the vocal folds is composed of the coupled oscillations of the body and cover layers.

In the two-mass model, the lower mass is made thicker (vertical dimension in the coronal plane) and more massive than the upper element in an attempt to include the effects of the body layer. But, because a provision does not exist for coupled oscillation of both layers, the two-mass model is essentially a "cover" model rather than a "body-cover" model. In order to more realistically represent the body-cover vocal fold structure, Story and Titze (1995) extended the two-mass model into a three-mass model (Figure A10.21).

The three-mass model consists of two "cover" masses coupled laterally to a "body" mass by nonlinear springs and viscous damping elements. The body mass, which represents muscle tissue, is

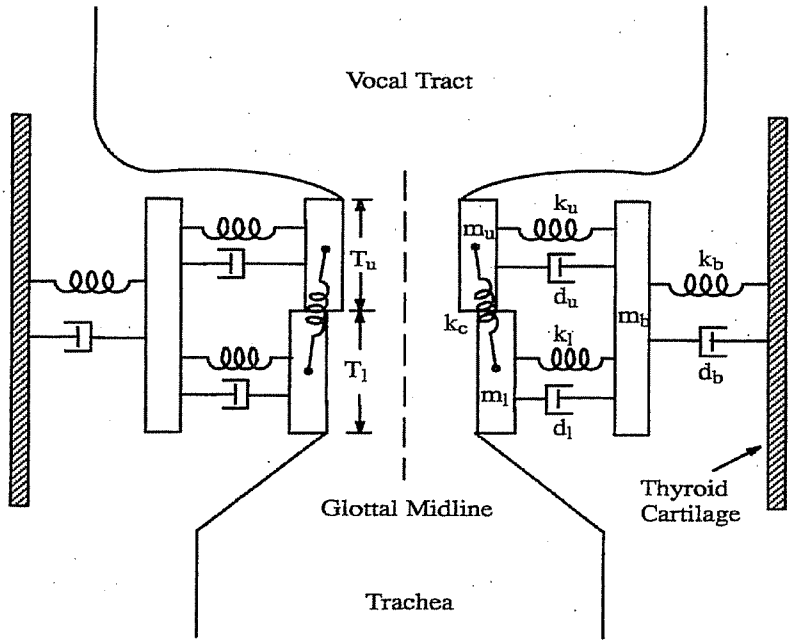


FIGURE A10.21 Body-cover model of the vocal folds.

further coupled laterally to a rigid wall (assumed to represent the thyroid cartilage) by a nonlinear spring and a damping element. The two cover springs are intended to represent the elastic properties of the epithelium and the lamina propria, while the body spring simulates the tension produced by contraction of the thyroarytenoid muscle. Thus, contractions of the cricothyroid and thyroarytenoid muscles are incorporated in the values used for the stiffness parameters of the body and cover springs. The two cover masses are coupled to each other through a linear spring, which can represent vertical mucosal wave propagation.

A10.5 THE TWO-MASS VOCAL FOLD MODEL: THEORY

This section describes the purpose for and the implementation of the two-mass model (Ishizaka and Flanagan, 1972). This is followed by the development of an acoustic model, and some examples.

A10.5.1 Purpose

The two-mass model provides sufficient theory to investigate variations of and the relationship among kinematic parameters of a vocal fold model as well as selected physiologic parameters; for example, pre-phonatory shape of the glottis, fold tension, and lung pressure. The kinematic parameters include fundamental frequency, vertical phase difference, vertical equilibrium shape, vertical amplitude profile, and maximum excursion of the vocal folds during vibration. There are limitations to the model. For example, variations of the vocal fold dynamics along the length of the glottis cannot be simulated because the two-mass model can only account for vertical dissimilarities.

A10.5.2 Acoustic Model for the Implementation of Two-Mass Model

An acoustic model is required to implement the two-mass model (Ishizaka and Flanagan, 1972). The acoustic model is a simple articulatory speech synthesizer, which does not include a noise source model or a nasal tract model. Usually, an articulatory speech synthesizer contains an articulatory model and an acoustic model. The articulatory model maps the positions of the key articulators, such as the jaw, tongue, lips, and velum, to the cross-sectional area function of the vocal tract. The

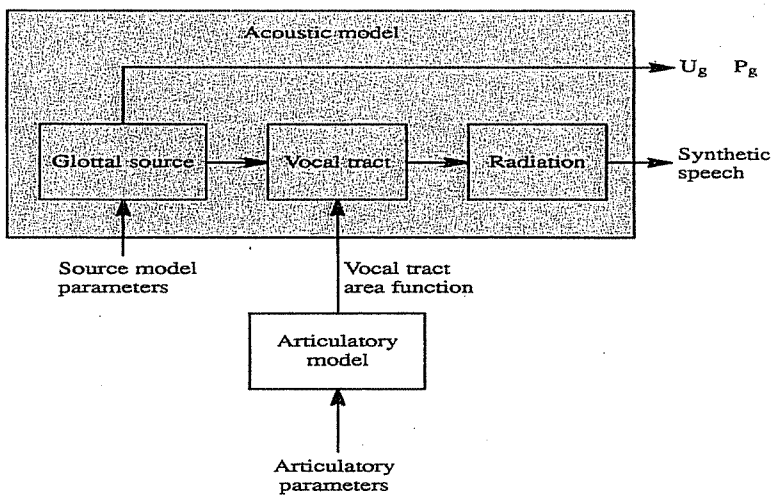


FIGURE A10.22 Acoustic model for the two-mass vocal fold model.

acoustic model is a set of ordinary differential equations (acoustic equations) that describe the acoustic properties of the vocal system. To obtain synthetic speech, one can solve the acoustic equations using numerical methods. The basic structure of the acoustic model for the implementation of the two-mass model is shown in Figure A10.22.

A10.5.2.1 Acoustic Model of the Glottal Source The schematic diagram of the two-mass model (Ishizaka and Flanagan, 1972) appears in Figure A10.17. The motions of the masses in the two-mass model are governed by the aerodynamic forces that activate the larynx as well as the myoelastic forces of the springs and the dampers. The basic equations of motion are

$$m_1 \frac{d^2x_1}{dt^2} + r_1 \frac{dx_1}{dt} + k_1 x_1 + k_c(x_1 - x_2) + F_1 = 0 \qquad \text{(A10.5.2.1.1)}$$

$$m_2 \frac{d^2x_2}{dt^2} + r_2 \frac{dx_2}{dt} + k_2 x_2 + k_c(x_2 - x_1) + F_2 = 0 \qquad \text{(A10.5.2.1.2)}$$

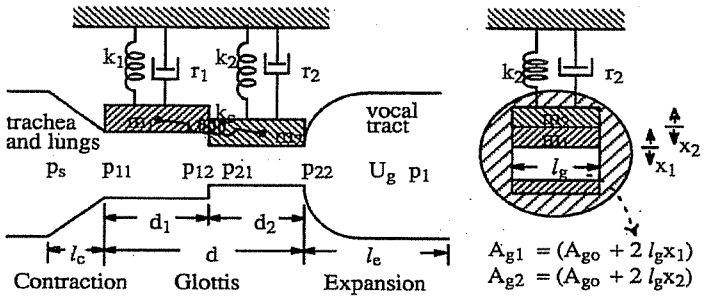
where x_i represents the lateral displacement of the two masses, F_i represents the aerodynamic forces exerted on each mass, r_i represents the viscous loss (resistance), $i = 1$ for the lower mass, $i = 2$ for the upper mass, $r_i = 2\zeta_i\sqrt{m_i k_i}$, ζ is the damping ratio, m is the mass, and k is the spring constant. In the model, the springs are given a nonlinear characteristic to conform to the stiffness as measured from excised human vocal folds (Ishizaka and Flanagan, 1972). During closure of the glottis, there is a contact force that results in additional deformation. The spring restoration forces are represented as

$$f_{si} = k_i x_i(1 + \eta_{ki} x_i), \quad \text{for } i = 1, 2 \qquad \text{(A10.5.2.1.3)}$$

$$f_{hi} = h_i(x_i + x_{oi})(1 + \eta_{hi}(x_i + x_{oi})), \quad \text{for } (x_i + x_{oi}) < 0 \qquad \text{(A10.5.2.1.4)}$$

where $i = 1$ for the lower mass, $i = 2$ for the upper mass, k and η_k are the linear and nonlinear stiffness parameters of the spring, and h and η_h are additional linear and nonlinear stiffness parameters of the spring during the closed glottal phase.

The physiologic constants of the two-mass model are set to values suggested by Ishizaka and Flanagan (1972) and are summarized in the Figure A10.23. Note that the lower mass (the vocalis–ligament combination) is five times as massive and five times as thick as the upper mass (the mucous membrane). For simulation of normal voices, the parameter values in Figure A10.23 are held constant, except for control of the fundamental frequency. Ishizaka and Flanagan (1972) proposed a tension parameter, Q, to control the fundamental frequency. The fundamental frequency in the range of 120 to 220 Hz varies almost linearly with the tension parameter Q. For other vocal registers such as falsetto and vocal fry, the parameter values in Figure A10.23 are no longer valid. Similar problems occur when attempts are made to simulate the vibratory pattern of abnormal vocal folds. Therefore, the parameters in the two-mass model must be changed according to laryngeal adjustments for other voicing modes.



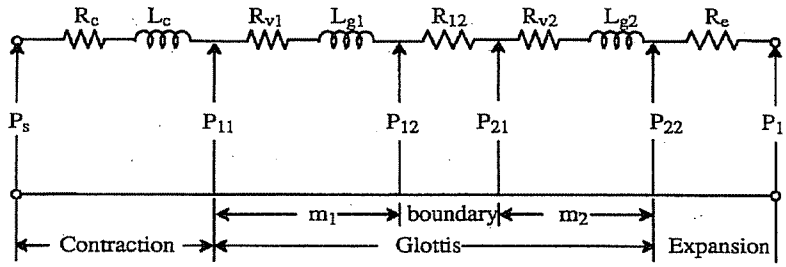
$m_1 = 0.125$ g	lower mass
$m_2 = 0.025$ g	upper mass
$d_1 = 0.25$ cm	thickness of m_1
$d_2 = 0.05$ cm	thickness of m_2
$l_g = 1.5$ cm	effective length of vocal folds
$k_2 = 8000$ dyne/cm	linear stiffness of spring 2
$k_1 = 80000$ dyne/cm	linear stiffness of spring 1
$k_c = 25000$ dyne/cm	stiffness of coupled spring
$\eta_{k1} = \eta_{k2} = 100$ dyne/cm	nonlinear stiffness of spring 1 and 2
$\eta_{h1} = \eta_{h2} = 500$ dyne/cm	nonlinear stiffness of contact springs
$h_1 = 3 \cdot k_1$	linear stiffness of contact spring 1
$h_2 = 3 \cdot k_2$	linear stiffness of contact spring 2
$\zeta_1 = 0.1$	damping ratio for open phase
$\zeta_2 = 0.6$	damping ratio for open phase
$\zeta_1 = 1.1$	damping ratio for closed phase
$\zeta_2 = 1.6$	damping ratio for closed phase

FIGURE A10.23 The physiologic constants for the two-mass model.

The acoustic impedance elements of the glottal orifice constitute an equivalent circuit of the glottis shown in the Figure A10.24, where R_c represents the abrupt contraction at the inlet to the glottis; R_{v1} and R_{v2} represent viscous losses at the lower-fold edge, upper-fold edge, respectively; R_{12} represents the change in kinetic energy per volume of fluid at the junction between masses m_1 and m_2; R_e represents the expansion of the glottal outlet; L_c, L_{g1}, and L_{g2} represent the inertances of the air masses (Ishizaka and Flanagan, 1972).

A10.5.2.2 Acoustic Model of the Vocal Tract The vocal tract is a three-dimensional lossy cavity composed of nonuniform cross-sections and nonrigid walls (Sondhi, 1974; Sondhi, 1986). Although the appropriate Navier–Stokes equations with the boundary conditions for nonrigid walls describe the acoustic properties of the vocal tract, a large number of calculations are required to solve such equations and neither the shape of the vocal tract nor the physical properties of the walls are known with sufficient accuracy to establish a reliable model. These limitations suggest the need for a simplified version of the acoustic model of the vocal tract. One simplification is to assume plane wave propagation in the vocal tract. This is reasonable since, first, the soft tissue along the vocal tract prevents radial propagation of the sound wave, and, second, the average lateral (cross-sectional) dimension of the vocal tract is about 2 cm, which is smaller than the wavelength of a sound wave at 4 kHz, which is $\lambda = c/f = 34{,}300/4000 = 8.6$ cm. Strictly speaking, this assumption is valid only for frequencies below 4 kHz. But for speech, where 5 kHz is considered to be an appropriate bandwidth, the plane wave propagation assumption is quite adequate.

With the assumption of plane wave propagation in the vocal tract, then only the cross-sectional area and the perimeter along the length of the vocal tract determine the acoustic characteristics of the vocal tract. Thus, the acoustic equations can be described in one dimension instead of three, which is a significant simplification. The area function of the vocal tract is then approximated by a sufficiently small number of successive sections with each section having a constant cross-sectional area.



$$R_c = 1.37\frac{\varrho}{2}\frac{|U_g|}{A^2_{g1}} \qquad L_c = \int_0^{l_c}\frac{\varrho}{A_c(x)}dx$$

$$R_{v1} = 12\frac{\mu l^2_g d_1}{A^3_{g1}} \qquad L_{g1} = \frac{\varrho d_1}{A_{g1}}$$

$$R_{12} = \frac{\varrho}{2}\left(\frac{1}{A^2_{g2}} - \frac{1}{A^2_{g1}}\right)|U_g| \qquad R_e = -\frac{\varrho}{2}\frac{2}{A_{g2}A_1}\left(1 - \frac{A_{g2}}{A_1}\right)|U_g|$$

$$R_{v2} = 12\frac{\mu l^2_g d_2}{A^3_{g2}} \qquad L_{g2} = \frac{\varrho d_2}{A_{g2}}$$

U_g = glottal flow — P_s = lung pressure

ϱ = air density — μ = shear viscosity coefficient

l_c = the length of contraction area — A_g = glottal area

A_c = cross-sectional area of the contraction region

A_1 = cross-sectional area of the 1st vocal tract element

FIGURE A10.24 Equivalent circuit for the glottis.

For each section of the vocal tract, the acoustic model is derived as follows. Portnoff (1973) has shown that sound waves in the lossless tube satisfy the following equations.

$$-\frac{\partial p}{\partial x} = \rho\frac{\partial(u/A)}{\partial t} \qquad \text{(A10.5.2.2.1)}$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2}\frac{\partial(pA)}{\partial t} + \frac{\partial A}{\partial t} \qquad \text{(A10.5.2.2.2)}$$

where $p = p(x, t)$ is the sound pressure, $u = u(x, t)$ is the volume velocity, ρ is the density of air, c is the velocity of sound, and $A = A(x, t)$ is the area function of the tube. Applying Equations (A10.5.2.2.1) and (A10.5.2.2.2) to the section specified by the cross-sectional area A yields

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A}\frac{\partial(u)}{\partial t} \qquad \text{(A10.5.2.2.3)}$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2}\frac{\partial(p)}{\partial t} \qquad \text{(A10.5.2.2.4)}$$

Based on the similarity between these equations and the equations for lossless, uniform electrical transmission lines, the tube with length l can be represented by an inductance, $L = \frac{\rho l}{A}$, followed by a shunt capacitance, $C = \frac{Al}{\rho c^2}$. The similarity between the acoustic wave propagation in a cylindrical tube and the propagation of an electrical wave along a transmission line are summarized in Chapter 5.

The effects of the vibration of the vocal tract wall can be added to the above model. The pressure variations inside the vocal tract will cause the cross-sectional area to change, since it exerts a force on the tract's elastic walls. Assuming that the walls are subject to local reactions (i.e., the motion of one portion of the wall is dependent only upon the acoustic pressure on that portion and independent of the motion of any other part of the wall), the area A(x, t) will be a function of the pressure p(x, t). Since the pressure variations are very small, the resulting variation in the cross-sectional area can be treated as a small perturbation,

$$A(x, t) = A_0 + \Delta A = A_0 + yS_0 \qquad \text{(A10.5.2.2.5)}$$

where A_0 is the nominal area, ΔA is a small perturbation, S_0 is the circumference of the tube, and y is the displacement of the yielding walls due to the sound pressure inside the tube. The wall vibration is modeled as a mass–compliance–viscosity mechanical model and is governed by Newton's law. Let m, b, and k represent the mass, the mechanical resistance; and the stiffness of the wall per unit length of the tube, respectively. According to Newton's law

$$m\frac{\partial^2 y}{\partial t^2} + b\frac{\partial y}{\partial t} + ky = pS_0 \qquad \text{(A10.5.2.2.6)}$$

Define the volume velocity generated by the wall vibration as

$$u_w = \frac{\partial(yS_0\ell)}{\partial t} \qquad \text{(A10.5.2.2.7)}$$

These two equations can be combined to obtain

$$p = \frac{m}{S_0^2\ell}\frac{\partial u_m}{\partial t} + \frac{b}{S_0^2\ell}u_m + \frac{k}{S_0^2\ell}\int u_m\,dt \qquad \text{(A10.5.2.2.8)}$$

The equivalent circuit for this equation is an RLC series circuit where

$$L_w = \frac{m}{S_0^2\ell}$$

$$R_w = \frac{b}{S_0^2\ell}$$

and

$$C_w = \frac{S_0^2\ell}{k}$$

are the components of the wall vibration impedance.

The wall impedance can be included in every elemental section of the vocal tract as a distributed element (Flanagan, 1972a; Flanagan and Ishizaka, 1976; Flanagan et al., 1975, 1980; Ishizaka et al., 1975; Maeda, 1982a) or inserted as a lumped shunt element, one for the pharynx and one at the level of the cheeks (Badin and Fant, 1984; Lin, 1990; Wakita and Fant, 1978). As Wakita and Fant (1978) indicated, the lumped wall impedance, which is independent of the vocal tract configurations may not give satisfactory results.

Figure A10.25 presents data concerning the wall mass, viscosity, and compliance found in the literature. In some cases, the compliance is not used since it has no effect on the resonances of the model (Wakita and Fant, 1978). Maeda (1982a) pointed out that the total mass of the walls may vary unrealistically if the yielding wall parameters are specified in terms of a unit surface area. Thus, the per unit length specification was used in his vocal tract simulation (Hsieh, 1994).

The effects of viscous friction and thermal conduction at the wall sites are much less pronounced than those for the wall vibration. Flanagan (1972) considered these losses in detail and showed that the effects of viscous friction can be accounted for by including a frequency-dependent resistor, R, in series with the inductor, L. The effects of heat conduction through the vocal tract wall can be accounted for by adding a frequency-dependent resistor, $\frac{1}{G}$, in parallel with the capacitor, C. The resistor, R,

	b (gm/sec)	m (gm)	k (dyne/cm)
Ishizaka et al. (1975), at cheek	1060	1.5	33,300
Ishizaka et al. (1975), at neck	2320	2.4	49,100
Flanagan et al. (1975)	1600	1.5	–
Maeda (1982)	1400	1.5	30,000
Lin (1990a)	1600	1.4	–

FIGURE A10.25 Per unit area yielding wall parameters.

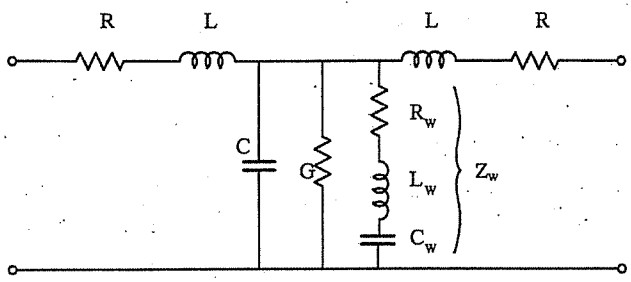


FIGURE A10.26 Equivalent circuit of an analog transmission line element of the vocal tract.

is significant in time domain simulations; when a constriction occurs, the resistance becomes very large and the air flow is blocked. As a result, a section of the vocal tract may be represented by a finite number of transmission line elements whose structure is given in Figure A10.26. The definitions of the circuit components are given in Figure A10.27. The series resistor R is used to represent the acoustic loss due to viscous drag in which the energy loss is proportional to the square of the volume velocity. The shunt conductance G represents the loss due to heat conduction, which is proportional to the pressure squared. The shunt impedance is the acoustic equivalent mechanical impedance of the

$$R = \frac{S\sqrt{\varrho\mu\omega}}{2\sqrt{2}A^2}l \quad ; \text{Series resistance}$$

$$L = \frac{\varrho}{2A}l \quad ; \text{Series inductance}$$

$$C = \frac{A}{\varrho c^2}l \quad ; \text{Shunt capacitance}$$

$$G = \frac{(\eta - 1)S}{\varrho c^2}\sqrt{\frac{\lambda\omega}{2\xi\varrho}}l \quad ; \text{Shunt conductance}$$

$$R_w = \frac{b}{S^2 l} \quad ; \text{Resistance in wall impedance}$$

$$L_w = \frac{m}{S^2 l} \quad ; \text{Inductance in wall impedance}$$

$$C_w = \frac{S^2 l}{k} \quad ; \text{Capacitance in wall impedance}$$

where

$S = 2S_A\sqrt{A\pi}$: circumference of element.
S_A: section shape factor, for a circular cross-section, S_A=1;
for an elliptic cross-section, S_A=2.
l : length of elemental tube.
A : cross-sectional area of element.
ϱ : density of air, 1.14×10^{-3} gm/cm^3 (moist air at body temperature, 37°C).
c : sound velocity, 3.53×10^4 cm/sec (moist air at body temperature, 37°C).
μ : viscosity, 1.86×10^{-4} dyne-sec/cm^2 (20°C, 0.76 m.Hg).
λ : coefficient of heat conduction of air, 0.055×10^{-3} cal/cm-sec-deg (0°C).
η : adiabatic gas constant, 1.4.
ξ : specific heat, 0.24 cal/gm-degree (0°C, 1 atmos.).
ω : radian frequency.

FIGURE A10.27 Physical definitions of the components in Figure A10.26.

yielding wall. This wall impedance, which represents a mass–compliance–viscosity loss of the soft tissue, has three components, R_w, L_w, and C_w. The acoustic model of the vocal tract is established by concatenating these element models. No standardized model of the vocal tract has been established to date and the choices for the component values vary among researchers. The values representing the best choices remain to be determined (Wakita and Fant, 1978).

A10.5.3 Acoustic Model of the Radiation

Acoustic energy escapes from the vocal tract via the lips. From the transmission-line analogs, the lips are treated as a radiation impedance that loads the vocal tract. The radiation impedance contains a resistive part that represents acoustic energy loss and a reactance part that represents the mass inertia of air at the lips (Fant, 1960). Radiation from a spherical baffle is one model for the radiation impedance that is represented by nonlinear functions (Morse, 1948; Morse and Ingard, 1968). Stevens et al. (1953) made approximations and represented the radiation impedance by a resistive load with three other frequency-dependent components. Fant made another approximation and modeled the impedance by two frequency-dependent components, one being resistive and the other inductive (Fant, 1960; Wakita and Fant, 1978).

Another simplified radiation model is to assume that the radiating surface is set in a plane baffle of infinite extent. In this case, the radiation impedance is formed by a first order Bessel function and a Struve function (Flanagan, 1972; Rayleigh, 1945; Wakita and Fant, 1978). Flanagan (1972) provided a good approximation to this complicated representation by a parallel connection of a resistance and an inductance. The most important feature of Flanagan's model (1972) is that both circuit components are frequency independent. Figure A10.28 illustrates the Stevens et al. (1953) model and Flanagan (1972) model.

Comparisons between models have been made by researchers (Badin and Fant, 1984; Lin, 1990; Wakita and Fant, 1978). The Stevens et al. (1953) model yields the most accurate result. However, the Flanagan (1972) model is usually preferred for time-domain synthesis (Flanagan and Ishizaka, 1976; Flanagan et al., 1975, 1980; Maeda, 1982a) and is used in our acoustic model of radiation.

A10.5.4 Example of a Synthetic Vowel

One example of a synthesized vowel, /AA/ as in father, is presented. The vocal tract area function is adopted from the optimized results of Hsieh (1994), and is similar to the data in Appendix 5. The synthetic glottal area, glottal flow, speech waveforms, and the wideband spectrograms for vowel /AA/ are shown in the Figure A10.29.

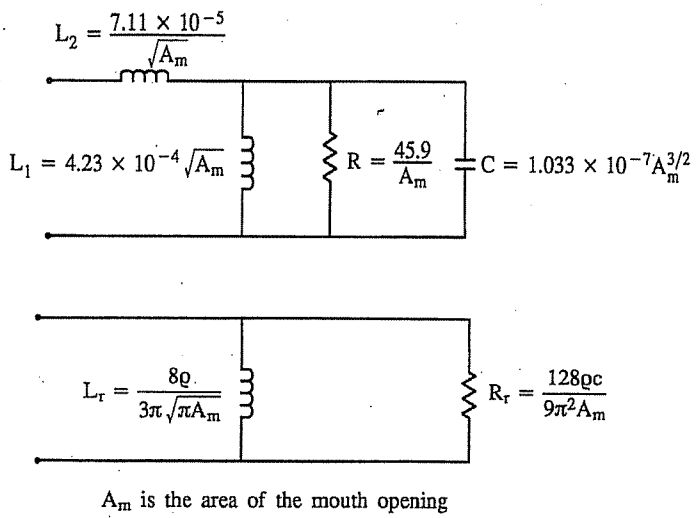


FIGURE A10.28 Radiation models. (*a*) Model by Stevens et al. (1953). (*b*) Flanagan model (1972).

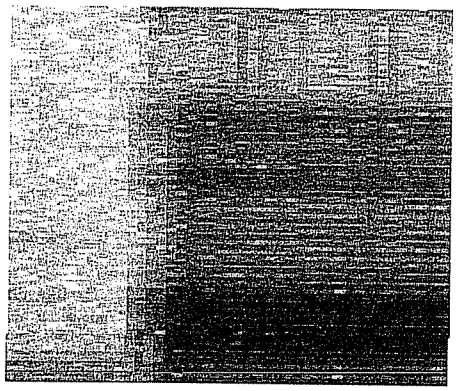


FIGURE A10.29 Synthesized glottal area, glottal flow, speech waveform, and wideband spectrogram for the vowel /AA/.

A time-domain comparison of the synthetic waveforms with the results of Ishizaka and Flanagan (1972) and a frequency-domain comparison of the formant frequencies and bandwidths, measured from the wideband spectrograms, with the vocal tract frequency responses from the optimized results of Hsieh (1994), are satisfactory. Thus, the two-mass vocal fold model and the synthesizer are appropriate tools for modeling the vibratory motion of the vocal folds.

In the implementation of the two-mass model in Chapter 9, the two masses are shown as two semirigid, connected panels or ribbons rather than as two blocks as seen in Figure A10.17. This convention is used since it is similar in form to the ribbon model, which is discussed next.

A10.6 THE RIBBON MODEL

A brief introduction to the ribbon model is given at the beginning of this appendix. The ribbon model (model I) is based on Titze (1984). One modification to the model includes the design of a flexible three-dimensional model of the vocal fold vibrations.

A10.6.1 Basic Objectives of the Model

Due to the relative inaccessibility of the larynx, imaging and modeling of the vocal folds are difficult, as described previously. Thus, one aim of the model is to characterize the vocal fold vibratory characteristics in terms of a three-dimensional glottal configuration with spatially varying tissue properties. Another aim of the model is to estimate several glottographic waveforms. The three glottographic waveforms obtained with the model are the projected glottal area, the vocal fold contact area, and the electroglottographic waveform.

A10.6.2 Assumptions for the Model

Several assumptions are made to simplify the mathematical expressions in the vocal fold vibratory model. The vibratory movement of vocal folds is confined to the lateral direction. Although motion of the vocal fold tissue does occur in other directions, especially after collision between the two folds, the projected area of the glottis is primarily determined by the vocal fold lateral motion. An algorithm accounts for changes in lateral contact area due to vertical movement when the folds collide. Each point on the medial vocal fold surface vibrates in a sinusoidal fashion; that is, as a harmonic oscillator (Titze, 1984). This is a first-order approximation to the true displacement function of vocal folds. Despite its simplicity, the model based on this approximation can simulate most features of the glottographic waveforms. On the other hand, the vibratory pattern of the vocal folds for abnormal voices can be quite irregular and other surface vibratory functions need to be examined. Uniform travelling speed is assumed for all travelling wave phenomena in the proposed model. Thus, the time delay in movement of different points on the vocal fold surface is proportional to their location in the direction of wave propagation.

A10.6.3 Choice of Model Parameters

Since each glottographic waveform reflects an average measurement of the vocal fold motion, it is natural to formulate the glottographic waveforms in terms of vocal fold kinematics. The choice of model parameters is based on their importance in developing and adjusting the self-oscillation of the vocal folds and their appropriateness in describing the kinematics of the vocal folds. Model building is essentially a systematic coordination of theoretical and empirical elements of knowledge into a joint construct. Empirical observations of the vocal fold vibrations suggest that specific information about the glottal configuration and vocal fold kinematics should be contained in the parameter set. Ultra-high speed photography of the vibrating folds (Childers and Krishnamurthy, 1985; Childers et al., 1986; Childers et al., 1990; Hirano et al., 1983; Moore and von Leden, 1958) and the results from previous vocal fold models (Ishizaka and Flanagan, 1972; Titze and Talkin, 1979a) have demonstrated that five factors are important in determining the vibratory patterns. First, there must be proper abduction and adduction of the vocal folds. It is well known that proper abduction must be achieved before vibration. Different levels of abduction may result in different vibratory modes (Ishizaka and Flanagan, 1972). Second, the vertical pre-phonatory shape of the glottis is important. Titze (1988) has shown that the vertical shape may be of significance in register control in human phonation and in determining the waveshape of the glottographic waveforms. Third, vertical phasing is important. The mucosal wave is also known as vertical phasing. A lag between the movements of the upper and lower portions of the vocal folds has been observed during phonation, except for falsetto. This phenomenon is more obvious during low pitch phonation. Photography of the vibration of the vocal folds in the frontal planes confirms the wave-like nature of the vibration (Moore and von Leden, 1958). This lag reflects the degree of coupling along the depth of the vocal folds. Fourth, there is also longitudinal phasing. A lag between the movements of the anterior and the posterior portions of the vocal folds is known as longitudinal phasing. During the opening phase, the vocal folds first separate along their posterior section and the separation progresses toward the anterior (Childers et al., 1986; Childers et al., 1990; Hirano et al., 1983). Fifth, the profiles of maximum excursion of the vibration along the length and the depth of the glottis is another factor. The maximum displacement from its equilibrium position is called the amplitude of oscillation or maximum excursion. The maximum excursion of the vocal folds and equilibrium positions may vary according to the glottal configuration.

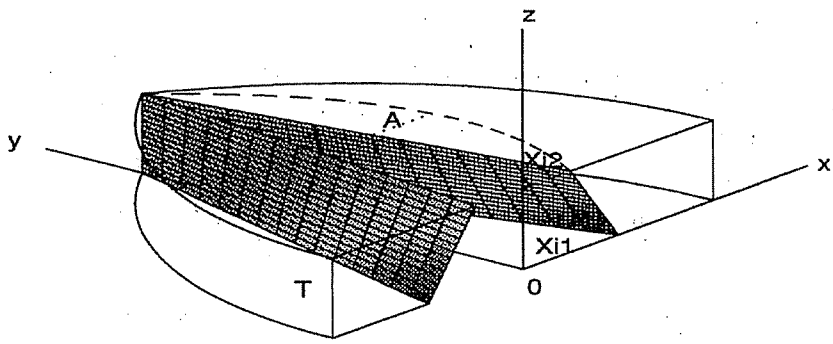
The model parameters of the ribbon model is summarized in Figure A10.30. Note that there is an adjustment phasing in the displacement function, which is described in the following paragraph.

The maximum excursion profile reflects the degree of compliance of the vocal fold tissue along each dimension. Along its length, the vocal folds appear to be most pliable near the midpoint of the membranous portion. The posterior portion of the vocal fold appears to be slightly more pliant than the anterior portion. At the anterior and the posterior ends, the tissues appear to be firm. The largest amplitude of lateral excursion, however, occurs not at the midpoint but at a place posterior to the midpoint (Hirano et al., 1983). This can be accounted for, at least in part, by the fact that the anterior end of the vocal folds can be considered fixed, while the posterior end is movable. In order to compensate for the fact that the largest amplitude of lateral excursion occurs not at the midpoint, we chose one more parameter, the adjustment phasing ϕ_m. The value of ϕ_m is set to $\frac{\pi}{4}$ in our model. However, the user can change this value.

A10.6.4 Displacement Function for the Ribbon Model

The configurational adjustments of the ribbon model can be described geometrically as illustrated in the Figure A10.30(a). Let ξ_{01} and ξ_{02} be the pre-phonatory displacements (from the glottal midline) of the inferior and superior edges of the vocal folds at the level of the vocal process. Assume that the pre-phonatory glottal shape decreases linearly toward zero at the anterior commissure and that the vertical shape of the glottis is trapezoidal. Thus, we can define the pre-phonatory glottal configuration by the equation

$$\xi_0 = \left\{\xi_{01} - (\xi_{01} - \xi_{02})\frac{z}{T}\right\}\left(1 - \frac{y}{L}\right) \qquad \text{(A10.6.4.1)}$$



Pre-phonatory Glottal Configuration :

$$\xi_0(y,z) = \left\{\xi_{01} - (\xi_{01} - \xi_{02})\frac{z}{T}\right\}\left(1 - \frac{y}{L}\right)$$

where $\xi_{01} =$ Xi1; $\xi_{02} =$ Xi2; $A = \xi_m$

L is the length of the glottis; T is vocal fold thickness

(a)

Displacement Function :

$$\xi(y,z,t) = \xi_0(y,z) + \xi_1(y,z,t) = \xi_m\left(1 - \frac{y}{L}\right)\left(Q_a + Q_s - Q_s\frac{z}{T}\right) + \xi_1(y,z,t)$$

where $$\xi_1(y,z,t) = \xi_m \sin\left(\frac{\pi y}{L}\right)\sin\left(2\pi ft - \phi_v\frac{z}{T} - \phi_h\frac{y}{L} - \phi_m\right)$$

Model Parameters :

abduction quotient $= Q_a = \dfrac{\xi_{02}}{\xi_m}$ shape quotient $= Q_s = \dfrac{(\xi_{01} - \xi_{02})}{\xi_m}$

ξ_m is the maximum excursion amplitude $\phi_m = \dfrac{\pi}{4}$ is the adjustment phasing

ϕ_v = vpd is the vertical phase delay ϕ_h = hpd is the horizontal phase delay

(b)

FIGURE A10.30 The ribbon model. (*a*) Configuration of the glottis and vocal folds. (*b*) Displacement function and model parameters.

where T is the vocal fold thickness, L is the length of the glottis, and y and z are spatial dimensions as indicated.

Making the further assumption that the displacement at the anterior and posterior boundaries is zero (i.e., fixed), and that the displacement between these boundary points is sinusoidal (Titze, 1976; Titze and Story, 1975), we approximate the dynamic displacement (from the pre-phonatory position) to be

$$\xi_1(y,z,t) = \xi_m \sin\left(2\pi ft - \phi_v\frac{z}{T} - \phi_h\frac{y}{L} - \phi_m\right) \qquad \text{(A10.6.4.2)}$$

where ξ_m is the common amplitude for the upper and lower edges of the vocal folds, ϕ_v is the vertical phase delay between the folds, ϕ_h is the longitudinal phasing, $\phi_m = \frac{\pi}{4}$ is the adjustment phasing, f is the fundament frequency of vibration, and t is time. Left-right vocal fold symmetry is also assumed.

Two configurational parameters are defined as follows.

$$\text{abduction quotient} = Q_a = \frac{\xi_{02}}{\xi_m} \qquad \text{(A10.6.4.3)}$$

$$\text{shape quotient} = Q_s = \frac{\xi_{01} - \xi_{02}}{\xi_m} \qquad \text{(A10.6.4.4)}$$

Combining the pre-phonatory displacement Equation (A10.6.4.1) with the dynamic displacement

Equation (A10.6.4.2) and substituting Q_a and Q_s, we have the following expression

$$\xi(y, z, t) = \xi_m\left(1 - \frac{y}{L}\right)\left(Q_a + Q_s - Q_s\frac{z}{T}\right) + \xi_m \sin\left(2\pi ft - \phi_v\frac{z}{T} - \phi_h\frac{y}{L} - \phi_m\right) \quad \text{(A10.6.4.5)}$$

where $\xi(y, z, t)$ is the total displacement of each vocal fold from the midline described in terms of the configurational parameters. All negative values of $\xi(y, z, t)$ are set to zero for the glottal area computations. The displacement function, model parameters, and the ribbon model are summarized in Figure A10.30.

The normalization of the pre-phonatory glottal widths ξ_{01} and ξ_{02} with respect to the amplitude of vibration ξ_m, is not only convenient, but meaningful, since it reminds us of the DC/AC ratios of glottal area and flow used in glottal leakage assessment. In other words, the relative width of the glottal chink with respect to the maximum glottal width is more important than the absolute width of the chink (Titze, 1984).

A10.7 RELATING THE VOCAL FOLD VIBRATORY MOTION TO THE GLOTTOGRAPHIC WAVEFORMS

Three glottographic waveforms, electroglottographic, photoglottographic, and inverse-filtered glottal waveforms, have been related to glottal characteristics; that is, lateral contact area (Childers, Smith, and Moore, 1984) and glottal volume velocity (Wong, Markel, and Gray, 1979), respectively. In order to establish the relation between the glottographic waveforms and the vibratory patterns of the vocal folds, we first express these glottal characteristics in terms of the glottal displacement. Formulas are then derived for the simulation of the glottographic waveforms. The comparisons between the synthetic glottographic waveforms and the measured waveforms allow an evaluation of the vibratory model and the transduction formula.

A10.7.1 Derivation of the Projected Glottal Area

The formulation of the projected glottal area follows the work by Titze (1984). The projected glottal area is the area outlined by the glottis when seen from above the glottis. The glottal area is believed to be the primary descriptor of the glottal excitation. The glottal area becomes zero when the glottis is completely closed along its length.

Consider the length L of the glottis to be divided into a series of differential lengths dy. A differential projected glottal area can then be written as

$$dA_g(t) = 2\xi_{min}(y, t)\,dy \quad \text{(A10.7.1.1)}$$

where ξ_{min} is the minimum positive value of ξ in a differential vertical duct (negative values are set to zero). ξ_m can, in principle, be found by differentiating $\xi(y, z, t)$ with respect to z and setting the result equal to zero. This produces a contour of values along the length of the vocal folds at the vertical positions z_{min}. The total projected glottal area is then

$$A_g(t) = \int_0^L 2\xi_{min}(y, t)\,dy \quad \text{(A10.7.1.2)}$$

In practice, this glottal area function is determined numerically. In discrete form, Equation (A10.7.1.2) becomes

$$A_g(n) = \frac{L}{M}\sum_{j=1}^{M} 2\xi_{min,j,n} \quad \text{(A10.7.1.3)}$$

for the nth time step and M finite ducts along the length. The minimum value in each of the vertical ducts is found by simple comparison of N discretized values of vertical displacements within the duct ($z = k\Delta z$, $k = 1, N$).

The locations of z_{min} may not be the same along the length. For example, if there are large phase differences in the tissue movement, the possibility that z_{min} is near the top anteriorly and near the bottom posterily (or vice versa) cannot be excluded. The minimum glottal area that is relevant for glottal airflow may not be the same as the projected glottal area. For a computation of the glottal volume velocity, it may be more appropriate to define the minimum glottal area as

$$A_g(n) = \min_k \left(\frac{L}{M} \sum_{j=1}^{M} 2\xi_{j,k,n} \right) \qquad \text{(A10.7.1.4)}$$

where $\min_k$ indicates that the minimum value of k glottal areas stacked vertically is selected as the minimum glottal area.

A10.7.2 Derivation of the Vocal Fold Contact Area

The vocal fold contact area is the lateral area of contact between the folds when they come together. To compute the total contact area, an infinitesimal amount of contact area dy dz is added whenever the glottal width goes to zero at any coordinate (y, z) on the glottal mid-plane. The total contact area is then the summation of the partial contact area along the length and the depth of the vocal fold as

$$A_c = \sum\sum c(y, z)\, dy\, dz \qquad \text{(A10.7.2.1)}$$

where $c(y, z) = 1$ for $\xi(y, z) \leq 0$ or $c(y, z) = 0$ for $\xi(y, z) > 0$.

In the above formulation, the vocal folds are allowed to overlap. We may consider the vocal fold tissue to be incompressible and every movement of the vocal folds results in their deformation in another direction. During vocal fold collision, the vocal folds press against each other and cause a change in the thickness of the vocal folds. For a better approximation to the collision process, the thickness at the collision surface should be dynamically adjusted according to the degree of overlap.

The approach for calculating the varying thickness during the collision is to use the incompressibility of the vocal tissue. The incompressibility property requires that the total volume of the vocal folds remains the same. This implies that

$$(x(t) - x_0)\, d_0 = (x_c(t) - x_0)\, d(t) \qquad \text{(A10.7.2.2)}$$

where $x(t)$ is the lateral displacement of a unit area, x_0 is the position of the boundary, d_0 is the nominal unit contact area if there is no overlap, $x_c(t)$ is the location where contact occurs, and $d(t)$ the adjusted contact area of the vocal folds. One immediate result of this approximation is that the electroglottographic (EGG) waveform in the vicinity of the EGG minimum is rounded rather than flat. This result is a consequence of the change in contact area after the first collision.

A10.7.3 The Simulation of EGG Waveform

Electroglottography (EGG) is described in Chapter 4. It measures the electrical impedance of the tissue in the vicinity of the larynx. It is generally accepted that the EGG reflects the changes in the lateral area of contact (Childers and Krishnamurthy, 1985; Childers et al., 1986; Childers et al., 1990; Kitzing, 1986). Phonation alters the impedance, most likely due to changes in the current paths within the larynx. These changes occur when vocal fold motion alters the glottal configuration. A closed glottis creates a relatively small impedance compared to an open glottis. However, the vocal fold contact area has never been measured *in vivo*. Titze (1989) used excised canine larynges to examine the relation between EGG signals and the corresponding dynamic vocal fold contact area. His results did not refute the hypothesis of a linear relation between contact area and the EGG signal. Thus, it is assumed that a change in electrical impedance measured by the EGG is inversely proportional to the changes in lateral contact area. The lateral contact area and EGG signal are then related as follows.

$$EGG(t) = \frac{k_1}{c(t) + k_2} \qquad \text{(A10.7.3.1)}$$

where $c(t)$ is the contact area, k_2 represents the shunt impedance of the adjacent tissues, and k_1 is a scaling constant.

A10.8 SELECTING VALUES FOR THE RIBBON MODEL PARAMETERS

The choice of model parameter values is based on empirical data, with each parameter setting being made separately. In fact, parameters such as vertical phase difference, vertical shape of the glottis, and fundamental frequency of vibration are correlated and should not be assigned arbitrarily. Knowledge of such correlations would help us exclude physically impossible combinations of parameters. In order to have a more systematic method for parameter settings, one could explore combinations and ranges of the parameter values using the data and model described earlier in this appendix. The approach is to systematically vary the control of the input, for example, lung pressure, vocal fold tension (or abduction of the glottis), and determine the corresponding changes in kinematics of the vocal folds. To obtain optimum model parameters by this method is difficult. The optimum model parameters can be extracted from the measured glottographic waveforms through a waveform-matching, model-fitting-data, optimization procedure. A simulated annealing algorithm is used to estimate the model parameters and the vocal fold configurations that minimize the error between the measured and the model-generated glottal area and EGG waveforms (Wu, 1996).

The ribbon model basically describes the kinematics of vocal fold vibration. For example, the lag between movements of the anterior and the posterior portions of the vocal folds is taken into account, and an adjustment phasing provides a method to compensate for the fact that the largest amplitude of lateral excursion does not occur at the midpoint of vocal fold length. A formulation to relate the lateral contact area to the EGG is also included in our model for the simulation of the EGG waveform.

A graphical user interface and animation are available in Chapter 9 for the two-mass and ribbon vocal fold models. The animation of vocal fold vibration can be viewed on the computer screen. These tools provide the user a view of the effects and interrelations between the model parameters and vocal fold vibrations.

A comparison between a measured and a model-derived glottal area waveform is shown in the Figure A10.31. The percentage error distance between model and data is 0.05. This example illustrates that the model does capture the vibratory patterns of the vocal folds quite well.

Additional data and models are available in Wu (1996). A movie of one vibratory cycle of the motion of actual vocal folds is also available on the CDs accompanying this text. See the README file on the CDs.

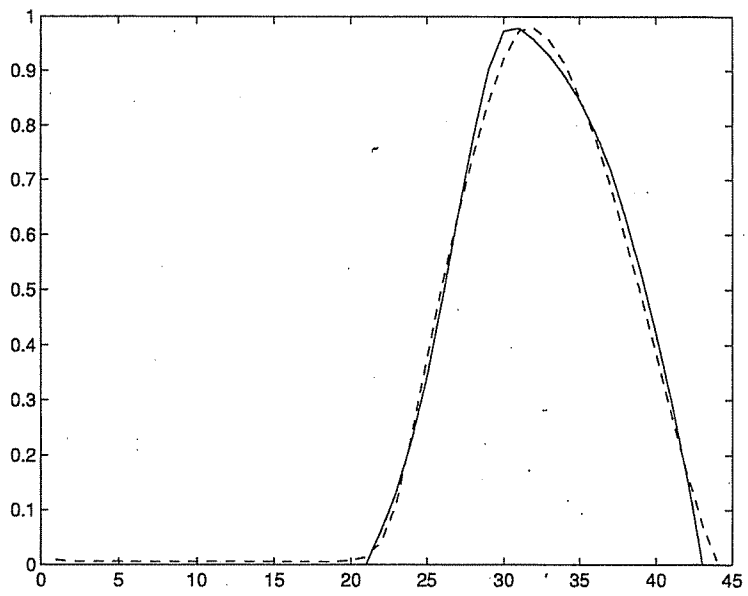


FIGURE A10.31 A comparison between a measured and a model-derived projected glottal for the ribbon model. Model is dashedline. Measured data is solidline.