Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 667–690

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

A review of lumped-element models of voiced speech

Byron D. Erath [a,*], Matías Zañartu [b], Kelley C. Stewart [c], Michael W. Plesniak [c], David E. Sommer [d], Sean D. Peterson [d]

[a] Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, NY 13699, United States
[b] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[c] Department of Mechanical and Aerospace Engineering, The George Washington University, Washington DC 20052, United States
[d] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Received 11 September 2012; received in revised form 6 February 2013; accepted 8 February 2013
Available online 21 February 2013

Abstract

Voiced speech is a highly complex process involving coupled interactions between the vocal fold structure, aerodynamics, and acoustic field. Reduced-order lumped-element models of the vocal fold structure, coupled with various aerodynamic and acoustic models, have proven useful in a wide array of speech investigations. These simplified models of speech, in which the vocal folds are approximated as arrays of lumped masses connected to one another via springs and dampers to simulate the viscoelastic tissue properties, have been used to study phenomena ranging from sustained vowels and pitch glides to polyps and vocal fold paralysis. Over the past several decades a variety of structural, aerodynamic, and acoustic models have been developed and deployed into the lumped-element modeling framework. This paper aims to provide an overview of advances in lumped-element models and their constituents, with particular emphasis on their physical foundations and limitations. Examples of the application of lumped-element models to speech studies will also be addressed, as well as an outlook on the direction and future of these models.

Keywords: Lumped-mass models; Glottal aerodynamics; Acoustics; Vocal fold models; Vocal tract; Subglottal system; Acoustic interaction

Contents

* Corresponding author. Tel.: +1 3152686584.
E-mail addresses: berath@clarkson.edu (B.D. Erath), matias.zanartu@usm.cl (M. Zañartu), kstewart@gwu.edu (K.C. Stewart), plesniak@gwu.edu (M.W. Plesniak), peterson@mme.uwaterloo.ca (S.D. Peterson).

0167-6393/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2013.02.002

1. Introduction

The widely accepted myoelastic-aerodynamic theory of vibration states that vocal fold (VF) vibrations are produced by a coupling between the aerodynamic forces and the tissue parameters (van den Berg, 1958), which generates a rich spectrum of acoustical sound that can feed back and influence both the VF dynamics and the fluid flow. Interest in the field of voiced speech production can be traced back nearly a century (Wegel, 1930) and arises primarily from a desire to understand the basic physics of phonation, as well as to provide diagnosis and treatment to individuals that suffer from any number of vocal pathologies. Due to the complexity of voiced speech, simplified models are often employed for both scientific and clinical investigations. As with all modeling practices, scientific speech investigations seek to reduce the complex problem down to the essential physical foundation to gain traction in understanding the phenomena. This is, of course, a fine balancing act between reducing the model enough to be tractable, but not so far as to be trivial.

The most common modeling framework in voiced speech investigations is the so-called lumped-element approach, wherein the VF structure is modeled as a collection of discrete coupled mass-spring-damper systems subjected to some external aerodynamic and/or acoustical loading function. Generally speaking, in a damped vibrating system such as the VFs, the energy imparted to the system must be as great as the energy dissipated by it in order to sustain self-oscillation. This occurs physiologically, as the aerodynamic loading is physically coupled to the VFs and supplies energy to the system during each oscillation cycle. Lumped-element models have proven capable of emulating the physiological VF kinematics and acoustic output. In the lumped-element framework, the dynamical response of the VFs is governed by a system of coupled nonhomogeneous autonomous ordinary differential equations. The number of equations needed to describe the system is directly related to the number of degrees of freedom employed in the model. The system of equations can usually be solved with relative ease using an explicit, forward time-marching scheme, with appropriate initial conditions.

Lumped-element VF model investigations were initially utilized because of the low computational cost that enables quick, efficient calculations. In spite of ever-increasing computer speeds, they still remain an attractive approach for scientific speech investigations due to their clinical significance and their ability to perform broad parametric investigations relatively quickly (Erath et al., 2011b; Steinecke and Herzel, 1995). In addition, as lumped-element models become more refined, the prospect of developing short run-time models that can mimic patient-specific phonatory conditions becomes more realistic.

The objective of this paper is to review advancements in lumped-element models of speech, focusing specifically on

the push towards physics-based models. This article aims to be a comprehensive reference for, and review of, the state-of-the-art in lumped-element modeling for researchers seeking to employ these research. While a broad overview is provided, more recent findings are emphasized.

The paper begins with a brief overview of the anatomy and physiology of the larynx in Section 2. This is followed by a discussion of the strategies that have been implemented for modeling the geometric structure (Section 3), tissue properties (Section 4), contact forces (Section 5), fluid loading (Section 6), and acoustics (Section 7) that occur during voiced speech. The utility and limitations of these strategies are given due discourse in each section. Examples of the application of lumped-element models to scientific speech investigations of both normal and pathological speech are provided in Section 8. The article closes with outlooks for the field of lumped-element speech modeling in Section 9.

2. Anatomy and physiology of the vocal folds

The larynx is situated in the anterior portion of the neck at the junction of the trachea and the hypopharynx, see Fig. 1 for a schematic of anatomical reference planes and orientations. It serves to protect the airway from pulmonary aspiration, regulate lung pressure, and generate sound. Colloquially referred to as the "voice box", the larynx houses the VFs, which as a unit, comprise the primary acoustic source in voiced speech, as well as the more membranous, less active ventricular folds. In this section we briefly introduce the anatomical structure and physiological function of the larynx as it pertains to phonation and the conception of the models of interest. For a more comprehensive discussion of laryngeal anatomy and physiology, see for example Titze (1994b).

2.1. Cartilages and intrinsic muscles of the larynx

The larynx consists of nine principle cartilages; three individual cartilages (epiglottis, thyroid, and cricoid), and three paired cartilages (arytenoid, cuneiform, and corniculate). The geometry, location, and function of the primary cartilages are shown schematically in Fig. 2 and tabulated in Table 1.

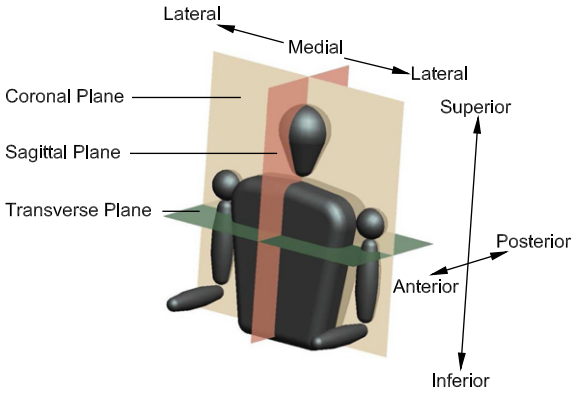

Fig. 1. Description of physiological data planes and orientations. Adapted from Mittal et al. (2013).

The intrinsic muscles of the larynx are those that have both origin, defined as the fixed point of attachment where the muscle begins, and insertion, defined as the point where the muscle attaches to a moving joint, within the larynx. The muscle locations are shown schematically in Fig. 2. These muscles are essential for both VF posturing and control of VF vibration. The thyroarytenoid muscles constitute the bulk of the VF structure and primarily serve to control tension, elasticity, and, to some degree, VF adduction. The cricothyroid muscles control pitch by both lengthening and tensioning the VFs, while adduction and abduction of the VFs is controlled by the arytenoid muscles. A brief summary of the muscles and their function and innervation is provided in Table 2.

2.2. The vocal fold structure

Each VF is a multi-layered structure consisting of the thyroarytenoid muscle as the deepest layer, the intermediate lamina propria and the epithelium as the superficial layer, as shown in Fig. 2(d). The lamina propria is further subdivided into deep, intermediate, and superficial layers based on the density of the constituent elastic and collagenous fibers (Hirano, 1975). A coarser, yet arguably more functional, division of the VFs separates them into "body" and "cover" layers, where the body comprises the thyroarytenoid muscle and the deep and intermediate layers of the lamina propria, and the "cover" consists of the superficial layer of the lamina propria and the epithelium. While the body is a muscular structure (formally), the stiffness of which can be modulated via muscle activation, the cover is pliable and viscoelastic, sitting loosely on the VF body much like the scalp sits on the head (Baken, 1997). In preparation for speech, the VFs adduct and seal the trachea. When a critical lung pressure is achieved, the VFs are pushed apart by aerodynamic forces and begin self-sustained oscillations, which are characterized by the propagation of a mucosal wave in the cover. This wave travels along the medial surface of each VF. During each phonatory cycle, propagation of the mucosal wave causes the space between the VFs, referred to as the glottis, to transition from a convergent to uniform, and finally to a divergent passage before closing and repeating the cycle, see Fig. 3.

2.3. Phonatory control

Phonatory output is regulated by coordinated control of VF tension and elasticity and respiratory muscles; for a comprehensive discussion, see for example Jiang et al. (2000). The fundamental physical properties that determine VF motion are tissue mass, stiffness, and viscosity. The vibrating mass of the VFs can be adjusted through coordinated contraction of the intrinsic and extrinsic muscles of the larynx, thereby altering the fundamental frequency of the structure. VF tension may be altered by passive stretching of the VFs due to contraction of the cricothyroid, as

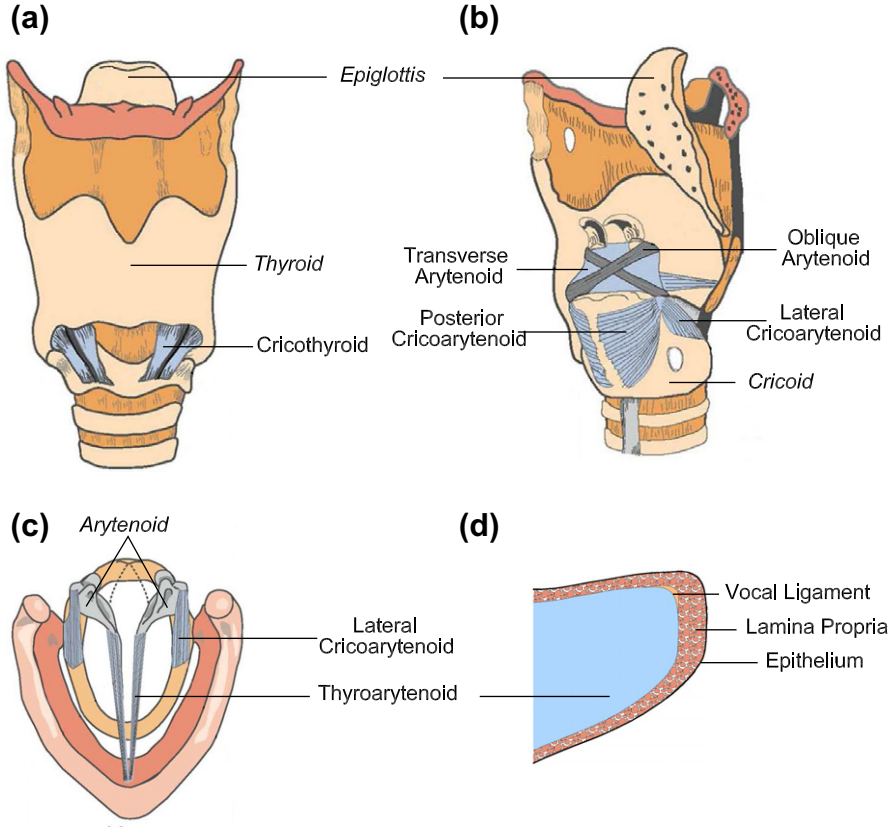

Fig. 2. Detailed view of the muscles (normal font) and cartilages (italicized font) of the larynx shown from (a) anterior, (b) oblique, (c) and superior views. (d) A midcoronal section of the VF structure is also shown. Adapted from Mittal et al. (2013).

well as by activation of the thyroarytenoid muscle. The latter allows changes in the internal stiffness without any alteration of VF length (Hirano, 1974; Titze, 1979). While these muscles primarily influence the body of the VF structure, the stiffness of the cover is also affected due to longitudinal strains imposed by surrounding structures. Alteration of the stiffness in both the body and the cover may also change the fundamental frequency of oscillation. The VF tissue viscosity determines the ease with which the mucosal wave propagates through the VF cover; high tissue viscosity damps oscillations and impairs phonation (Jiang et al., 2000).

3. Components of lumped-element vocal fold models: Functionality and geometry

A significant challenge in modeling voiced speech is to sufficiently simplify the VF structures so that the dynamics can be reasonably assessed while maintaining sufficient geometric and functional fidelity; that is, to borrow a famous aphorism, the VF structure must be made as simple as possible, but no simpler. Consequently, an array of VF representations has been developed with the number of degrees of freedom ranging from one to millions (Titze, 2006). The most ubiquitous VF models are lumped-element models, in which the VFs are represented by a small number of discrete, rigid masses connected via a series of springs and dampers that account for the viscoelastic nature of the VF structure. A comprehensive survey of the lumped-element configurations that have been utilized for scientific speech investigations is presented in Birkholz (2011). Hence, the objective of this section is not to introduce the gamut of existing models, but rather to discuss the progression of model development from a functional perspective, highlighting the advantages and shortcomings of various implementations for voiced speech studies.

3.1. Single layer block models

As introduced in Section 2.2 and discussed in detail in Section 4, the VFs can be functionally split into "body" and "cover" layers. Lumped-element models can be similarly classified by the number of layers employed to model the VFs. The preponderance of models conform to either a single or dual layer structure. The general layered VF structure may consist of an array of m_i masses in the anterior–posterior direction and n_i masses in the inferior–superior direction, while the lateral VF layers are denoted by the subscript $i = 1$ or 2, with $i = 1$ corresponding to the "cover" layer, and $i = 2$ the "body" layer. Single layer models condense the layered VF structure into a single layer of masses $i = 1$; that is, the "body" and "cover" are not separated. This section focuses on this class of models, see Fig. 4(a).

Table 1
Cartilages of the larynx.

Name	Function
Epiglottis	Seals the airway to direct food to the esophagus
Thyroid	Largest cartilage; forms the main structure and protective framework of the larynx
Cricoid	Inferior to the thyroid; together the two create a point of rotation to allow VF lengthening and tensioning by the attached muscles
Arytenoid (paired)	Enable VF positioning through adduction and abduction as well as rotation
Corniculate (paired)	Small cartilage located at the superior tip of the arytenoid
Cuneiform (paired)	Small cartilage located anterior to the corniculate

One mass models are the simplest form of single layered models, with the entire vocal fold structure condensed into a single mass ($m_1 = n_1 = 1$). Some of the earliest lumped-element VF investigations used one mass with a single degree of freedom (Flanagan and Landgraf, 1968; Wegel, 1930). Generally speaking, for a viscoelastic VF model to exhibit self-sustained oscillations, energy must be continuously (or periodically) added to the system to overcome internal friction in the tissue. Energy is added to the system by an external load when the load is in phase with the velocity of the VFs. Self-sustained oscillation will only occur in the presence of acoustical loading or some other imposed loading mechanisms (McGowan and Howe, 2010). Energy imparted to the VFs can be provided by the glottal air flow, acoustic pressures, or a combination thereof (Titze, 1988). The modification of the aerodynamic pressures due to the changing convergent-to-divergent glottal orientation is the primary mechanism by which this occurs physiologically. For multi-mass lumped-element models the time-varying glottal geometry can yield an asymmetry in the pressure loading and thus enable self-sustained oscillations (Titze, 1988). Single mass, single degree of freedom systems by definition cannot develop the time-varying glottal orientation necessary for asymmetric pressure loading. To circumvent this constraint, Flanagan and Landgraf (1968) impose an acoustic loading on a single mass model via inertive and resistive loads in the supraglottal tract, thereby introducing a lag between the glottal velocity and area. In so doing, less energy is imparted to the system during closing than during opening. Alternatively, a capacitive load can be prescribed in the subglottal tract in conjunction with a resistive load in the upstream tract. This causes the subglottal pressure to deplete more slowly as the VFs open and to remain lower as the VFs close, thus producing the requisite asymmetry. However, there must be a sufficiently high flow rate to ensure the capacitive load is recharged each cycle (Gupta et al., 1973).

If the single-mass model is instead allowed two degrees of freedom (one translational and one rotational), then a time-varying glottal geometry is possible, which can produce self-sustained oscillations (Horáček et al., 2005; Liljencrants, 1991). Another approach to obtain the necessary energy exchange is to employ time-varying orifice discharge coefficients for the glottis. This technique assumes the flow separation point within the glottis is fixed, however the losses are assumed to vary throughout the cycle (Zañartu et al., 2007). More recently, potential flow theory has been applied to a single-mass system, demonstrating that a flow separation point that shifts from the upstream edge of the mass during opening to the downstream edge during closure can induce self-sustained oscillations (McGowan and Howe, 2010). Non-traditional forcing can also incite self-sustained oscillations in a single-mass model through the mechanism of "negative Coulomb damping" (Fulcher et al., 2006). Finally, a "pseudo" one mass model has also been proposed, whereby the phasing difference between the inferior and superior edges is simulated by assuming that for large amplitude oscillations one-to-one modal entrainment occurs, allowing the displacement of the upper fold to be described in terms of the lower mass by prescribing a time delay and scaling factor. In this manner, only the dynamics of the lower mass are tracked, thereby decreasing computation time (Avanzini, 2008).

Ishizaka and Flanagan (1972) added a second coupled superior mass to each VF in the traditional one mass model while restricting the masses to move in pure translation, resulting in the popular two mass, two degree of freedom model ($m_1 = 1, n_1 = 2$). The second mass enables the phasing difference observed between the inferior and superior edges of the VFs and thus captures the mucosal wave, however in this model the fluid loading is generally applied only to the larger inferior mass. Thus the second mass exists solely to capture the structural dynamics and is not influenced directly by the fluid loading. Since its introduction, the two-mass configuration has been one of the most oft-employed models for the VF structure (Bailly et al., 2008; Erath et al., 2011a,b; Lous et al., 1998; Lucero, 2005; Mergell and Herzel, 1997; Jiang et al., 2001; Steinecke and Herzel, 1995). Tokuda et al. (2007, 2008) append a third superior mass to each VF ($m_1 = 1, n_1 = 3$) in order to simulate register transition between chest and falsetto. The addition of the third (superior) mass allows the model to more accurately replicate the mucosal wave during falsetto with the middle and superior masses vibrating out of phase.

A 10 mass three-dimensional block model ($m_1 = 5, n_1 = 2$) is introduced by Wong et al. (1991), which enables longitudinal tensioning of the VFs by coupling the anterior–posterior masses (Titze, 1973). The masses are constrained to move only in the medial-lateral direction, however. More recently, a similar approach has been utilized to reproduce the three-dimensional dynamics of the VFs during *in vivo* oscillations using a 6 mass model

Table 2
Intrinsic muscles of the larynx. For innervation: RLN = recurrent laryngeal nerve; SLN = superior laryngeal nerve.

Name	Function	Innervation
Thyroarytenoid	The "body" of the true VF; subdivided into the thyrovocalis and thyromusclaris, which aid in VF tension and adduction, respectively	RLN
Cricothyroid	Further subdivided into the pars recta and pars oblique, contraction can both lengthen and tension the VF	SLN
Interarytenoid	Consists of both a transverse and oblique component, which adduct the arytenoid cartilages	RLN
Lateral cricoarytenoid	Adduct and rotate the arytenoid cartilages, bringing the VFs together	RLN
Posterior cricoarytenoid	Abduct the arytnoid cartilages, causing the VFs to open	RLN

($m_1 = 2, n_1 = 3$) (Schwarz et al., 2008; Wurzbacher et al., 2008). Similar to the 10 mass model of Wong et al. (1991), this 6 mass model enables varying longitudinal stiffnesses, as well as the ability to vary anchor spring stiffnesses in the longitudinal direction. Furthermore, it incorporates an added degree of freedom, allowing the masses to move in both the lateral and longitudinal directions. Yang et al. (2010) introduced a 25 mass model ($m_1 = 5, n_1 = 5$) to provide higher resolution of the VF geometry and enable parameter extraction based on high-speed video endoscopy and excised human larynx motion tracking of the medial VF surface (Yang et al., 2010, 2011, 2012). Similarly, the original model from Ishizaka and Flanagan (1972) has been segmented in the anterior–posterior direction using a variable number of masses to model vocal posturing and some pathologies (Kob et al., 2002; Fraile et al., 2012). An alternative strategy for incorporating anterior–posterior variations in the glottal area while minimizing the degrees of freedom is introduced by Birkholz (2011). In this model, two masses are used to represent each VF ($m_1 = 1, n_1 = 2$), however these masses are angled with respect to a sagittal plane through the glottal midline. That is, the model allows for varying levels of VF adduction by allowing for different spacing between the opposing masses at the anterior and posterior ends.

3.2. Dual layer models

The two functional layers of the VFs are structurally different, a feature that single layer lumped-element models do not incorporate. The differentiated structural layering of the VFs is incorporated into a second class of "body-cover", or dual layered, models, see Fig. 4(b). Titze (1973, 1974) proposed a 16 mass model with 8 masses in both the body and cover layers ($m_1 = 8, n_1 = 1; m_2 = 8, n_2 = 1$). This approach is attractive as it provides appropriate boundary constraints and allows for modeling coarse spatial variations in tissue properties. The initial version described the geometry in the streamwise flow direction with only one mass, but accounted for the string-like modes of the ligaments and muscle fibers with more masses. At the time the model was proposed, there was confusion about whether the first layer should represent the ligament or the mucosa, as the model came before the fundamental work of Hirano on the topic (Hirano, 1974). For this reason, the original version of the model (Titze, 1973, 1974) was abandoned for years. More recently, Titze (2006) updated the model's layered structure, leading to a 175 point-mass model. The model now features 7 layers to describe the vibrating tissue (mucosa, ligament, and thyroarytenoid muscle), 5 divisions in the inferior–superior direction, and 5 divisions in the anterior–posterior direction (i.e., $m_i = 5$ and $n_i = 5$ for $i = 1, \ldots, 7$). This current version is widely accepted and has been used to model various voice qualities, pitch glides, singing, voice instabilities, etc. (Titze, 2006, 2008; Titze and Worley, 2009).

A model closer in spirit to the two-mass model of Ishizaka and Flanagan (1972) couples a two-mass representation of the cover with a single body mass ($m_1 = 1, n_1 = 2, m_2 = n_2 = 1$) (Story and Titze, 1995). This model configuration enables a more natural connection between contraction of the cricothyroid and thyroarytenoid muscles to the model (cover and body) stiffness parameters, see Section 4. From a model comparison standpoint, this framework is attractive because the standard two-mass model ($m_1 = 1, n_1 = 2$) can be recovered by making the mass and stiffness of the base approach infinity (Story and Titze, 1995; Titze, 2002). Tokuda et al. (2010) have also introduced a dual-layer version of their three-mass model, in which the three cover masses are coupled with a single body mass ($m_1 = 1, n_1 = 3, m_2 = n_2 = 1$).

3.3. Aerodynamically smooth models

From a VF dynamics perspective, models in which the masses are represented as blocks are fully capable of capturing the structural vibration modes within the limits of the degrees of freedom on the model. However, from an aerodynamic loading perspective, there are some subtle, but significant, concerns that arise from the discontinuous area change at the block junctions. Specifically, severely asymmetric tissue properties, as occur in some pathologies, can lead to vibration regimes where an inferior mass of one VF may momentarily overlap with a neighboring superior mass of the opposing VF, thus "closing" the glottis. This leads to a large spike in the pressure loading during a nominally open portion of the cycle, which can have a significant impact on the VF dynamics, as discussed by Sommer et al. (2013).

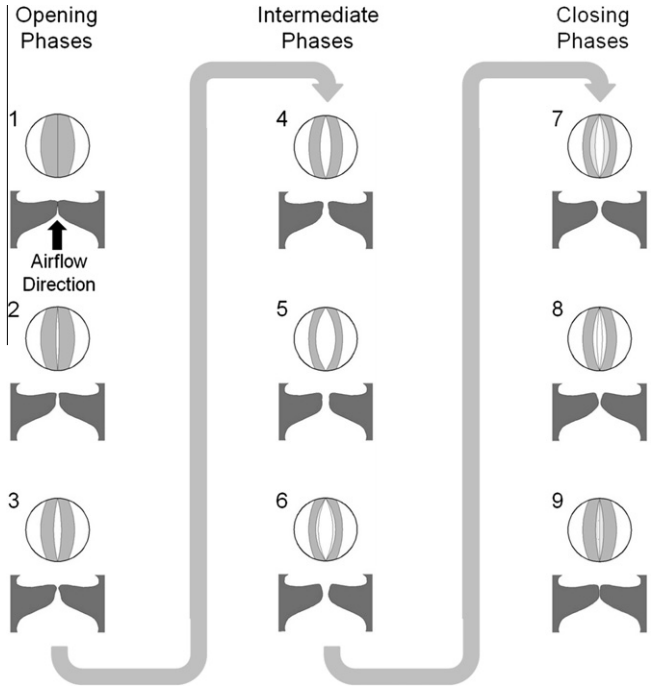

Fig. 3. Nominal configuration of the glottis through one vibration cycle representative of normal phonation at 9 instances during the cycle. In each panel, a superior view is shown above a midcoronal section. During the opening phases the glottis is pushed apart and assumes a convergent configuration. It transitions from a convergent to divergent orientation during the intermediate phases, and then during the closing phases the medial VF surfaces come together and close.

A more physically satisfying VF model construction from an aerodynamic standpoint has a smooth medial surface contour, irrespective of the number of discrete masses employed for the structural dynamics. Indeed, a number of aerodynamically smooth VF models have been introduced. A simple modification of the two-mass model ($m_1 = 1$, $n_1 = 2$), whereby a plane representing the medial surface of each VF is defined by connecting the two point masses is explored in Erath et al. (2011a,b), Lous et al. (1998), and Ruty et al. (1997), see Fig. 4(c). In this configuration, the medial surface angle can change throughout the phonatory cycle, however closure of the glottis can only occur if the leading or trailing edges of the VFs are in contact. A more sophisticated approach is obtained by incorporating smoothly-contoured radii at the VF inlet and exit, and providing a smooth boundary that adjoins both masses (Dresel et al., 2006; Pelorson et al., 1994).

While identical to the block-type two-mass models ($m_1 = 1, n_1 = 2$) in terms of the tissue dynamics, smooth contour models are fundamentally more appealing in terms of the fluid mechanics, and experience subtle yet important differences with regards to the fluid loading. For example, aerodynamically smooth models enable temporally-varying flow separation points, as will be discussed in Section 6. The inclusion of a variable separation point along the smoothly-contoured geometry causes, for example, a significant decrease in the fundamental frequency (Pelorson et al., 1994). Alternatively, the upper VF layer can be represented as a single mass with two degrees of freedom; that is, a plate that translates and rotates (Liljencrants, 1991; Titze, 2002; Titze and Story, 2002). This plate is connected to the body via both linear and torsional springs and dampers as shown in Fig. 4(d) ($m_1 = n_1 = m_2 = n_2 = 1$). From a structural dynamics perspective, it can be shown that this orientation is identical to a $m_1 = 1, n_1 = 2, m_2 = n_2 = 1$ block representation (Titze, 2002). The aerodynamic loading is different, however, due to the smooth nature of the medial surface in the bar-plate model.

4. Components of lumped-element vocal fold models: Structural characteristics

As discussed in Section 2, the human VFs are viscoelastic structures whose stiffness and damping properties determine the vibration patterns of voiced speech (Chan and Titze, 1999, 2000). In lumped-element models, the tissue viscoelasticity is represented as springs and dampers in parallel configurations. While *ad hoc* tissue properties can enable lumped-element models to accurately replicate VF kinematics during sustained vowel phonation, relating physiological tissue properties to the lumped-element system parameters remains considerably challenging. This is particularly difficult when attempting to incorporate the intricate articulatory control that occurs during running speech, where muscle activation states and vocal posturing are important. Ultimately, direct correlation between actual tissue properties and their lumped-element representations is necessary in order to use reduced-order models to relate physiological control with vocal function. Consequently, a number of strategies have been developed for both selecting and optimizing lumped-element model parameters.

4.1. Direct measurement and parameter tuning

Ishizaka and Flanagan (1972) sought lumped-element parameters based upon physiologically measured values for their two-mass model. They assigned lateral spring values from static tensile stress measurements of the thyroarytenoid muscle from excised human larynges (Ishizaka and Kaneko, 1968). This neglects the varying tissue properties between the body and cover, as well as any modification during muscle activation. Damping parameters were extracted from the damped free vibration of human VF tissue, while the masses were chosen such that their sum represented the total measured mass. The distribution of the mass between the two blocks correlates with the physiological variation in the thickness of the VFs (Ishizaka and Flanagan, 1972). Similar approaches have been used in subsequent model developments as well (Lous et al., 1998; Story and Titze, 1995), where system parameters are chosen based on physiological measures of VF length (Hirano et al., 1981), body/cover thickness (Baer, 1981; Hirano, 1977; Hirano et al., 1981), tissue density (Perlman

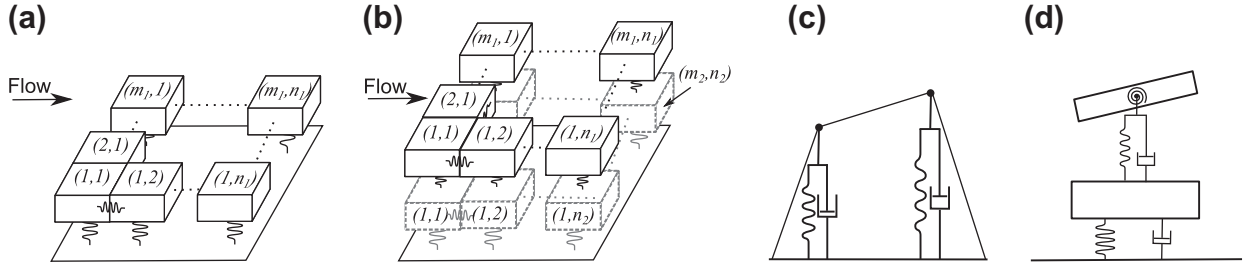

Fig. 4. (a) Single layer model configuration, (b) dual layer model configuration, (c) single layer aerodynamically smooth model, (d) dual layer "bar-plate" model.

et al., 1985), stiffness (Alipour-Haghighi and Titze, 1991), and damping of the thyroarytenoid muscle (Isshiki, 1977; Kaneko et al., 1972). As a result of the paucity of available experimental data and the limitations inherent to the measured data, lumped-element models employing directly measured physiological parameter values are often "tuned" as well; that is, the parameters are adjusted to elicit the desired response. (Avanzini, 2008; Cisonni et al., 2011; Dresel et al., 2006; Lous et al., 1998; Ruty et al., 1997; Story and Titze, 1995). Parametric variations of the lumped-element parameters have also been performed to correlate them to the system behavior (Ishizaka and Flanagan, 1972).

4.2. Correlation to muscle activation

The cricothyroid and thyroarytenoid muscles are the primary regulators of fundamental frequency (Gay et al., 1972; Hirano et al., 1970; van den Berg and Tan, 1959). Fundamental frequency control is important for expressing linguistical qualities such as syllable and word emphasis, as well as speaker identification (Brown et al., 1974). Recent investigations have sought to provide "rules" for controlling system parameters based on muscle activations in order to provide real-time variations in the VF properties. The pioneering work in this field by Titze and Story (2002) resulted in a set of rules for controlling lumped-element VF parameters based on cricothyroid, thyroarytenoid, and lateral cricoarytenoid muscle activity. Muscle activation controls VF elongation, nodal location of mucosal upheaval, which determines voice register, thickness and depth of the body and cover, VF adduction and abduction, and glottal orientation in both a "body-cover" ($m_1 = 1, n_1 = 2, m_2 = n_2 = 1$) and "bar-plate" ($m_1 = n_1 = m_2 = n_2 = 1$) lumped-element VF model (Titze and Story, 2002). Lowell and Story (2006) further study cricothyroid and thyroarytenoid activation associated with fundamental frequency determination using the same set of "rules". They show that muscle activation affects not only tissue properties but also the aerodynamic quantities of glottal flow rate and maximum flow declination rate.

This approach has also been used to explore mappings between muscle activation and features of the voice source signal. Avanzini et al. (2006) introduce a "codebook" that fits the output of a two-mass model ($m_1 = 1, n_1 = 2$) to voice waveforms, which is related to speech synthesis, a topic further discussed in Section 8.3. Although fitting capabilities were only qualitative in nature, controlling muscle activation is significantly more stable than prior attempts at fitting physiological waveforms using regressor-based mappings to control glottal pressure (Avanzini et al., 2001; Drioli and Avanzini, 2002). In addition, this approach shows promising scalability to lower dimensional (e.g., one mass) VF models.

4.3. Inverse analysis methods

In the past decade, several researchers have utilized inverse analysis methods to determine model system parameters based on clinical high-speed video recordings of normal (Döllinger et al., 2002; Qin et al., 2009; Rupitsch et al., 2011; Schwarz et al., 2008; Tao and Jiang, 2007a; Wurzbacher et al., 2008; Yang et al., 2011, 2012) and pathological voice (Schwarz et al., 2006). Typically, VF motion in the transverse plane is first extracted from high-speed video using image processing algorithms (Lohscheller et al., 2007, 2009; Qin et al., 2009; Schwarz et al., 2008; Voigt et al., 2010). An objective function is then created by comparing the clinical trajectories with the theoretical, lumped-element model trajectories. A parameter set that allows variation in the system parameters (e.g., subglottal pressure, block mass, spring stiffness, etc.) is then optimized based on minimizing the objective function. In order for the objective function to fit the criteria of a "well-posed" problem, each solution must: (1) exist; (2) be unique; and (3) be stable when subject to small perturbations in the data or problem parameters. Generating a well-posed problem for inverse analysis of tissue properties and glottal kinematics using a lumped-element VF model poses a significant challenge. Overly simple models are unable to recreate the observed experimental data, thereby violating the first criterion. Conversely, overly-elaborate models may have multiple solutions that accurately reproduce the experimental data within experimental uncertainty, or a very large set of parameters that very closely explain the observed data, thereby violating the second criterion. This is the challenge observed in the previously reported methods, whereby the least-squares approach

produces a non-convex objective function. This is exacerbated as models include additional degrees of freedom. To address this problem, the objective function can be minimized by using genetic algorithms (Goldberg, 1989), although care must be taken in interpreting the results because uniqueness is still not guaranteed.

Initial inverse analysis investigations only tracked trajectories at either one point along the medial surface (Döllinger et al., 2002), or the glottal area as a whole (Tao and Jiang, 2007a). Subsequent models have utilized higher-order geometrical descriptions of the VF geometry, tracking three points along each VF, as well as anchor points at the anterior and posterior commissures (Schwarz et al., 2008; Wurzbacher et al., 2008). More recently, Yang et al. (2011, 2012) have employed a higher spatial resolution three-dimensional cover model ($m_1 = n_1 = 5$). They generate an objective function by comparing the motion of the lumped-element model blocks with corresponding surface points on a hemi-laryngeal excised larynx. By extracting system parameters in both the inferior–superior and anterior–posterior directions, they show that the lateral stiffness is lower than the longitudinal and vertical stiffnesses, while the local "effective" mass and stiffness decrease in the vertical direction.

4.4. Reduction of finite-element to lumped-element vocal fold models

Reduction of order from finite-element models to lumped-element models is another potential avenue for determining lumped-element VF system parameters (Cook and Zañartu, 2010; Chen et al., 2008; de Vries et al., 1999; Qiu et al., 2002). The first work to propose this idea (de Vries et al., 1999) begins by specifying two locations in a mid coronal plane on the finite-element model at which the kinematics will be matched. The displacement at these locations is then determined based on a static pressure loading, representative of voiced speech. Given the same aerodynamic loading, the spring stiffnesses are then calculated by the known displacement, after which the mass can be computed such that the normal modes of both systems match (de Vries et al., 1999). A shortcoming of this approach is that the location at which displacements are measured are chosen *ad hoc*, whereas displacements vary based on location. Furthermore, while the finite-element model used a body-cover (dual-layer) representation, the lumped-element models did not incorporate this distinction, using instead a two-mass ($m_1 = 1, n_1 = 2$) arrangement. Consequently, the computed masses and stiffnesses are significantly lower than in previously reported lumped-element models (Ishizaka and Flanagan, 1972; Pelorson et al., 1994), since the extraction points are located in the superior surface of the model.

Subsequent efforts have pursued the inverse approach, whereby the mass is first determined by integrating the density over the volume of the model, with weighting factors added to account for the displacement of the finite-element model. Instead of describing displacements at discrete points, the displacements along the entire surface are extracted for a specific modal response of the system. With the lumped mass known, the spring stiffness is chosen such that the natural frequency of the lumped-element model matches that of the finite-element model. Initially proposed for only a one-mass ($m_1 = n_1 = 1$) model (Chen et al., 2008), this method has been extended to two-mass ($m_1 = 1, n_1 = 2$) VF models (Cook and Zañartu, 2010). For two-mass orientations, the lumped-element masses are determined by computing the kinetic energy and average velocities from the finite-element model for the translational and rotational kinetic modes and then computing the corresponding masses based on these properties. By computing modal frequencies next, the appropriate stiffnesses is determined (Cook and Zañartu, 2010). This approach, however, is limited since the displacements are computed based on a modal, rather than the dynamical response of the structure.

5. Components of lumped-element vocal fold models: Contact forces

In several phonatory modalities, including chest register, the dynamics are such that the two opposing VFs collide during the course of self-sustained oscillations. Mechanically, as the VFs collide, the tissues of the folds deform, inducing a stress field in the tissue. Deformation results in elastic potential energy storage in the tissue, which contributes to forcing the VFs apart for the next oscillation cycle. Consequently, collision is important for good speech quality, evidenced by its impact on acoustic metrics such as the maximum flow declination rate (Holmberg et al., 1988). However, excessive contact stresses, both normal and shear, are hypothesized to be an impetus for structural VF pathologies, see for example Jiang and Titze (1994), Titze (1994a), and Gunter (2003).

5.1. Contact stress magnitudes

VF closure initiates at the anterior and posterior ends of the folds and progresses towards the midline, which is generally the location of largest contact stresses. *In vitro* experiments with a canine hemi-larynx by Jiang and Titze (1994) indicate that peak contact pressure along the midline during collision is in the range of approximately 3 kPa. Intra-glottal measurements in human subjects using cannula-mounted sensors have reported contact pressures ranging from 1 to 4 kPa (Hess et al., 1998; Yamana and Kitajima, 2000), though the force sensors are typically relatively large, which impacts phonation and consequently the collision stresses. Numerical simulations by Tao and Jiang (2007b) using a finite element VF representation predict peak impact pressures near 4 kPa, which agrees well with the experimental studies. In general, the collision pressure increases with increasing subglottal pressure and with

decreasing glottal gap. Thus, breathy voice has lower impact stresses than does pressed voice, for example (Jiang and Titze, 1994).

5.2. Lumped-element model implementation

In the context of lumped-element models, the elastic energy stored in the tissues during collision is typically modeled as additional elastic members that are only activated when the folds are in contact. As discussed in Section 4, the lumped masses are assumed perfectly rigid, so in order to model collision, the masses are allowed to overlap. During the overlap, an additional restoring force is incorporated into the governing dynamical equations. A few different collision models have been proposed, the most common of which, typically employed in block-type models, incorporates an additional set of springs to model the collision elasticity. Both linear (Steinecke and Herzel, 1995) and nonlinear springs (Ishizaka and Flanagan, 1972; Story and Titze, 1995) have been used. For symmetric tissue properties and mass deflections, the "deformation" of each VF is equal to one-half of the total overlap distance, or equivalently, the distance of the medial block surface from the glottal centerline. In the case of asymmetric tissue properties, ensuring equal contact forces on the two VFs requires that the overlap distance for each fold be scaled by the inverse of the asymmetry parameter (Sommer et al., 2012). That is, the stiffer VF should deform by a relatively smaller amount than the softer fold. This necessitates a slight modification to the governing equations presented by Steinecke and Herzel for investigating asymmetric tissue parameters (Steinecke and Herzel, 1995).

Horáček et al. (2005) introduce a Hertz contact model for VF collision. They employ a single-mass two degree of freedom model with a curved, converging geometry. The curved mass shape is allowed to translate and rotate, while the glottal midline is treated as a symmetry plane. They compute the local radius of curvature of their model, then employ the classical Hertz contact force model of a spherical shape impacting a rigid wall. While Horáček et al. (2005) present a viscoelastic collision model, they ultimately neglect the damping term. Zañartu et al. (2007) apply the Hertz contact model with the viscous damping term to a single mass one degree of freedom curved model.

While for traditional block models the medial planes of opposing blocks always remain parallel, the model of Birkholz et al. (2011c), which has an anterior–posterior variation, requires special consideration. In this case, the overlap during collision varies in the anterior–posterior direction, which necessitates a contact area weighting be included in the governing equations. Similar contact area weighting is required in bar-plate type models. In these models, the overlap of the two folds depends not only upon the location of the cover plate, but also upon its angle. As discussed by Titze (2002), the bar-plate ($m_1 = n_1 = m_2 = n_2 = 1$) model does not incorporate collision springs, but rather applies an average of the subglottal and supraglottal pressures, which he refers to as a hydrostatic pressure, to the overlapped region. Thus, the force is not dependent upon the degree of overlap. It may be argued then, that the bar-plate model does not include contact forces.

6. Components of lumped-element vocal fold models: Fluid mechanics

6.1. Bernoulli flow solvers

During voiced speech the aerodynamic and acoustic pressures drive VF motion. The flow field is highly complex, comprised of many viscous flow behaviors, including unsteady flow separation, vortex shedding, transition to turbulence, and asymmetric flow instabilities (Mittal et al., 2013). Capturing the salient flow characteristics in a low computational cost modeling framework is a significant challenge. The earliest lumped-element flow solvers modeled the flow field as one-dimensional, symmetric, quasi-steady, incompressible, and inviscid, allowing simple solution via the steady Bernoulli equation (Flanagan and Landgraf, 1968). The seminal paper by Ishizaka and Flanagan (1972) utilizes the Bernoulli equation as a flow solver, but with added correction terms to include both viscous losses and pressure recovery in the supraglottal tract. The viscous losses are included by superimposing a pressure loss from a Poiseuille flow approximation through the glottal channel.

The majority of lumped-element flow solvers still invoke the quasi-steady assumption (Erath et al., 2011a; Ishizaka and Flanagan, 1972; Liljencrants, 1991; Pelorson et al., 1994; Vilain et al., 2004), whereby it is assumed that the Strouhal number, which indicates the relative importance of unsteady inertial terms to viscous terms in the governing Navier-Stokes equations, is sufficiently small that the time-dependent terms are negligible. In voiced speech, the Strouhal number is $O(10^{-4} - 10^{-3})$. While this range of Strouhal numbers suggests that a quasi-steady assumption is valid, at glottal opening and closure the flow experiences very large accelerations that may violate the quasi-steady assumption (Park and Mongeau, 2007). Furthermore, flow separation investigations have demonstrated that it is a Reynolds number-Strouhal number dependence that is critical for predicting quasi-steady behavior, as opposed to a Strouhal number dependence alone (Sobey, 1983).

6.2. Corrections for flow separation

Early flow solvers prescribed flow separation to occur at the minimal glottal area throughout the cycle, impeding any pressure recovery when the glottis forms a divergent channel. By definition, this assumption precludes the possibility of any variation in the flow separation location. Liljencrants (1991) proposes an *ad hoc* correction, suggesting that the point of flow separation for a divergent passage should occur at a constant ratio of the glottal area at flow

separation to the minimal glottal area equal to 1.3. A physics-based approach for determining flow separation based upon solving the von Kármán equation (Schlichting, 1968) using the Polhausen cubic method to find the separation location is also commonly used (Hirschberg et al., 1996; Hofmans et al., 2003; Pelorson et al., 1994, 1995, 1996). Due to numerical instabilities, it is necessary, however, to solve the flow in two steps, whereby the Polhausen method is used for the glottal inlet, but downstream a Blasius solution for flow over a flat plate is applied. This approach enables a moving separation point to form during the closing phases of the phonatory cycle, which is observed to first move upstream as the VFs first form a divergent channel, and to then move downstream shortly preceding complete glottal closure (Pelorson et al., 1994). Surprisingly, this more complex approach yields a solution that is very similar to the *ad hoc* area ratio criteria of Liljencrants (1991) when that value is set to 1.1 (Hirschberg et al., 1996). In order to circumvent the previously described complexity of solving the von Kármán equation in two sections, subsequent methods have proposed using Thwaites' method to predict flow separation. Although it slightly over-predicts the pressure recovery in the divergent glottis when compared to experimental investigations, it still provides good agreement and is significantly easier to implement numerically than the Polhausen approach (Vilain et al., 2004).

6.3. Computational fluid dynamics solvers

With increased computing power, more realistic flow solvers have been coupled with lumped-element VF models in order to avoid the need for *ad hoc* assumptions about the flow field. Although this approach eliminates the advantage from a computational cost standpoint of using a lumped-element VF model, it does elucidate the manner in which flow features impact VF dynamics. de Vries et al. (2002) couple a computational fluid dynamics (CFD) Navier–Stokes flow solver with a lumped-element VF model and compare the results to a traditional Bernoulli flow solver. They note that that the CFD solver produces a lower fundamental frequency and a peak flow that is twice as high as the Bernoulli solver. The increase in peak flow is attributed to the changing flow separation points throughout the phonatory cycle. In addition, the phonation threshold pressure is much more closely correlated with clinical results. However, this study, as well as prior works already discussed, assumes that the flow separates symmetrically within the glottis, whereas two-dimensional experimental investigations have demonstrated that during the divergent phases of the phonatory cycle, the flow fully separates from one wall and an attached wall jet forms on the opposing wall (Erath and Plesniak, 2006a, 2010a,b; Krane et al., 2007). This additional complexity has been investigated by coupling lumped-element VF models with full CFD solvers that are solved in tandem with the VF dynamics (Tao and Jiang, 2008; Tao et al., 2007; Xue et al., 2010). These studies show that even for symmetric tissue properties, asymmetric airflow can induce asymmetric VF motion, although the degree of asymmetry is relatively minor. Asymmetries in the VF motion increase with subglottal pressure, becoming more pronounced when tissue properties are also asymmetric, and can incite chaotic VF motion (Tao and Jiang, 2008).

6.4. A simplified asymmetric flow solver

Recently, a theoretical asymmetric intraglottal flow solver has been developed for lumped-element VF models. Using a Falkner–Skan-like similarity solution for flow over a rotating and translating flat plate, a model of the asymmetric pressure loading on the VFs is derived, termed the boundary-layer estimation of the asymmetric pressures (BLEAP) model (Erath et al., 2011a). This solution models the aerodynamic loading that arises from an attached wall jet. This scenario occurs when the glottis forms a divergent passage, causing the glottal jet to asymmetrically attach to one VF wall while completely separating from the opposing wall (Erath and Plesniak, 2006a, 2010c). This approach specifies flow separation *a priori* from the non-flow wall to occur at the minimal glottal area, while on the flow wall it occurs at the glottal exit. The flow wall is specified as the VF wall that has a shallower divergence angle when the total included glottal angle becomes divergent. This behavior has been confirmed with experimental measurements (Erath and Plesniak, 2006c, 2010a). The asymmetric wall pressure that is produced by the attached wall jet induces only slight asymmetries in the VF dynamics for symmetrically tensioned tissues. However, for asymmetric model parameters, indicative of unilateral paralysis, significant disruption of the dynamics occurs, which can introduce chaotic VF motion (Erath et al., 2011b). When a smooth glottal geometry is used, as in the coupled CFD/lumped-element model investigations of Tao and Jiang (2008) and Tao et al. (2007), very good agreement is observed between the CFD simulations and the BLEAP solution, suggesting that the BLEAP flow solver captures the pertinent physics of asymmetric intraglottal flow for both normal and pathological models with minimal additional computational cost.

6.5. Impact of the ventricular folds

While the ventricular folds significantly impact the fluid dynamics of phonation from both flow stability (Agarwal et al., 2003, 2004; Drechsel et al., 2008; Li, 2007; Triep et al., 2005; Zheng et al., 2009) and acoustical standpoints (Alipour, 2013; Zhang et al., 2002), very little work has been undertaken to model these effects from a fluid mechanics perspective. The sole effort (Bailly et al., 2008) assumes Bernoulli flow through the glottis, and then investigates both uniform and turbulent supraglottal jet development before passing through a ventricular gap, for which the pressure drop is again solved from Bernoulli's equation with corrections for flow inertia and viscous effects. When

compared with experimental results the theory provides good qualitative agreement. Furthermore, it successfully predicts that the ratio of ventricular to glottal area has a very significant impact on phonation onset pressure. For ventricular areas less than the glottal area, flow resistance is increased, which increases phonation onset pressure. As the ratio of ventricular to glottal area increases, an additional pressure recovery occurs, causing the phonation onset pressure to decrease below the level of that found for flow through the glottis with no ventricular folds. For area ratios greater than about 7, the ventricular folds have negligible impact on the flow and the phonation onset pressure acts as if there are no ventricular folds.

6.6. Three-dimensionality and higher-order flow effects

A shortcoming of existing analytical flow solvers is their inability to capture higher-order viscous flow effects that may have a significant impact on aerodynamic pressure loadings and acoustic output. The glottal jet is three-dimensional in nature (Triep and Brücker, 2010), and exhibits complex viscous behavior including axis-switching (Khosla et al., 2008a; Triep and Brücker, 2010), intraglottal and supraglottal vortex shedding (Khosla et al., 2007), shear layer instabilities (Luo et al., 2009), and cycle-to-cycle perturbations in flow trajectory (Erath and Plesniak, 2006c; Erath and Plesniak, 2010b; Neubauer et al., 2007). Although VF geometries have been developed to incorporate three-dimensional contours as discussed in Section 3, flow solvers applied to these three-dimensional geometries still largely assume one-dimensional inviscid flow. Since the geometry is commonly prescribed as a matrix of masses in both the inferior–superior (streamwise) and anterior–posterior (cross-stream) directions, each anterior–posterior section is assumed independent of the other sections from the fluid mechanics standpoint. It is further assumed that the subglottal and supraglottal pressures in each plane are the same, but due to the changing geometries for each plane, the intraglottal pressures varies in the anterior–posterior direction (Wong et al., 1991; Yang et al., 2010, 2011, 2012). This approach neglects viscous effects, and the observation that velocity magnitudes in the anterior–posterior (cross-stream) direction can be as high as 10% of the streamwise velocity in three-dimensional geometries (Scherer et al., 2010; Triep and Brücker, 2010; Triep et al., 2005; Zheng et al., 2011).

6.7. A water hammer-based solver

Other novel approaches also exist for modeling the fluid mechanics of voiced speech in lumped-element VF models, one of the more interesting of which is based upon water hammer theory. In Ishizaka et al. (1976) suggested that the water hammer phenomena may be used in the study of human phonation. Sciamarella and Artana (2009) formalize the water hammer theory for modeling human phonation, in which the vocal tract is approximated as a pipeline system with the VFs acting as a valve. The water hammer formulation solves for the flow as well as the unsteady pressure fluctuations in a naturally coupled manner, as opposed to the more traditional methods discussed in Section 7, in which a wave source model must be coupled to an acoustic analogy. While Sciamarella and Artana (2009) acknowledge this one-dimensional approach does not produce more accurate results, per se, it is a more intuitive and direct method for computing unsteady intraglottal pressures under the assumption of one dimensional flow. A shortcoming of this approach is, however, the reliance upon empirical orifice discharge coefficients, which have only been measured for limited scenarios (Mongeau et al., 1997; Park and Mongeau, 2007, 2008).

7. Components of lumped-element vocal fold models: Acoustics

7.1. Sound generation

Although a small fraction of the sound emitted by the vibrations of the VFs may be radiated through structural vibrations of the bones and cartilages, most of the acoustic energy is obtained from the pulsating glottal jet flow and its interaction with the human airways (Mongeau et al., 1997). The sound produced by the pulsing jet can be broadly categorized as monopole, dipole, and quadrupole sources. A periodic monopole source is produced due to displacement of the fluid by the moving tissue of the VFs (Zhao et al., 2002). A dipole source is generated by the net force that the surfaces of the glottal walls exert on the fluid, which is governed by the instantaneous pressure drop across the glottal opening (McGowan, 1988; Zhao et al., 2002). Finally, a quadrupole source, which is random and broadband in nature, arises from the turbulent kinetic energy of the flow (Zhao et al., 2002).

Self-oscillating lumped-mass models are capable of readily incorporating multiple sound sources. By specifying subglottal and supraglottal vocal tract geometries, the periodic dipole source is inertly present due to the presence of a subglottal tract and the continuity of volume below and above the glottis, which yields phase-inverted pressures on each side of the glottis. Additional sources can also be incorporated into these models. A monopole source can be added by calculating the instantaneous volume occupied by the folds, and quadrupole sources can be included by estimating the rate of change in the axial kinetic energy in the source region (Zhang and Mongeau, 2006; Zhao et al., 2002). Even though a simple addition of these components will better represent the physics of phonation, these higher-order terms are normally neglected and only the predominant dipole source is utilized. The assumption of a periodic dipole-only source may be sufficient to study modal phonation, but it falls short in the instances of higher pitch frequencies due to an increased monopole component (Zhao et al., 2002). Furthermore, the presence of ventricular folds or vocal tract occlusions will give rise

to an additional dipole component (Zhang et al., 2002), and incomplete glottal closure will increase the contribution of the quadrupole component (Park and Mongeau, 2008; Zhao et al., 2002).

7.2. Sound propagation

In general, sound propagation in the subglottal and supraglottal tracts is modeled in the time domain under a one-dimensional planar wave assumption. The three-dimensional tracts are discretized into a finite number of short length cylindrical sections with variable cross-sectional areas. The tract shapes are typically generated from magnetic resonance imaging (MRI) data (Story, 2008; Story et al., 1996; Takemoto et al., 2006). Sound pressure waves are tracked over time across each cylindrical section, where the acoustic pressure is assumed to propagate as a planar wave. The plane wave assumption in voiced speech is a reasonable approximation for frequencies less than approximately $4700 - 5700$ Hz (Eriksson, 1980). Tube curvature, which is most prevalent in the vocal tract, can either increase or decrease the phase velocity of acoustic waves around the bending region, thereby altering the resonances of the tube. However, as discussed by Story et al. (1995), these effects are minor and can be reasonably neglected.

Two standard approaches exist for tracking acoustic wave propagation in the subglottal and supraglottal regions. The wave reflection analog (WRA) scheme (Kelly and Lochbaum, 1973; Liljencrants et al., 1985; Story et al., 1995) is a time-domain transmission model, where the solution of the plane wave equation for each tube section is expressed as the superposition of forward and backward traveling acoustic pressures, see Fig. 5. In this approach, the surface area of each section determines reflection coefficients and each transmitted and reflected pressure is computed using scattering equations. The technique is conceptually simple and of low complexity for simple geometries. It allows for radiation impedance, compliant walls, and various other loss factors for the subglottal and supraglottal tracts (Story et al., 1995; Zañartu, 2006). When utilizing MRI data to construct the vocal tracts, the WRA model shows good agreement of the first three formant locations when compared with clinical data (Story et al., 1996; Titze, 1994b). The technique, however, becomes less intuitive and less efficient when handling tube branching, frequency dependent losses, and temporal variations in the vocal tract. Thus, it is less attractive for modeling running speech and physiological representations of the subglottal pulmonary system.

An alternative sound propagation model uses a distributed transmission line method based upon a cascade of *T*-sections. Each section is constructed using electrical analogs of the acoustic components, see Fig. 6. This approach has been used extensively in the frequency domain (Fant, 1960; Flanagan et al., 1972; Stevens, 1998; Stevens and House, 1955; Wodicka et al., 1989). It was first applied in a time domain scheme by solving loop equations for each node in the *T*-sections in early self-oscillating lumped-mass models (Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972). In order to simplify the computations, Maeda (1982) proposes a matrix representation of a time-domain transmission line network for sound wave propagation in the vocal tract. This method has been extended to include multiple side branch cavities of the supraglottal tract (Mokhtari et al., 2008) and the subglottal system (Ho et al., 2011). Even though current transmission line methods overcome the limitations of WRA methods for complex tract representations, the simplicity of the latter still prevails when studying static vowel production using self-sustained models.

7.3. Acoustic interactions

The seminal *linear* source-filter theory (Fant, 1960) defines the vocal tract as a simple filter that convolves the glottal sound source. No interaction between the VF tissue, glottal airflow, or acoustic waves is considered, and any effects arising from the subglottal tract are neglected. Subsequent theoretical studies, however, observed changes in the glottal airflow due to fluid-acoustic interactions with the subglottal and supraglottal vocal tract (Fant and Lin, 1987; Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972; Rothenberg, 1981; van den Berg, 1968). Clinical evidence of these early observations are also reported in the literature (Hanson, 1997; Klatt et al., 1990; Rothenberg, 1973; Rothenberg and Zahorian, 1977).

Recent studies suggest that source-filter interactions can be much stronger than initially believed, and completely alter the dynamics of VF tissue motion in addition to producing fluid-acoustic interactions. This claim is primarily supported by observations made possible with advances in vocal tract imaging (Honda et al., 2004; Story, 2008; Story et al., 1996; Takemoto et al., 2006) that show vocal tract impedance can be comparable to that of the glottis (Titze, 2004, 2008; Titze and Story, 1997). The *nonlinear* source-filter coupling theory (Titze, 2008) seeks to incorporate these observations and introduces the concept of interactions due to impedance relations. Supporting evidence for this theory can be observed in register transitions in singers (Miller and Schutte, 2005), dynamic vocal exercises

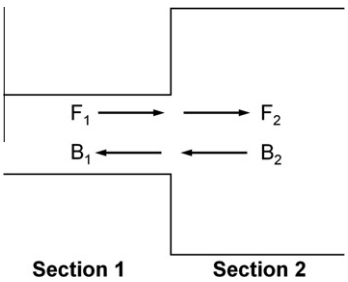

Fig. 5. Schematic of the WRA principle, where F_1 and F_2 are forward traveling waves upstream and downstream of the junction, respectively, while B_1 and B_2 are the backwards traveling analogs.

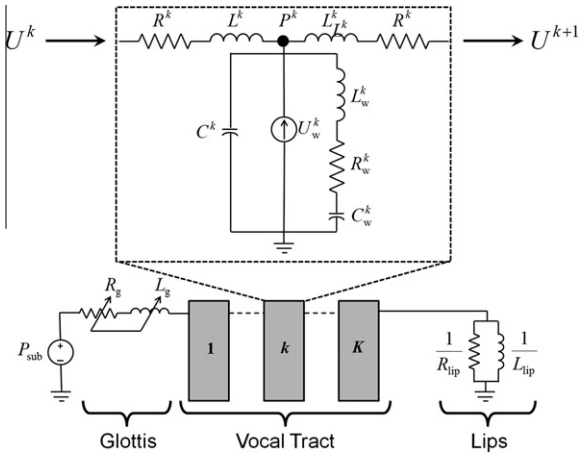

Fig. 6. Schematic of the transmission line model. Each circuit node consists of a series of resistors R, capacitors C, and inductors L that model flow through the tube. The wall impedance is similarly modeled and denoted by the subscript w. The glottal airflow into the kth node, U^k, is combined with a flow U_w to produce the outflow U^{k+1}. The flow U_w accounts for air flow generated due to the articulated volume changes within each segment. The pressure at the node is P^k.

(Titze et al., 2008; Titze and Worley, 2009), and high-speed visualization of acoustically-induced bifurcations (Zañartu et al., 2011).

The nonlinear source-filter theory proposes two types of source-filter interactions (Titze, 2008): a "level 1" interaction, in which only the flow and sound interact; and a "level 2" interaction, in which the tissue is now part of a three-way interaction that affects the tissue motion. These two classifications are not purely academic, as it has been suggested that humans have the ability to operate their voices with different levels of source-filter interactions by changing the degree of adduction of the VFs and narrowing the epilarynx tube (Titze, 2008). This implies that a "level 2" interaction could be simplified into a "level 1" interaction under certain conditions.

When implementing "level 1" interactions into lumped-element VF models, the acoustic pressures are not included in the differential equation that defines the motion of the VFs, but are applied separately to the equations that compute the glottal flow rate. Thus the VF dynamics are unaltered, but the volumetric flow rate (i.e., the acoustic source) is influenced by acoustic interactions in the subglottal and supraglottal tracts. Consequently, acoustic interactions arise from the relation between tract impedances and the harmonic composition of the glottal source. Harmonics falling within the inertive portion of the vocal tract are amplified, while those falling within the compliant portion are reduced. "Level 1" interactions also introduce harmonic distortion, increasing the frequency content of the glottal flow signal.

In "level 2" interactions the subglottal and supraglottal tracts impact both the glottal airflow and the VF dynamics through the governing differential equations. This type of acoustic interaction can yield favorable conditions for self-sustained oscillations, producing increased amplitudes of vibration, increased radiated sound pressure levels, and reduced phonation threshold pressures (Alipour et al., 2001; Titze, 2008). Likewise, unfavorable conditions raise the phonation threshold, yield smaller amplitudes, and can produce instabilities and bifurcations, such as frequency jumps, deterministic chaos, aphonation, and subharmonics (Titze, 2008; Titze et al., 2008; Tokuda et al., 2010; Zhang et al., 2006; Zañartu et al., 2007, 2011). The general "level 2" conditions where the acoustic interactions favor or hamper the oscillations are described for each frequency in terms of the series superposition of the tract reactances (Titze, 2008). Favorable loading conditions are produced by an inertive impedance (positive reactance) in the supraglottal tract and compliant impedance (negative reactance) in the subglottal tract (Titze, 2008; Zañartu et al., 2007). For normal voiced speech at the fundamental frequency, the vocal tract and the subglottal tract provide inertive impedances. Consequently, the ideal condition is not met and the subglottal tract acts to hamper the oscillations to some degree. However, because harmonics are frequency dependent and different impedance relations may exist, some frequency components may be amplified while others are attenuated.

8. Applications of lumped-element models in speech research

In Sections 3–7 we introduced the constituents of the lumped-element modeling framework. Therein we focused primarily on the functional aspects of each constituent separately. In this section we shift the focus to what lumped-element models can do and how they can do it. This section will highlight a few key demonstrations of lumped-element models in various speech studies, and as such is not meant to serve as a comprehensive literature review on the subject.

8.1. Normal phonation

In general, under normal phonation conditions the left and right VFs are structurally equivalent; that is, the tissue properties of the two VFs are identical. The myriad sounds associated with voiced speech result from control and alteration of the vocal tract configuration and VF tensioning and posturing, see Section 7.3. Herein we discuss speech simulations ranging in complexity from sustained vowels to running speech.

8.1.1. Sustained vowels

In sustained vowel vocalization, the VF fundamental frequency and vocal tract configuration remain constant during phonation. The fundamental frequency is set by the VF vibration while the sound spectrum is influenced by the resonances of the vocal tract. Consequently, modeling sustained vowel phonation requires an appropriate glottal flow, a VF model with the proper fundamental

vibration frequency, and a vocal tract for acoustic filtering (Story, 2005).

Lumped-element models of the vocal folds are well-suited to sustained vowel simulations since the structural parameters are constant and can be tuned to obtain the desired frequency, see Section 4. As discussed in Section 7.2, MRI data is frequently used to extract the vocal tract configurations for vowel intonations, which can be coupled with the VF model and solved using the wave reflection analog method. Story et al. (1996, 2002), for example, use a three-mass body-cover model ($m_1 = 1$, $n_1 = 2, m_2 = n_2 = 1$) connected with MRI-measured subglottal and supraglottal tracts and the WRA method to investigate a range of sustained vowels. The pressure waveforms for the /i/ and /a/ phonemes display a clear influence of the vocal tract resonances.

As an example of the practical utility of sustained vowel simulations, new voice production techniques can be evaluated using data produced from lumped-element models. For instance, a three-mass body-cover model ($m_1 = 1, n_1 = 2, m_2 = n_2 = 1$) and the WRA method are employed to evaluate the performance of an improved closed-phase covariance method for glottal inverse filtering by Alku et al. (2009). The new inverse filtering technique is evaluated by comparing the glottal volume velocity waveform estimated by applying the new inverse filtering method to an acoustic signal generated from the lumped-element model to the "actual" glottal waveform produced by the model.

8.1.2. Vocal registers and pitch glide

Vocal registers are associated with specific VF vibration regimes and are often described by their glottal orientation and muscular tension. Sounds produced in a given register will be similar due to the consistent VF vibration patterns. For example, in the modal register, the VFs are fully adducted and a relatively large portion of the folds vibrate. In contrast, in the falsetto register the majority of the motion is relegated to the trailing edge of the folds with little motion in the VF body. A pitch glide is a steady change (typically rise) of pitch at constant loudness and effort. Increasing pitch is largely accomplished by increasing vocal fold tension, particularly in modal voice. Physiologically, as the cricothyroid contracts, the VF tension increases, which can result in the VF vibration pattern transitioning from chest to falsetto register.

Lumped-element models are configured to simulate various registers, including vocal fry, chest, and falsetto, via alteration of the model parameters and/or model structure. Vocal fry is characterized by low-frequency glottal pulses with a short open phase, which is mimicked by using less stiff linear spring constants. In modal voice, or chest register, the VFs are large and close completely during each glottal cycle. In lumped-element models, this is achieved by increasing the VF mass. The bulk of the oscillations occur at the edge of the VFs during falsetto voice due to increased tension and decreased mass. Correspondingly, falsetto is modeled by decreasing VF mass, and increasing the linear and coupling spring constants (Sciamarella and d'Alessandro, 2003, 2004).

Lumped-element models are capable of simulating pitch glides by adaptively modifying model parameters. For example, a pitch raise is modeled using a multi-mass model ($m_1 = 3, n_1 = 2$) with time-dependent parameters by Wurzbacher et al. (2004, 2006). They use an iterative adaptation method that modulates the frequency and amplitude of the model to match clinical vibration patterns measured using high-speed glottography. The model very accurately tracks the increasing frequency and subglottal pressures of the simulated pitch raise and is able to replicate the pitch raise acquired using high-speed glottography with only slight discrepancies around the time of VF impact.

The increased cricothyroid tension needed for register transition can be introduced to single layer lumped-element models via a tensioning parameter, which increases the stiffness and decreases the mass of the superior mass in the model. Sciamarella and d'Alessandro (2004) use a two-mass model ($m_1 = 1, n_1 = 2$) to study vocal register transitions from chest to falsetto. The three-mass model ($m_1 = 1, n_1 = 3$) introduced by Tokuda et al. (2007, 2008) and discussed in Section 3.1 is specifically designed to simulate both chest and falsetto registers, as well as the transition between them, assuming no influence of vocal tract resonances. Using Lyapunov exponents, they conclude that the VF vibration during register transition, which occurs over a range of tensioning parameters, is chaotic. Their results demonstrate good agreement with analogous excised larynx experiments, in which the longitudinal tension is varied to cause register transition. Their improved dual-layer version of the model ($m_1 = 1, n_1 = 3, m_2 = n_2 = 1$), discussed in Section 3.2, coupled with subglottal and supraglottal resonators is used to replicate glissando singing of untrained male singers by varying the tension parameter (Tokuda et al., 2010). This study displays very similar register transitions and frequency jumps between chest and falsetto registers. The study finds that coupling of the subglottal and supraglottal resonances to the VF model results in a drastic shift in the frequencies at which the register transitions occur, with the transition dependency being dominated by the supraglottal resonance.

8.1.3. Running speech

Running speech involves a sequence of sounds generated in myriad ways, including fricatives, plosives, and the voiced sounds generated at the VFs. Simulating full running speech requires models capable of modifying the vocal tract, articulating and posturing the VFs, modulating VF parameters, as well as incorporating other sound sources for non-voiced sounds. Lumped mass models utilizing time-varying laryngeal parameters in combination with source-tract acoustical interactions are capable of simulating running speech to some degree. Running speech simulations can be achieved by dynamically fitting the model to oral airflow measurements acquired using a Rothenberg mask (Lofqvist et al., 1995; Lucero, 2005; McGowan

et al., 1995). McGowan et al. (1995) simulate an /aCa/ utterance, where C is replaced with a range of consonants, using a two-mass model ($m_1 = 1, n_1 = 2$) and a low-frequency aerodynamic model. Although the model is able to generally replicate the glottal dynamics, the authors suggest that the two-mass model is too simplified for accurate simulation of glottal pulses during the transition from unvoiced to voice speech. Lucero (2005) simulate the specific utterance /aha/ for instances where the /h/ is both voiced and unvoiced. Their iterative model is able to recreate the voiced syllables with good agreement in the voiced /h/ cases. In the cases containing an unvoiced /h/, the model is less accurate in restarting the oscillation after its cessation, resulting in a nonphysical elongated unvoiced period. Generally speaking, lumped-element models are not good at modeling unvoiced sounds, in part due to neglecting higher order sound sources, see Section 7.1.

8.2. Pathological phonation

Normal phonation is generally periodic, however small perturbations are often present in the sound signal. The introduction of pathological conditions most often result in much more substantial perturbations or irregularities of the vocal signal. A subset of pathological phonation conditions, in which lumped-element phonation modeling is utilized for scientific understanding, advancements, and treatment, is presented in this section. A summary of the vocal pathologies discussed herein, including their physiological manifestations and the requisite lumped-element model adaptations, is presented in Table 3.

8.2.1. Incomplete glottal closure

Incomplete glottal closure, or glottal leakage, is a common condition in both normal and disordered voices (Hanson, 1997; Hillman et al., 1989; Holmberg et al., 1988; Stevens, 1998). As the name suggests, incomplete glottal closure is a condition such that even when the VFs are completely adducted there remains a gap through which air can pass. It may present in healthy normal voice, as is often observed in females, see for example Fig. 7. If severe enough, the leakage flow can result in diminished vocal quality to the point of being pathological. Incomplete glottal closure may also arise as a manifestation of some other underlying pathology, such as polyps or paralysis, see Fig. 9. Indeed, incomplete glottal closure is related to changes in the harmonic decay of the source spectra and changes in the vocal tract formant and bandwidths (Hanson, 1997; Klatt et al., 1990; Rothenberg, 1984), the introduction of turbulent noise (Klatt et al., 1990; Stevens, 1998), fluctuations during the closed phase of the cycle (Cranen and Boves, 1987; Cranen and Schroeter, 1995), large open quotients (Sciamarella and d'Alessandro, 2003, 2004), and the need for an increased vocal effort that may lead to vocal hyperfunction (Hillman et al., 1989).

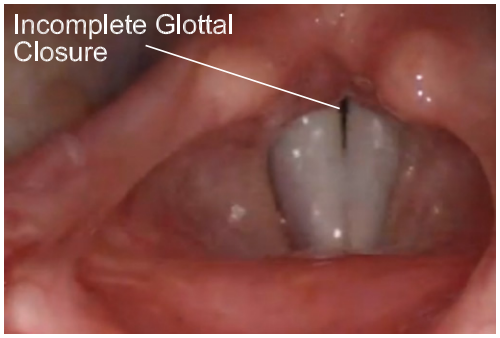

Fig. 7. Laryngocopic image of adducted VFs of a normal female showing non-pathological incomplete glottal closure. Figure taken from http://www.voicemedicine.com, used with permission from Lucian Sulica, MD.

Table 3
Lumped-element model adaptations for modeling pathologies.

Pathology	Physiological manifestations	Model adaptations
Incomplete glottal closure	Incomplete closure of the posterior glottis resulting in leakage during VF closure	Three-dimensional models allowing abduction; bypass channel to allow flow during closure
Unilateral paralysis		
RLN	Inability to adduct, position, and tension affected VF; incomplete closure	Decrease mass and increase all stiffness except coupling springs of damaged VF
SLN	Inability to tension affected VF	Decrease mass and increase all stiffnesses of damaged VF
Polyps and nodules	Geometric abnormalities on VF medial surface	Increase mass of affected fold; modified nonlinear collision force; modified pressure loading
Parkinson's disease	Impairment of muscle voluntary movement, reduced muscle speed, muscle rigidity; presents with vocal tremors, subharmonics, vocal irregularities	Time-varying structural parameters; increased stiffness

While a common occurrence, very few lumped-element model studies have incorporated incomplete glottal closure. Many models, particularly those with no anterior–posterior variation, restrict the glottal area to be rectangular, and thus when the opposing masses collide the glottis is completely closed. In order to incorporate incomplete glottal closure, there must be either anterior–posterior variation or an *ad hoc* correction to a planar model (Birkholz et al., 2011a; Childers et al., 1986). Different extensions or corrections are proposed that include: an anterior–posterior feature that restricts the vibration of the vocal folds (McGowan et al., 1995); use of nonlinear damping coefficients (Lucero and Koenig, 2005), a pre-collision scheme (Kuo et al., 1998; Pelorson et al., 1994), a separate flow channel to mimic a posterior parallel gap (Zañartu, 2010), and inclined masses with respect to the dorso-ventral axis that affects the flow solver and contact forces (Birkholz et al., 2011c). These extensions or corrections to models with no anterior–posterior variation are capable of mimicking the perceived effects of incomplete glottal closure, although no comprehensive comparison among them has been performed. Some of these extensions also explore the effect of incomplete glottal closure on the energy transfer that leads to hyperfunction (Zañartu, 2010) and the production of different perceived voice qualities, ranging from pressed to breathy voice (Birkholz et al., 2011c).

On the other hand, lumped-element models that include anterior–posterior variations (Fraile et al., 2012; Kob et al., 2002; Titze, 2006; Wong et al., 1991) can allow for incomplete glottal closure by means of separate Bernoulli flow channels for each section (Titze, 2006). Even though it increases the numerical complexity, this modeling approach can mimic various normal and pathological conditions and voice qualities, given the added degrees of freedom in the anterior–posterior direction.

8.2.2. Unilateral laryngeal nerve paralysis

As discussed in Section 2, the RLN innervates all intrinsic muscles of the larynx except the cricothyroid, which is innervated by the SLN. The cricothyroid controls VF elongation and tension, while the remaining muscles largely serve to position the VFs in preparation for voiced speech. Unilateral damage to the laryngeal nerve, which may arise from neurological disease, viral infection, or thyroid surgery (Lo et al., 1960; Schwarz et al., 2006; Yumoto et al., 2002), inhibits proper voice production by either (i) preventing adduction of one VF (RLN paralysis), see Fig. 8, or (ii) inhibiting the ability to control VF tension when both VFs are adducted (SLN paralysis). In addition, both types of paralysis often create a persistent posterior glottal gap. RLN paralysis is typically manifest by decreased mucosal wave propagation, a phase lag caused by the healthy VF leading the damaged VF (Sercarz et al., 1992), and exaggerated lateral motion in the damaged VF (Hanson et al., 1988). In SLN paralysis, phase lag between the healthy and damaged fold can be as high as 90° (Hanson et al., 1988). Due to the lack of tension, the damaged VF is also usually at a lower vertical level (Arnold, 1961; Dursun et al., 1996).

Prescribed tension asymmetries in lumped-element investigations vary based on the pathology being modeled. By applying different tension parameters for the left and right VFs, which alter the masses and springs of one VF in inverse proportion, Ishizaka and Isshiki (1976) investigate vibratory regimes as a function of degree of asymmetry. Using the two-mass model ($m_1 = 1, n_1 = 2$) of Koizumi et al. (1987), Smith et al. (1992) introduce additional asymmetry parameters to model the size of the vibratory gap and thyrovocalis contraction and tension, in addition to cricothyroid contraction, enabling the ability to model both RLN, SLN, and combined RLN/SLN paralysis. A simple and popular method for modeling VF paralysis is presented by Steinecke and Herzel (1995). SLN paralysis is modeled with all springs scaling linearly with an asymmetry parameter, and masses scaling inversely. RLN paralysis is also modeled by scaling the coupling and collision springs and mass of one fold. Lumped-mass investigations of VF paralysis have demonstrated that tension asymmetries can introduce highly irregular vocal fold kinematics characterized by biphonation and multiple period oscillations (Steinecke and Herzel, 1995; Xue et al., 2010), as well as highly-nonlinear, chaotic motion (Erath et al., 2011b; Herzel and Knudsen, 1995; Herzel et al., 1995). Lumped-element models of unilateral paralysis are also used successfully to extract physiological asymmetry parameters by correlating high-speed video with model kinematics using the methods described in Section 4.3 (Mehta et al., 2011; Wurzbacher et al., 2008).

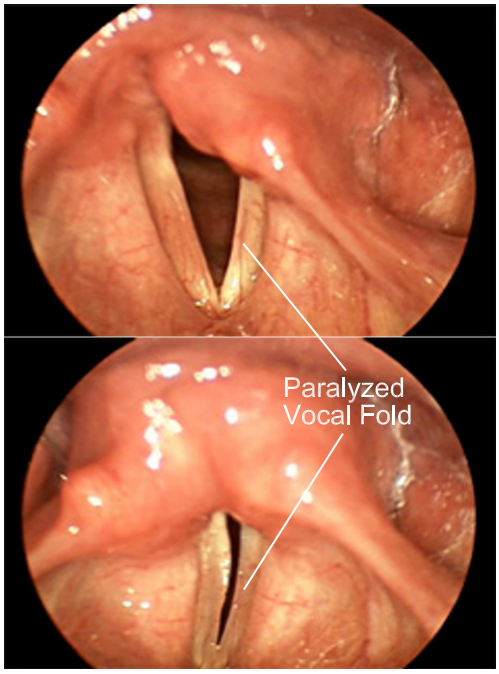

Fig. 8. Laryngoscopic images VFs with unilateral paralysis in the abducted (top) and adducted (bottom) configuration. Figures taken from http://www.voicemedicine.com, used with permission from Lucian Sulica, MD.

8.2.3. Polyps and nodules

Polyps and nodules are common pathologies that present as geometric abnormalities on the medial surface of the VFs (Benjamin and Croxson, 1987; Hirano et al., 1983), see Fig. 9. Polyps range in size, and may be classified as one of three types: sessile, which results in a distributed balloon like appearance; pedunculated, which is more stalk-like, with a base and body; and polypocorditis, which consists of a lesion that is distributed over the entire VF surface (Titze, 1994B). The geometric protuberance often prevents complete closure of the glottis, resulting in breathiness, and vocal fatigue (Johns et al., 2003), as well as increased phonation threshold pressure (Zhuang et al., 2009). Size, stiffness, and damping coefficients are all important polyp parameters that alter VF dynamics due to the impediment of the mucosal wave (Zhang et al., 2004). Investigations of a polyp-like mass added to excised canine VFs further reveal the inception of spatiotemporal asymmetry and chaos in the vibratory patterns (Zhang and Jiang, 2008a). In addition to structural modifications, aerodynamic pressure loadings may also be disrupted due to a polyp obstructing the glottal flow (Erath and Plesniak, 2012).

Dejonckere and Kob (2009) utilize a single layer model ($m_1 = 20, n_1 = 2$) to investigate the pathogenesis of vocal nodules, showing that anatomical differences in VF posturing can lead to localized zones of high impact along the VF surface. The incorporation of nodules and polyps into lumped element investigations has been limited to structural variations of the VF model. By altering the cover stiffness in a single-layer model ($m_1 = 5, n_1 = 2$) Wong et al. (1991) demonstrate that localized increases in stiffness damp subharmonics and increase irregularity. A two-mass model ($m_1 = 1, n_1 = 2$) with an added mass on the medial surface of the inferior mass, representative of a sessile polyp, is shown to incite aperiodic vibrations that can be attributed to low-dimensional chaos, while also producing a non-zero mean flow due to inhibited glottal closure (Zhang and Jiang, 2004). The more extreme case of a "golf-ball-like" pedunculated polyp is presented by Koizumi et al. (1993), where fluctuations in the fundamental period and amplitude are modeled well using a two-mass model ($m_1 = 1, n_1 = 2$) with a coupled mass on the superior surface of one side ($m_1 = 1, n_1 = 3$) to represent the pedunculated polyp. In spite of the relative success in modeling structural variations that arise due to polyps, the disruption of the aerodynamic loading due to the protruding geometric abnormality remains unknown. Although recent experiments have identified significant spatial and temporal variations in the glottal flow patterns due to periodically shed horseshoe vortices from a polyp, the impact on the pressure loadings remains unknown (Erath and Plesniak, 2012).

8.2.4. Parkinson's disease

Parkinson's disease is a neurological degenerative disorder. Approximately 90% of patients with Parkinson's disease exhibit laryngeal disorders including breathiness, and vocal tremor (Logemann et al., 1978), which can lead to pathological hoarseness (Massey and P.G.W., 1985). Tremulous voice as a result of Parkinson's disease, which typically occurs in the range of $4-7$ Hz, may also present with elevated frequency and amplitude perturbations (Hartman and Vishwanat, 1984).

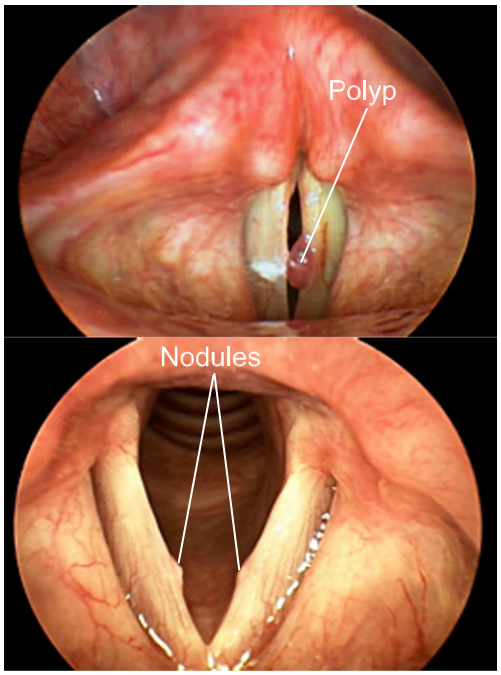

Fig. 9. Laryngoscopic images of VFs with polyps (top) and nodules (bottom). Figures taken from http://www.voicemedicine.com, used with permission from Lucian Sulica, MD.

Parkinsonian voice is modeled by including an additional normalized stiffness coefficient to account for increased tissue stiffness, and to simulate low-frequency fluctuations (Zhang et al., 2005). The stiffness coefficient utilizes the standard tissue properties of Ishizaka and Flanagan (1972) while also allowing for a linear increase in stiffness, with a superimposed sinusoidal oscillation with variable amplitude and frequency. This approach enables investigation of clinical manifestations of Parkinsonian voice including incomplete glottal closure, increased phonation threshold pressure, glottal tremor, and the emergence of chaos for sufficiently large tremor amplitude (Zhang and Jiang, 2008b; Zhang et al., 2005). Two-mass models ($m_1 = 1, n_1 = 2$) also show promise as a prosodic analysis tool for Parkinsonian voice (Bocklet et al., 2011).

8.3. Speech synthesis

Speech synthesis using physically-based modeling simulates the radiated speech signal by mimicking the behavior of the whole speech production system for running speech conditions. This approach differs from the more traditional concatenative speech synthesis (Schroete, 2008) in that the speech is originated from a physiologically-relevant

description rather than an *ad hoc* collection of short samples of recorded voice. Due to its high computational load, this type of synthesis is feasible only by means of low-order lumped-element models of the VFs and plane wave acoustic propagation (via WRA or time-domain transmission line) in a time-varying vocal tract.

Physically-based speech synthesis is primarily directed toward understanding speech production and speech perception. Having access to the complete system behavior allows the study of numerous research problems, such as the dynamics of co-articulation (Birkholz et al., 2011b; Story, 2007a; Story et al., 2010), acoustically-induced instabilities (Titze, 2008), the effects of turbulence in time-varying vocal systems (Birkholz et al., 2007), speech perception (Bunton and Story, 2010; Kröger et al., 2007), speech development in children (Bunton and Story, 2011; Story, 2009), speech development in robotic speech (Kröger et al., 2011a), and singing voice (Bailly et al., 2010; Titze, 2008; Titze and Worley, 2009). This method also has the potential to mimic emotional variations, different voice qualities, and other transformations that are not possible to achieve for a single subject via concatenative speech synthesis (Birkholz et al., 2011c; Story, 2007b).

As expected, physically-based speech synthesis requires knowledge of numerous time-varying inputs and physical phenomena that are generally not known, often producing uncertain results and unintelligible speech. Efforts are under way to better understand articulatory and laryngeal movements in running speech to improve the current limitations of this synthesis approach (Birkholz et al., 2011b; Kröger et al., 2011b,c).

9. Outlooks

Currently, there is a need to advance the understanding of normal and pathological voice in order to assist in clinical assessment and treatment of vocal function. Difficulties arise when attempting to correlate irregularities in the acoustical output with the underlying physical pathology producing the anomaly. That is to say, in pathological voice what is "heard" does not always correlate with what is "seen" (e.g., high-speed video) and/or "measured" (e.g., airflow). One patient may have a normal acoustical output despite irregular VF motion, while another with similar VF motion may present severely degraded speech quality. These seemingly incongruous observations give rise to questions such as; up to what point are pathological manifestations (e.g., incomplete glottal closure, polyps/nodules, asymmetric motion, etc.) detrimental to acoustical output; when should surgical intervention be considered in place of clinical rehabilitation; and how can surgical procedures be optimized to maximize vocal restoration? Lumped-element investigations have the potential to provide all of these answers since they are capable of modeling the fully-coupled fluid-structure-acoustic interactions that occur in voiced speech. However, in order to provide reliable and accurate insight, it is necessary that the models first capture the pertinent physics of the problem in order to relate abnormal vocal outcomes with pathological physical behaviors. For this reason, there is a need to advance the current state-of-the-art in lumped-element investigations to include higher-order physics modeling. Areas where advancements can be made have been discussed in this manuscript, and are briefly summarized below.

Recent experimental and numerical work has highlighted the importance of three-dimensional effects in voiced speech (Triep and Brücker, 2010; Yang et al., 2012; Zheng et al., 2011). These findings extend to both fluid and tissue mechanics. Glottal flow fields are highly three-dimensional, giving rise to anterior–posterior pressure gradients (Scherer et al., 2010), axis-switching in the glottal jet (Khosla et al., 2008a; Triep and Brücker, 2010), vortex shedding (Khosla et al., 2007, 2008b; Triep and Brücker, 2010), and glottal jet instabilities (Erath and Plesniak, 2006b; Erath and Plesniak, 2010b; Mittal et al., 2013; Neubauer et al., 2007; Zheng et al., 2011). Similarly, VF motion is not planar, but three-dimensional. Furthermore, active control of the larynx and surrounding structures as a means of reproducing posturing, control, and articulation of the supraglottal VF tract is needed to mimic higher-order vocal outcomes (Birkholz et al., 2011a). Closely related to the structural mechanics of the VFs, the contact mechanics of VF collision are not yet well understood, where accurate contact force modeling has implications for determining VF dynamics as well as injury initiation and development (Titze, 1994a). The underlying assumptions for some, or all three, acoustic components (generation, propagation, and interactions) may also differ considerably under pathological conditions. Since the contribution of acoustic sources and the energy exchange due to acoustic interactions can vary under different laryngeal configurations, it is not unlikely that sound propagation for pathologies would introduce considerable differences, including the emergence of completely new sound sources (e.g., a vortex sound component produced by hairpin vortices shedding from a polyp (Erath and Plesniak, 2012)).

In order to address and validate all of the aforementioned shortcomings, more direct comparisons between clinical data and lumped-element model findings are needed. To date, there has largely been a disconnect between model observations and clinical data. By comparing model behavior with human subject data, the field has the potential to reach the point where we can fully-explain and predict both normal and pathological voice behaviors, thereby creating a path by which scientific modeling can aid and support clinical and surgical decisions.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. CBET 1036280. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the

authors and do not necessarily reflect the views of the National Science Foundation. The work of Matías Zañartu was supported by UTFSM and CONICYT, Grant FONDECYT 11110147

References

Agarwal, M., Scherer, R.C., Hollien, H., 2003. The false vocal folds: shape and size in frontal view during phonation based on laminagraphic tracings. J. Voice 17 (2), 97–113.

Agarwal, M., 2004. The false vocal folds and their effects on translaryngeal ariflow resistance, Ph.D. thesis, Bowling Green State University, Bowling Green, OH.

Alipour, F., Titze, I.R., 2013. Ventricular pressures in phonating excised larynges. J. Acoust. Soc. Am. 132 (2), 1017–1026.

Alipour, F., Montequin, D., Tayama, N., 2001. Aerodynamic profiles of a hemilarynx with a vocal tract. Ann. Otol. Rhinol. Laryngol. 110 (6), 550–555.

Alipour-Haghighi, F., Titze, I.R., 1991. Elastic models of vocal fold tissues. J. Acoust. Soc. Am. 90, 1326–1331.

Alku, P., Magi, C., Yrttiaho, S., Bäckström, T., Story, B., 2009. Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering. J. Acoust. Soc. Am. 125 (5), 3289–3305.

Arnold, G.E., 1961. Physiology and pathology of the cricothyroid muscle. Laryngoscope 71, 687–753.

Avanzini, F., 2008. Simulation of vocal fold oscillation with a pseudo-one-mass physical model. Speech Comm. 50, 95–108.

Avanzini, F., Alku, P., Karjalainen, M., 2001. One-delayed-mass model for efficient synthesis of glottal flow, In: Seventh European Conference on Speech Communication and Technology, pp. 51–54.

Avanzini, F., Maratea, S., Drioli, C., 2006. Physiological control of low-dimensional glottal models with application to voice source parameter matching. Acta Acust. 92, 731–740.

Baer, T., 1981. Investigation of the phonatory mechanism. ASHA Report 11, 38–46.

Bailly, L., Pelorson, X., Henrich, N., Ruty, N., 2008. Influence of a constriction in the near field of the vocal folds: physical modeling and experimental validation. J. Acoust. Soc. Am. 124, 3296–3308.

Bailly, L., Henrich, N., Pelorson, X., 2010. Vocal fold and ventricular fold vibration in period-doubling phonation: physiological description and aerodynamic modeling. J. Acoust. Soc. Am. 127, 3212–3222.

Baken, R.J., 1997, second ed.. In: Professional Voice: The Science and Art of Clinical Care Singular Publishing Group, San Diego, CA.

Benjamin, B., Croxson, G., 1987. Vocal nodules in children. Ann. Oto. Rhinol. Laryngol. 99 (5), 530–533.

Birkholz, P., 2011. A survey of self-oscillating lumped-element models of the vocal folds. In: Kröger, B.J., Birkholz, P., (Eds.), Studientexte zur Sprachkommunication: Elektronische Sprachsignalverarbeitung, Dresrmany, pp. 184–194.

Birkholz, P., Jackel, D., Kröger, B.J., 2007. Simulation of losses due to turbulence in the time-varying vocal system. IEEE Trans. Audio Speech Lang. Process. 15 (4), 1218–1226.

Birkholz, P., Kröger, B.J. Neuschaefer-Rube, C., 2011a. Articulatory synthesis of words in six voice qualities using a modified two-mass model of the vocal folds. In: First International Workshop on Performative Speech and Singing Synthesis.

Birkholz, P., Kröger, B.J., Neuscheafer-Rube, C., 2011b. Model-based reproduction of articulatory trajectories for consonant–vowel sequences. IEEE Trans. Audio Speech Lang. Process. 19 (5), 1422–1433.

Birkholz, P., Kröger, B.J., Neuscheafer-Rube, C., 2011c. Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis. In: Proc. of the Interspeech 2011, Florence, Italy, pp. 2681–2684.

Bocklet, T., Nöth, E., Stemmer, G., Ruzickova, H., Rusz, J., 2011. Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis. In: Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp. 478–483.

Brown, B.L., Strong, W.J., Rencher, A.C., 1974. Fifty-four voices from two: the effects of simultaneous manipulations of rate mean fundamental frequency and variance of fundamental frequency on ratings and personality from speech. J. Acoust. Soc. Am. 55, 313–318.

Bunton, K., Story, B.H., 2010. Identification of synthetic vowels based on a time-varying model of the vocal tract area function. J. Acoust. Soc. Am. 127 (4), EL146–EL152.

Bunton, K., Story, B.H., 2011. A test of formant frequency analyzes with simulated child-like vowels. J. Acoust. Soc. Am. 129 (4), 2626.

Chan, R.W., Titze, I.R., 1999. Viscoelastic shear properties of human vocal fold mucosa: meaurement methodology and empirical results. J. Acoust. Soc. Am. 106, 2008–2021.

Chan, R.W., Titze, I.R., 2000. Viscoelastic shear properties of human vocal fold mucosa: theoretical characterization based on constituitive modeling. J. Acoust. Soc. Am. 107, 565–580.

Chen, L.J., Zañartu, M., Cook, D.D., Mongeau, L., 2008. Effects of acoustic loading on the self-oscillations of a synthetic model of the vocal folds. In: Zolotarev, I., Horáček, J. (Eds.), Proceedings of the Ninth International Conference on Flow-Induced Vibrations, Prague, Czech Republic, pp. 1–6.

Childers, D.G., Hicks, D.M., Moore, G.P., Alsaka, Y.A., 1986. A model for vocal fold vibratory motion, contact area, and the electroglottogram. J. Acoust. Soc. Am. 80 (5), 1309–1320.

Cisonni, J., Van Hirtum, A., Pelorson, X., Lucero, J., 2011. The influence of geometrical and mechanical input parameters on theoretical models of phonation. Acta Acust. 97, 291–302.

Cook, D.D., Zañartu, M., 2010. Toward patient-specific vocal fold models: objective determination of lumped vocal fold model parameters from contiuum vocal fold models. In: Presented at the Seventh International Conference on Voice Physiology and Biomechanics, Madison, WI.

Cranen, B., Boves, L., 1987. On subglottal formant analysis. J. Acoust. Soc. Am. 81 (3), 734–746.

Cranen, B., Schroeter, J., 1995. Modeling a leaky glottis. J. Phon. 23, 165–177.

Dejonckere, P.H., Kob, M., 2009. Pathogenesis of vocal fold nodules: new insights from a modelling approach. Folia Phoniat. Logoped. 61, 171–179.

de Vries, M.P., Schutte, H.K., Verkerke, G.J., 1999. Determination of parameters for lumped parameter models of the vocal folds using a finite-element method approach. J. Acoust. Soc. Am. 106, 3620–3628.

de Vries, M.P., Schutte, H.K., Veldman, A.E.P., Verkerke, G.J., 2002. Glottal flow through a two-mass model: comparison of Navier–Stokes solutions with simplified models. J. Acoust. Soc. Am. 111, 1847–1853.

Döllinger, M., Hoppe, U., Hettlich, F., Lohscheller, J., Schuberth, S., Eysholdt, U., 2002. Vibration parameter extraction from endoscopic image series of the vocal folds. IEEE Trans. Biomed. Eng. 49, 773–781.

Drechsel, J.S., Thomson, S.L., 2008. Influence of supraglottal structures on the glottal jet exiting a two-layer synthetic, self-oscillating vocal fold model. J. Acoust. Soc. Am. 123, 4434–4445.

Dresel, C., Mergell, P., Hoppe, U., Eysholdt, U., 2006. An asymmetric smooth contour two-mass model for recurrent laryngeal nerve paralysis. Logoped. Phoniatr. Vocol. 31, 61–75.

Drioli, C., Avanzini, F., 2002. Hybrid parametric physiological glottal modeling wilth application to voice quality assessment. Med. Eng. Phys. 24, 453–460.

Dursun, G., Sataloff, R.T., Spiegel, J.R., Mandel, S., Heuer, R.J., Rosen, D.C., 1996. Superior laryngeal nerve paresis and paralysis. J. Voice 10, 206–211.

Erath, B.D., Plesniak, M.W., 2006a. The occurrence of the Coanda effect in pulsatile flow through static models of the human vocal folds. J. Acoust. Soc. Am. 120, 1000–1011.

Erath, B.D., Plesniak, M.W., 2006b. An investigation of bimodal jet trajectory in flow through scaled models of the human vocal folds. Exp. Fluids 40, 683–696.

Erath, B.D., Plesniak, M.W., 2006c. An investigation of jet trajectory in flow through scaled vocal fold models with asymmetric glottal passages. Exp. Fluids 41, 735–748.
Erath, B.D., Plesniak, M.W., 2010a. Viscous flow features in scaled-up physical models of normal and pathological phonation. Int. J. Heat Fluid Flow 31, 468–481.
Erath, B.D., Plesniak, M.W., 2010b. Impact of wall rotation on supraglottal jet instability in voiced speech. J. Acoust. Soc. Am. 129, EL64–EL70.
Erath, B.D., Plesniak, M.W., 2010c. An investigation of asymmetric flow features in a scaled-up model of the human vocal folds. Exp. Fluids 49, 131–146.
Erath, B.D., Plesniak, M.W., 2012. Three-dimensional laryngeal flow fields induced by a model vocal fold polyp. Int. J. Heat Fluid Flow 35, 93–101.
Erath, B.D., Peterson, S.D., Zañartu, M., Wodicka, G.R., Plesniak, M.W., 2011a. A theoretical model of the pressure field arising from asymmetric intraglottal flows applied to a two-mass model of the vocal folds. J. Acoust. Soc. Am. 130, 389–403.
Erath, B.D., Zañartu, M., Peterson, S.D., Plesniak, M.W., 2011b. Nonlinear vocal fold dynamics resulting from asymmetric fluid loading on a two-mass model of speech. Chaos 21, 033113.
Eriksson, L.J., 1980. Higher order mode effects in circular ducts and expansion chambers. J. Acoust. Soc. Am. 67 (2), 545–550.
Fant, G., 1960. In: Acoustic Theory of Speech Production, with Calculations Based on X-Ray Studies of Russian Articulations. Mouton and Co. N.V., The Hague.
Fant, G., Lin, Q., 1987. Glottal source – vocal tract acoustic interaction. STL-QPSR 28, 13–27.
Flanagan, J.L., 1972, second ed.. In: Speech Analysis, Synthesis, and Perception Springer-Verlag, New York.
Flanagan, J.L., Landgraf, L.L., 1968. Self-oscillating source for vocal tract synthesizers. IEEE Trans. Audio Electroacoust. AU-16, 57–64.
Fraile, R., Kob, M., Godino-Llorente, J.I., Saenz-Lechon, N., Osma-Ruiz, V.J., Gutierrez-Arriola, J.M., 2012. Physical simulation of laryngeal disorders using a multiple-mass vocal fold model. Biomed. Signal Process. & Control 7 (1), 65–78.
Fulcher, L.P., Scherer, R.C., Melnykov, A., Gateva, V., Limes, M.E., 2006. Negative coulomb damping, limit cycles, and self-oscillation of the vocal folds. Am. J. Phys. 74, 386–393.
Gay, T., Hirose, H., Strome, M., Sawashima, M., 1972. Electromyography of the instrinsic laryngeal muscles during phonation. Ann. Otolaryngol. 81, 401–409.
Goldberg, D., 1989. In: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA.
Gunter, H.E., 2003. A mechanical model of vocal-fold collision with high spatial and temporal resolution. J. Acoust. Soc. Am. 113, 994–1000.
Gupta, V., Wilson, T.A., Beavers, G.S., 1973. A model for vocal cord excitation. J. Acoust. Soc. Am. 54, 1607–1617.
Hanson, H.M., 1997. Glottal characteristics of female speakers: acoustic correlates. J. Acoust. Soc. Am. 101 (1), 466–481.
Hanson, D.G., Gerratt, B.R., Karin, R.R., Berke, G.S., 1988. Glottographic measures of vocal fold vibration: an examination of laryngeal paralysis. Laryngoscope 98, 541–549.
Hartman, D.E., Vishwanat, B., 1984. Spastic dysphonia and essential (voice) tremor treated with primidone. Arch. Otolaryngol. 110, 394–397.
Herzel, H., Knudsen, C., 1995. Bifurcations in a vocal fold model. Nonlinear Dyn. 7, 53–64.
Herzel, H., Berry, D., Titze, I.R., Steinecke, I., 1995. Nonlinear dynamics of the voice: signal analysis and biomechanical modeling. Chaos 5 (1), 30–34.
Hess, M.M., Verdolini, K., Bierhals, W., Mansmann, U., Gross, M., 1998. Endolaryngeal contact pressures. J. Voice 12, 50–67.
Hillman, R.E., Holmberg, E.B., Perkell, J.S., Walsh, M., Vaughan, C., 1989. Objective assessment of vocal hyperfunction: an experimental framework and initial results. J. Speech Hear. Res. 32, 373–392.
Hirano, M., 1974. Morphological structure of the vocal cord as a vibrator and its variations. Folia Phoniatr. 26, 89–94.
Hirano, M., 1975. Phonosurgery – basic and clinical investigations. Otologia 21, 239–240.
Hirano, M., 1977. Sturcture and vibratory behavior of the vocal folds. In: Sawashima, M., Cooper, F.S. (Eds.), . In: Dynamic Aspects of Speech Production. University of Tokyo, Tokyo, Japan, pp. 13–27.
Hirano, M., Vennard, W., Ohala, J., 1970. Regulation of register, pitch and intensity of voice: an electromyographic investigation of intrinsic laryngeal muscles. Folia Phoniatr. 22, 1–20.
Hirano, M., Kurita, S., Makashima, T., 1981. The structure of the vocal folds. In: Stevens, K., Hirano, M. (Eds.), . In: Vocal Fold Physiology. University of Tokyo Press, Tokyo, Japan, pp. 33–41.
Hirano, M., Kurita, S., Nakashima, T., 1983. Growth, development and aging of human vocal folds. In: Bless, D.M., Abbs, J.H. (Eds.), . In: Vocal Fold Physiology: Contemporary Research and Clinical Issues. College Hill Press, San Diego, CA, pp. 22–43.
Hirschberg, A., Pelorson, X., Hofmans, G.C.J., van Hassel, R.R., Wijnands, A.P.J., 1996. Starting transient of the flow through an in-vitro model of the vocal folds. In: Davis, P.J., Fletcher, N.H. (Eds.), Vocal Fold Physiology: Controlling Complexity and Chaos, Singular, San Diego, CA, pp. 31–46.
Ho, J.C., Zañartu, M., Wodicka, G.R., 2011. An anatomically based time-domain acoustic model of the subglottal system for speech production. J. Acoust. Soc. Am. 129 (3), 1531–1547.
Hofmans, G.C.J., Groot, G., Ranucci, M., Graziani, G., Hirschberg, A., 2003. Unsteady flow through in-vitro models of the glottis. J. Acoust. Soc. Am. 113, 1658–1675.
Holmberg, E.B., Hillman, R.E., Perkell, J.S., 1988. Glottal air-flow and transglottal air-pressure measurements for male and female speakers in soft normal and loud voice. J. Acoust. Soc. Am. 84, 511–529.
Honda, K., Takemoto, H., Kitamura, T., Fujita, S., Takano, S., 2004. Exploring human speech production mechanisms by MRI. IEICE Info. & Systems E87-D, 1050–1058.
Horáček, J., Šidlof, P., Švec, J.G., 2005. Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces. J. Fluid. Struct. 20, 853–869.
Ishizaka, K., Flanagan, J.L., 1972. Synthesis of voice sounds from a two-mass model of the vocal cords. Bell Systems Tech. J. 51, 1233–1268.
Ishizaka, K., Isshiki, N., 1976. Computer simulation of pathological vocal-cord vibration. J. Acoust. Soc. Am. 60 (5), 1193–1198.
Ishizaka, K., Kaneko, T., 1968. On equivalent mechanical constants of the vocal cords. J. Acoust. Soc. Jpn. 24, 312–313.
Ishizaka, K., Matsuidara, M., Kaneko, T., 1976. Input acoustic-impedance measurement of subglottal system. J. Acoust. Soc. Am. 60, 190–197.
Isshiki, N., 1977. In: Functional Surgery of the Larynx. Kyoto University.
Jiang, J.J., Titze, I.R., 1994. Measurement of vocal fold intraglottal pressure and impact stress. J. Voice 8 (2), 132–144.
Jiang, J.J., Lin, E., Hanson, D.G., 2000. Vocal fold physiology. Otolaryngol. Clin. North Am. 33, 699–717.
Jiang, J.J., Zhang, Y., Stern, J., 2001. Modeling of chaotic vibrations in symmetric vocal folds. J. Acoust. Soc. Am. 110 (4), 2120–2128.
Johns, M.M., 2003. Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. Otolarynol. Head Neck Surg. 11, 456–461.
Kaneko, T., Asano, H., Naito, J., Kobayashi, K., Kitamura, T., 1972. Biomechanics of the vocal cords – on damping ratio. J. Jpn. Bronchoesophagol. Soc. 25, 133–138.
Kelly, J.L., Lochbaum, C.C., 1973. Speech synthesis. In: Flanagan, J.L., Rabiner, L.R. (Eds.), Speech Synthesis, Dowden, Stroudsberg, PA, pp. 1–512.
Khosla, S.M., Murugappan, S., Gutmark, E.J., Scherer, R.C., 2007. Vortical flow field during phonation in an excised canine larynx model. Ann. Otol. Rhinol. Laryngol. 116, 217–228.
Khosla, S.M., Murugappan, S., Lakhamaraju, R., Gutmark, E.J., 2008a. Using particle imaging velocimetry to measure anterior–posterior velocity gradients in excised canine larynx model. Ann. Otol. Rhinol. Laryngol. 117, 134–144.
Khosla, S.M., Murugappan, S., Gutmark, E.J., 2008b. What can vortices tell us about vocal fold vibration and voice production. Curr. Opin. Otolaryngol. Head Neck Surg. 16, 183–187.

Klatt, D.H., Klatt, L.C., 1990. Analysis synthesis and perception of voice quality variations among male and female talkers. J. Acoust. Soc. Am. 87 (2), 820–856.

Kob, M., 2002. Physical modeling of the singing voice, Ph.D. thesis, Unversity of Technology, Aachen, Berlin.

Koizumi, T., Taniguchi, S., Hiromitsu, S., 1987. Two-mass models of the vocal cords for natural voice synthesis. J. Acoust. Soc. Am. 82, 1179–1192.

Koizumi, T., Taniguchi, S., Itakura, F., 1993. An analysis-by-synthesis approach to estimation of vocal cord polyp features. Laryngoscope 103, 1035–1042.

Krane, M.H., Barry, M., Wei, T., 2007. Unsteady behavior of flow in a scaled-up vocal folds model. J. Acoust. Soc. Am. 122, 3659–3670.

Kröger, B.J., Birkholz, P., Neuschaefer-Rube, C., 2007. Ein neuronales modell zur sensomotorischen entwicklung des sprechens. Laryngo-Rhino-Otologie 86, 365–370.

Kröger, B.J., Birkholz, P., Neuschaefer-Rube, C., 2011a. Towards an articulation-based developmental robotics approach for word processing in face-to-face communication. PALADYN J. Behav. Robot. 2 (2), 82–93.

Kröger, B.J., Birkholz, P., Kannampuzha, J., Kaufmann, E., Mittelberg, I., 2011b. Movements and holds in fluent sentence production of american sign language: the action-based approach. Cogn. Comput. 3, 449–465.

Kröger, B.J., Birkholz, P., Neuschaefer-Rube, C., 2011c. Categorical perception of consonants and vowels: evidence from a neurophonetic model of speech production and perception. In: Esposito, A., Esposito, A.M., Martone, R., Müller, V.C., Scarpetta, G. (Eds.), . In: Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues. Springer, Berlin, pp. 354–361.

Kuo, J., 1998. Voice source modeling and analysis of speakers with vocal-fold nodules, Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology.

Li, S., Wan, M., Wang, S., 2007. The effects of the false vocal fold gaps in a model of the larynx on pressures distributions and flows. In: Proceedings of the First International Conference on Digital human modeling, ICDHM'07. Springer-Verlag, Berlin, Heidelberg, pp. 147–156.

Liljencrants, J., 1985. Speech synthesis with a reflection-type line analog, Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden.

Liljencrants, J., 1991. A translating and rotating mass model of the vocal folds. STL-QPSR 32, 1–18.

Lo, C.Y., Kwok, K.F., Yuen, P.W., 1960. A prospective evaluation of recurrent laryngeal nerve paralysis during thyroidectomy. Arch. of Surg. 135 (2), 204–207 (Chicago, Ill. : 1960).

Lofqvist, A., Koenig, L.L., McGowan, R.S., 1995. Vocal tract aerodynamics in /aCa/ utterances: measurements. Speech Comm. 16, 49–66.

Logemann, J.A., Fisher, H.B., Boshes, B., Blonsky, E.R., 1978. Frequency and coocurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. J. Speech Hear. Disord. 43, 47–58.

Lohscheller, J., Toy, H., Rosanowski, F., Eyshold, U., Döllinger, M., 2007. Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. Med. Image Anal. 4, 400–413.

Lohscheller, J., Eyshold, U., Toy, H., Döllinger, M., 2009. Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-*D* diagrams for visualizing and analyzing the underlying laryngeal dynamics. IEEE Trans. Med. Imaging 27, 300–309.

Lous, N.J.C., Hofmans, G.C.J., Veldhuis, R.N.J., Hirschberg, A., 1998. A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design. Acta Acust. 84, 1135–1150.

Lowell, S.Y., Story, B.H., 2006. Simulated effects of criothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. J. Acoust. Soc. Am. 120, 386–397.

Lucero, J.C., 2005. Oscillation hysteresis in a two-mass model of the vocal folds. J. Sound Vib. 282, 1247–1254.

Lucero, J.C., Koenig, L.L., 2005. Simulations of temporal patterns of oral airflow in men and woman using a two-mass model of the vocal folds under dynamic control. J. Acoust. Soc. Am. 117 (3), 1362–1372.

Luo, H., Mittal, R., Bielamowicz, S., 2009. Analysis of flow-structure interaction in the larynx during phonation using an immersed-boundary method. J. Acoust. Soc. Am. 126, 816–824.

Maeda, S., 1982. A digital simulation method of the vocal-tract system. Speech Comm. 1, 199–229.

Massey, E.W., Paulson, G.W., 1985. Essential vocal tremor: clinical characteristics and response to therapy. South. Med. J. 78, 316–317.

McGowan, R.S., 1988. An aeroacoustic approach to phonation. J. Acoust. Soc. Am. 83, 696–704.

McGowan, R.S., Howe, M.S., 2010. Comments on single-mass models of vocal fold vibration. J. Acoust. Soc. Am. 127, EL215–EL221.

McGowan, R.S., Koenig, L.L., Lofqvist, A., 1995. Vocal tract aerodynamics in /aCa/ utterances: simulations. Speech Comm. 16, 67–88.

Mehta, D.D., Zañartu, M., Quatieri, T.F., Deliyski, D.D., Hillman, R.E., 2011. Investigating acoustic correlates of human vocal fold phase asymmetry through mathematical modeling and laryngeal high-speed videoendoscopy. J. Acoust. Soc. Am. 130, 3999–4009.

Mergell, P., Herzel, H., 1997. Modelling biphonation – the role of the vocal tract. Speech Comm. 22, 141–154.

Miller, D.G., Schutte, H.K., 2005. 'Mixing' the registers: glottal source or vocal tract? Folia Phoniatr. Logop. 57, 278–291.

Mittal, R., Erath, B.D., Plesniak, M.W., 2013. Fluid-dynamics of human phonation and speech. Ann. Rev. Fluid Mech. 45, 437–467.

Mokhtari, P., Takemoto, H., Kitamura, T., 2008. Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches. Speech Comm. 50, 179–190.

Mongeau, L., Franchek, N., Coker, C.H., Kubli, R.A., 1997. Characteristics of a pulsating jet through a small modulated orifice, with application to voice production. J. Acoust. Soc. Am. 102, 1121–1133.

Neubauer, J., Zhang, Z., Miraghie, R., Berry, D., 2007. Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds. J. Acoust. Soc. Am. 121, 1102–1118.

Park, J.B., Mongeau, L., 2007. Instantaneous orifice discharge coefficient of a physical driven model of the human larynx. J. Acoust. Soc. Am. 121, 442–455.

Park, J.B., Mongeau, L., 2008. Experimental investigation of the influence of a posterior gap on glottal flow and sound. J. Acoust. Soc. Am. 124, 1171–1179.

Pelorson, X., Hirschberg, A., Wijnands, A.P.J., Bailliet, H.M.A., 1994. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. J. Acoust. Soc. Am. 96, 3416–3431.

Pelorson, X., Hirschberg, A., Wijnands, A.P.J., Bailliet, H.M.A., 1995. Description of the flow through in-vitro models of the glottis during phonation. Acta Acust. 3, 191–202.

Pelorson, X., Vescovi, C., Castelli, E., Hirschberg, A., Wijnands, A.P.J., Bailliet, H.M.A., 1996. Description of the flow through in-vitro models of the glottis during phonation: application to voiced sound synthesis. Acta Acust. 82, 358–361.

Perlman, A.L., 1985. A technique for measuring the elastic properties of vocal fold tissue, Ph.D. thesis, The University of Iowa, Iowa City, IA.

Qin, X., Wang, S., Wan, M., 2009. Improving reliability and accuracy of vibration parameters of vocal folds based on high-speed video and electroglottography. IEEE Trans. Biomed. Eng. 56, 1744–1754.

Qiu, Q.J., Yu, Q.L., Jiang, J.Y., Xu, F., Cai, Y.J., 2002. A combined vocal fold model. In: Tan, J.B., Wen, X.F. (Eds.), . In: Proceedings of the Second International Symposium on Instrumentation Science and Technology, vol. 3. Harbin Institute Technology Publishers, Heilongjiang, China, pp. 541–546.

Rothenberg, M., 1973. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J. Acoust. Soc. Am. 53 (6), 1632–1645.

Rothenberg, M., 1981. An interactive model for the voice source. STL-QPSR 4, 1–17.

Rothenberg, M., 1984. Source-tract acoustic interaction in breathy voice. In: Titze, I.R., Scherer, R.C. (Eds.), . In: Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control. The Denver Center for the Performing Arts, pp. 465–481.

Rothenberg, M., Zahorian, S., 1977. Nonlinear inverse filtering technique for estimating the glottal-area waveform. J. Acoust. Soc. Am. 61 (4), 1063–1070.

Rupitsch, S.J., Ilg, J., Sutor, A., Lerch, R., Döllinger, M., 2011. Simulation based estimation of dynamic mechanical properties for viscoelastic materials used for vocal fold models. J. Sound Vib. 330, 4447–4459.

Ruty, N., Pelorson, X., Van Hirtum, A., Lopez-Arteaga, I., Hirschberg, A., 1997. An in vitro setup to test the relevance and accuracy of low-order vocal fold models. J. Acoust. Soc. Am. 121, 479–490.

Scherer, R.C., Torkaman, S., Kucinschi, B.R., Afjeh, A.A., 2010. Intraglottal pressures in a three-dimensional model with non-rectangular shape. J. Acoust. Soc. Am. 128, 828–838.

Schlichting, H., 1968, sixth ed.. In: Boundary Layer Theory McGraw-Hill, New York.

Schroete, J., 2008. Basic principles of speech synthesis. In: Benesty, J., Sondhi, M.M., Huang, Y. (Eds.), Handbook of Speech Processing. Springer-Verlag, Berlin, Heidelberg, pp. 413–428.

Schwarz, R., Hoppe, U., Schuster, M., Wurzbacher, T., Eysholdt, U., Lohscheller, J., 2006. Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recording and inversion of a biomechanical model. IEEE Trans. Biomed. Eng. 53, 1099–1108.

Schwarz, R., Döllinger, M., Wurzbacher, T., Eysholdt, U., Lohscheller, J., 2008. Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. J. Acoust. Soc. Am. 123, 2717–2732.

Sciamarella, D., d'Alessandro, C., 2003. Reproducing laryngeal mechanisms with a two-mass model. In: Eurospeech, no. 2, Geneva, Switzerland.

Sciamarella, D., d'Alessandro, C., 2004. On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters. Acta Acust. 90, 746–761.

Sciamarella, D., Artana, G., 2009. A water hammer analysis of pressure and flow in the voice production system. Speech Comm., 344–351.

Sercarz, J.A., Berke, G.S., Ming, Y., Gerratt, B.R., Natividad, M., 1992. Videostroboscopy of human vocal fold paralysis. Ann. Otol. Rhinol. Laryngol. 101, 567–577.

Smith, M.E., Berke, G.S., Gerratt, B.R., Kreiman, J., 1992. Laryngeal paralyses: theoretical considerations and effects on laryngeal vibration. J. Speech Hear. Res. 35, 545–554.

Sobey, I.J., 1983. The occurrence of separation in oscillatory flow. J. Fluid Mech. 134, 247–257.

Sommer, D., Erath, B.D., Zañartu, M., Peterson, S.D., 2012. Corrected contact dynamics for the Steinecke and Herzel asymmetric two-mass model of the vocal folds. J. Acoust. Soc. Am. 132, EL271–EL276.

Sommer, D., Erath, B.D., Zañartu, M., Peterson, S.D., 2013. The impact of glottal area discontinuities on block-type vocal fold models with asymmetric tissue properties. J. Acoust. Soc. Am., 133 (3), EL214–EL220.

Steinecke, I., Herzel, H., 1995. Bifurcations in an asymmetric vocal-fold model. J. Acoust. Soc. Am. 97, 1874–1884.

Stevens, K.N., 1998. In: Acoustic Phonetics. The MIT press, Cambridge, MA.

Stevens, K.N., House, A.S., 1955. Development of a quantitative description of vowel articulation. J. Acoust. Soc. Am. 27 (3), 484–493.

Story, B.H., 1995. Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract, Ph.D. thesis, The University of Iowa, Iowa City, IA.

Story, B.H., 2002. An overview of the physiology , physics and modeling of the sound source for vowels. Acoust. Sci. & Tech. 4, 195–206.

Story, B.H., 2005. A parametric model of the vocal tract area function for vowel and consonant simulation. J. Acoust. Soc. Am. 117 (5), 3231–3254.

Story, B.H., 2007a. Time dependence of vocal tract modes during production of vowels and vowel sequences. J. Acoust. Soc. Am. 121 (6), 3770–3789.

Story, B.H., 2007b. Modification of emotional speech and voice quality based on changes to the vocal tract structure, first ed.. In: Izdebski, K. (Ed.), . In: Emotions in the Human Voice, vol. 1 Plural Publishing, pp. 123–136.

Story, B.H., 2008. Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. J. Acoust. Soc. Am. 123 (1), 327–335.

Story, B.H., 2009. A possible role of nonlinear source-filter interaction in simulation of childlike speech. J. Acoust. Soc. Am. 125 (4), 2637.

Story, B.H., Titze, I.R., 1995. Voice simulation with a body-cover model of the vocal folds. J. Acoust. Soc. Am. 97, 1249–1260.

Story, B.H., Titze, I.R., Hoffman, E.A., 1996. Vocal tract area functions from magnetic resonance imaging. J. Acoust. Soc. Am. 100 (1), 537–554.

Story, B.H., Bunton, K., shape, Relation of vocal tract, transitions, formant, 2010. and stop consonant identification. J. Speech Lang. Hear. Res. 53, 1514–1528.

Takemoto, H., Honda, K., Masaki, S., Shimada, Y., Fujimoto, I., 2006. Measurement of temporal changes in vocal tract area function from 3*D* cine-MRI data. J. Acoust. Soc. Am. 119, 1037–1049.

Tao, C., Jiang, J.J., 2007a. Extracting physiologically relevant parameters of vocal folds from high-speed video image series. IEEE Trans. Biomed. Eng. 54, 794–801.

Tao, C., Jiang, J.J., 2007b. Mechanical stress during phonation in a self-oscillating finite-element vocal fold model. J. Biomech. 40, 2191–2198.

Tao, C., Jiang, J.J., 2008. Chaotic component obscured by strong periodicity in voice production system. Phys. Rev. E 77, 061922.

Tao, C., Zhang, Y., Hottinger, D.G., Jiang, J.J., 2007. Asymmetric airflow and vibration induced by the Coanda effect in a symmetric model of the vocal folds. J. Acoust. Soc. Am. 122, 2270–2278.

Titze, I.R., 1973. The human vocal cords: a mathematical model, Part I. Phonetica 28, 129–170.

Titze, I.R., 1974. The human vocal cords: a mathematical model Part II. Phonetica 29, 1–21.

Titze, I.R., 1979. The concept of muscular isometrics for optimizing vocal intensity and efficiency. In: Lawrence, V.L. (Ed.), . In: Transcripts of the Ninth Symposium: Care of the Professional Voice. The Voice Foundation, New York, pp. 23–28.

Titze, I.R., 1988. The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. 83, 1536–1552.

Titze, I.R., 1994a. Mechanical stress in phonation. J. Voice 8, 99–105.

Titze, I.R., 1994b. In: Principles of Voice Production. Prentice Hall.

Titze, I.R., 2002. Regulating glottal airflow in phonation: application of the maximum power transfer theorem to a low dimensional phonation model. J. Acoust. Soc. Am. 111, 367–376.

Titze, I.R., 2004. A theoretical study of F0–F1 interaction with application to resonant speaking and singing voice. J. Voice 18 (3), 292–298.

Titze, I.R., 2006. In: The Myoelastic Aerodynamic Theory of Phonation. The National Center for Voice and Speech.

Titze, I.R., 2008. Nonlinear source-filter coupling in phonation: theory. J. Acoust. Soc. Am. 123, 2733–2749.

Titze, I.R., Story, B.H., 1997. Acoustic interactions of the voice source with the lower vocal tract. J. Acoust. Soc. Am. 101 (4), 2234–2243.

Titze, I.R., Story, B.H., 2002. Rules for controlling low-dimensional vocal fold models with muscle activation. J. Acoust. Soc. Am. 112, 1064–1076.

Titze, I.R., Worley, A.S., 2009. Modeling source-filter interaction in belting and high-pitched operatic male singing. J. Acoust. Soc. Am. 126 (3), 1530–1540.

Titze, I.R., Riede, T., Popolo, P., 2008. Nonlinear source-filter coupling in phonation: vocal exercises. J. Acoust. Soc. Am. 123 (4), 1902–1915.

Tokuda, I.T., Horáček, J., Švec, J.G., Herzel, H., 2007. Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. J. Acoust. Soc. Am. 122 (1), 519–531.

Tokuda, I.T., Horáček, J.J., Švec, J.G., Herzel, H., 2008. Bifurcations and chaos in register transitions of excised larynx experiments. Chaos 18, 013102.

Tokuda, I., Zemke, M., Kob, M., 2010. Biomechanical modeling of register transitions and the role of vocal tract resonators. J. Acoust. Soc. Am. 127, 1528–1536.

Triep, M., Brücker, C., 2010. Three-dimensional nature of the glottal jet. J. Acoust. Soc. Am. 127, 1537–1547.

Triep, M., Brücker, C., Schröder, W., 2005. High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds. Exp. Fluids 39, 232–245.

van den Berg, J., 1958. Myoelastic-aerodynamic theory of voice production. J. Voice 1, 227–244.

van den Berg, J., 1968. Register problems. Ann. N.Y. Acad. Sci. 155 (1), 129–134.

van den Berg, J., Tan, T.S., 1959. Results of experiments with human larynxes. Pract. Otorhinolaryngol. 21, 425–450.

Vilain, C.E., Pelorson, X., Fraysse, C., Deverge, M., Hirschberg, A., Willems, J., 2004. Experimental validation of a quasi-steady theory for the flow through the glottis. J. Sound Vib. 276, 475–490.

Voigt, D., Döllinger, M., Eysholdt, U., Yang, A., Gürlek, E., Lohscheller, J., 2010. Objective detection and quantification of mucosal wave propagation. J. Acoust. Soc. Am. 128, EL347–EL353.

Wegel, R.L., 1930. Theory of vibration of the larynx. J. Acoust. Soc. Am. 1, 1–21.

Wodicka, G.R., Stevens, K.N., Golub, H.L., Cravalho, E.G., Shannon, D.C., 1989. A model of acoustic transmission in the respiratory system. IEEE Trans. Biomed. Eng. 36 (9), 925–934.

Wong, D., Ito, M.R., Cox, N.B., Titze, I.R., 1991. Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. J. Acoust. Soc. Am. 89, 383–394.

Wurzbacher, T., Schwarz, R., Hoppe, U., Eysholdt, U., Lohscheller, J., 2004. Non-stationary modeling of vocal fold vibrations during a pitch raise. In: International Conference on Voice Physiology and Biomechanics, Marseille, France.

Wurzbacher, T., Schwarz, R., Döllinger, M., Hoppe, U., Eysholdt, U., Lohscheller, J., 2006. Model-based classification of nonstationary vocal fold vibrations. J. Acoust. Soc. Am. 120, 1012–1027.

Wurzbacher, T., Döllinger, M., Schwarz, R., Hoppe, U., Eysholdt, U., Lohscheller, J., 2008. Spatiotemporal classification of vocal fold dynamics by a multimass model comprising time-dependent parameters. J. Acoust. Soc. Am. 123, 12324–12334.

Xue, Q., Mittal, R., Zheng, X., Bielamowicz, S., 2010. A computational study of the effect of vocal-fold asymmetry on phonation. J. Acoust. Soc. Am. 128 (2), 818–827.

Yamana, T., Kitajima, K., 2000. Laryngeal closure pressure during phonation in humans. J. Voice 14, 1–7.

Yang, A., Lohscheller, J., Berry, D.A., Becker, S., Eysholdt, U., Voigt, D., Döllinger, M., 2010. Biomechanical modeling of the three-dimensional aspects of human vocal fold dynamics. J. Acoust. Soc. Am. 127 (2), 1014–1031.

Yang, A., Stingl, M., Berry, D.A., Lohscheller, J., Voigt, D., Eysholdt, U., Döllinger, M., 2011. Computation of physiological human vocal fold parameters by mathematical optimzation of a biomechanical model. J. Acoust. Soc. Am. 130 (2), 948–964.

Yang, A., Berry, D.A., Kaltenbacher, M., Döllinger, M., 2012. Three-dimensional biomechanical properties of human vocal folds: parameter optimization of a numerical model to match in vitro dynamics. J. Acoust. Soc. Am. 131 (2), 1378–1390.

Yumoto, E., Minoda, R., Hyodo, M., Yamagata, T., 2002. Causes of recurrent laryngeal nerve paralysis. Auris. Nasus. Larynx 29 (1), 41–45.

Zañartu, M., 2006. Influence of acoustic loading on the flow-induced oscillations of single mass models of the human larynx, Master's thesis, School of Electrical and Computer Engineering, Purdue University.

Zañartu, M., 2010. Acoustic coupling in phonation its effect on inverse filtering of oral airflow neck surface acceleration, Ph.D. thesis, Purdue University, West Lafayette, IN.

Zañartu, M., Mongeau, L., Wodicka, G.R., 2007. Influence of acoustic loading on an effective single-mass model of the vocal folds. J. Acoust. Soc. Am. 121, 1119–1129.

Zañartu, M., Mehta, D.D., Ho, J.C., Hillman, R.E., Wodicka, G.R., 2011. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. J. Acoust. Soc. Am. 129, 326–339.

Zhang, Y., Jiang, J.J., 2004. Chaotic vibrations of a vocal fold model with a unilateral polyp. J. Acoust. Soc. Am. 115, 1266–1269.

Zhang, Y., Jiang, J.J., 2008a. Asymmetric spatiotemporal chaos induced by a polypoid mass in the excised canine larynx. Chaos 18, 043102.

Zhang, Y., Jiang, J.J., 2008b. Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations. J. Sound Vib. 316, 248–262.

Zhang, C., Zhao, W., Frankel, S.H., Mongeau, L., 2002. Computational aeroacoustics of phonation Part II: effects of flow parameters and ventricular folds. J. Acoust. Soc. Am. 112, 2147–2154.

Zhang, Y., McGilliagan, C., Zhou, L., Vig, M., Jiang, J.J., 2004. Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. J. Acoust. Soc. Am. 115, 2270–2277.

Zhang, Y., Jiang, J.J., Rahn, D.A., 2005. Studying vocal fold vibrations in Parkinson's disease with a nonlinear model. Chaos 15, 033903.

Zhang, Z., Mongeau, L.G., 2006. Broadband sound generation by confined pulsating jets in a mechanical model of the human larynx. J. Acoust. Soc. Am. 119 (6), 3995–4005.

Zhang, Z., Neubauer, J., Berry, D.A., 2006. The influence of subglottal acoustics on laboratory models of phonation. J. Acoust. Soc. Am. 120, 1558–1569.

Zhao, W., Zhang, C., Frankel, S.H., Mongeau, L., 2002. Computational aeroacoustics of phonation, Part I: computational methods and sound generation mechanisms. J. Acoust. Soc. Am. 112 (5), 2134–2146.

Zheng, X., Bielamowicz, S., Luo, H., Mittal, R., 2009. A computational study of the effect of false vocal folds on glottal flow and vocal fold vibration during phonation. Ann. Biomed. Eng. 37, 625–642.

Zheng, X., Mittal, R., Xue, Q., Bielamowicz, S., 2011. Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model. J. Acoust. Soc. Am. 130, 404–415.

Zhuang, P., Sprecher, A.J., Hoffman, M.R., Zhang, Y., Fourakis, M., Jiang, J.J., Wei, C.S., 2009. Phonation threshold flow measurements in normal and pathological phonation. Laryngoscope 119, 811–815.