Acoust. Sci. & Tech. **23**, 4 (2002)

TUTORIAL

An overview of the physiology, physics and modeling of the sound source for vowels

Brad H. Story*

Department of Speech and Hearing Sciences, University of Arizona, P. O. Box 210071 Tucson, AZ 85721–0071 USA

Abstract: The vibration of the vocal folds produces the primary sound source for vowels. This paper first reviews vocal fold anatomy and the kinematics associated with typical vibratory motion. A brief historical background is then presented on the basic physics of vocal fold vibration and various efforts directed at mathematical modeling of the vocal folds. Finally, a low-dimensional model is used to simulate the vocal fold vibration under various conditions of vocal tract loading. In particular, a ''no-tract'' case is compared to two cases in which the voice source is coupled to vocal tract area functions representing the vowels /i/ and /ɑ/, respectively.

Keywords: Vocal folds, Voice, Mucosal wave, Source-tract interaction

PACS number: 43.70.Aj, 43.70.Bk

1. INTRODUCTION

Vowel production is typically discussed in terms of vocal tract shape and corresponding formant (resonance) frequencies. However, the manifestation of a particular vowel as a *sound* depends on the acoustic excitation of the vocal tract cavities by a sound source. For vowels, the sound source is most often generated by the vibrating vocal folds that convert the steady (DC) airflow from the lower respiratory system into a periodic train of flow pulses. These pulses, referred to as the glottal flow, are then acoustically filtered by the vocal tract resonances. This process effectively redistributes the amplitudes of the frequency components of the source signal, resulting in a vowel sound. Figure 1 illustrates the ''source-filter'' representation [1] in both the time and frequency domains where the source is shown on the left side, the vocal tract filter in the center, and the resulting output (at the lips) on the right. In the time domain (top row of figures) the filter is shown as a vocal tract area function suggesting that the glottal flow propagates through it and is transformed into the output pressure (in a strict mathematical sense the source would be convolved with the impulse response of the area function). The frequency domain representation suggests that the output spectrum is the product of the glottal flow spectrum and the frequency response of the vocal tract. In either case, it is apparent that the vibration of the vocal folds provides an harmonically-rich acoustic excitation that the vocal tract resonances can ''mold'' into the respective vowel sounds.

While the sound source (glottal flow) produced by vocal fold vibration may appear to be relatively simple, the mechanisms responsible for inducing and sustaining the vibration involve complexities of anatomy, muscular control, nonlinear flow-induced oscillation, and the interaction of vocal tract pressures with both the glottal flow and the movement of the vocal folds themselves. To show how some of these complexities are integrated with vowel production this paper has three aims: (1) to review the anatomical structure of vocal folds and their vibratory kinematics, (2) to present a brief overview of vocal fold modeling, and (3) to use a low-dimensional model to illustrate some basic mechanisms of vocal fold oscillation, including the acoustic interaction with the vocal tract.

2. VOCAL FOLD STRUCTURE AND VIBRATORY KINEMATICS

The vocal folds are soft tissue structures contained within the cartilaginous framework of the larynx. Their location in the neck and their ability to abduct (move apart) during respiration and to adduct (move together) during phonation makes the vocal folds the point of division between the subglottal and supraglottal airways. Four views of the larynx and vocal folds are shown in Fig. 2. The first is a midsagittal sketch of the head and neck indicating the location of the larynx (Fig. 2a) and the second is a schematic drawing showing how the vocal folds are located at the superior end of the trachea and connected to the lower portion of the pharynx (Fig. 2b). During

*e-mail: bstory@u.arizona.edu

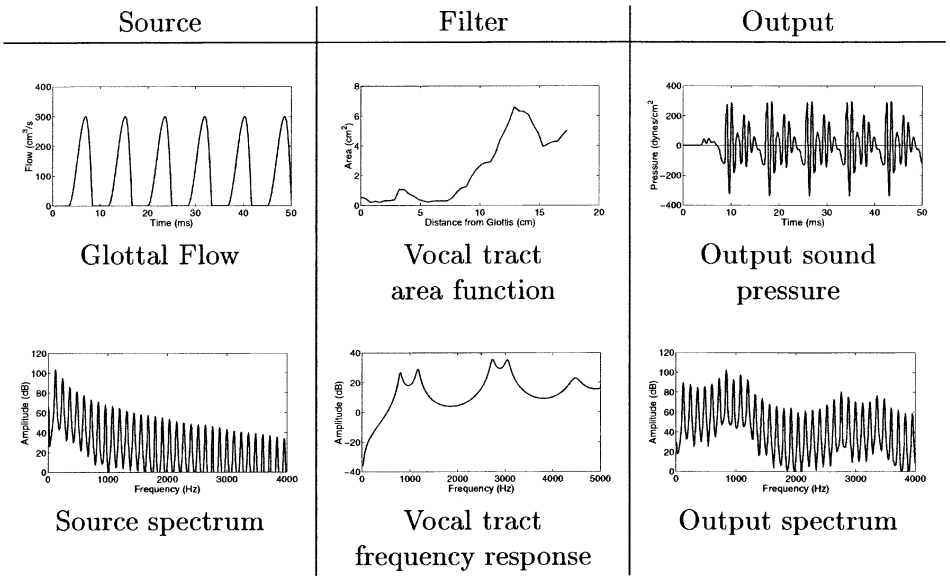


Fig. 1 Illustration of the source-filter representation of vowels. The upper row shows the source waveform, a vocal tract area function, and the output waveform, all in the time domain. The second row shows the frequency domain representation of each of the quantities in the first row.

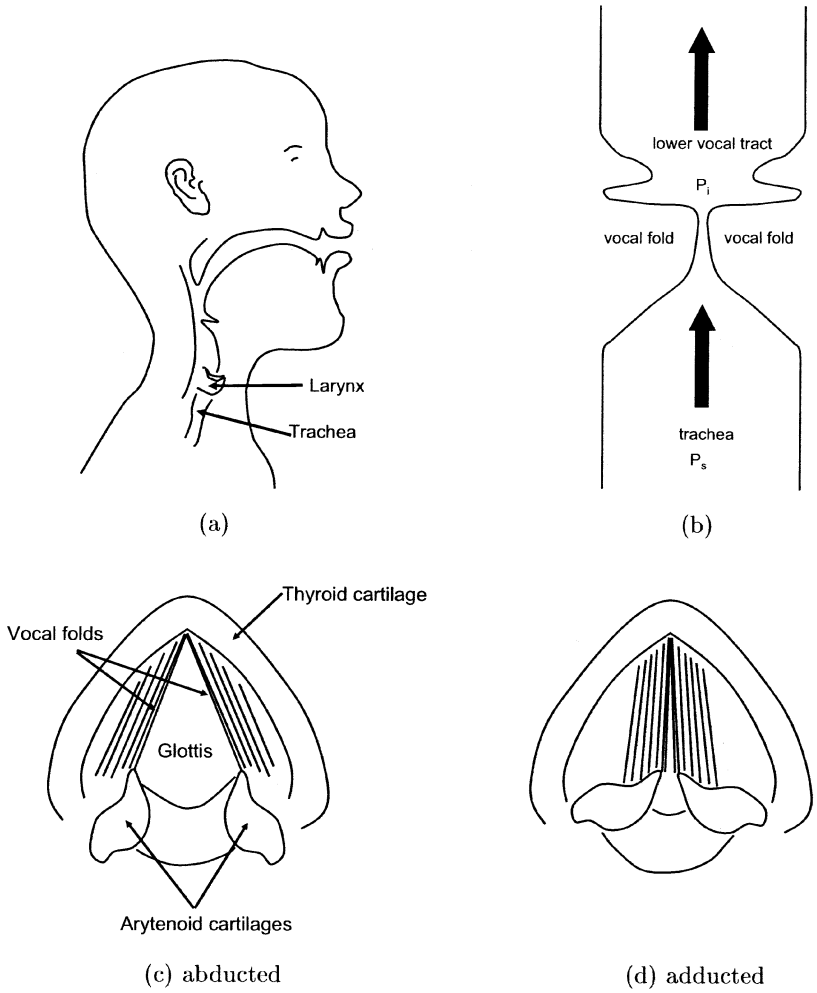


Fig. 2 Diagrams of the larynx and vocal folds. (a) Midsagittal view of the head and neck indicating the location of the larynx and tracheas, (b) schematic diagram showing how the vocal folds are located at the superior end of the trachea and coupled to the supraglottal vocal tract, (c) superior view of larynx when the vocal folds are abducted, as during respiration, and (d) superior view of larynx when the vocal folds are adducted, as during phonation.

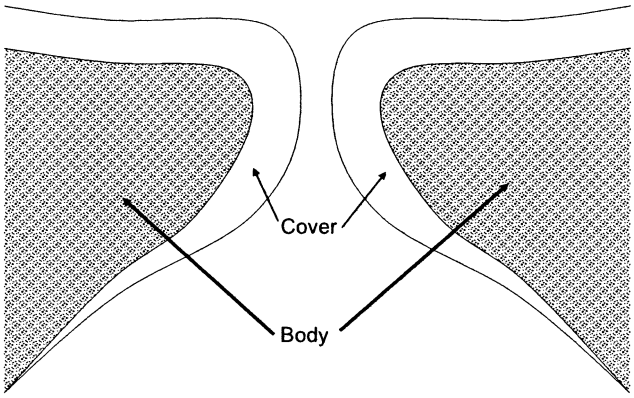


Fig. 3 Division of the vocal fold into the cover and body portions (based on Hirano [2]).

normal respiration the vocal folds are widely separated (*abducted*, Fig. 2c), but prior to and during phonation they are closely approximated (*adducted*, Fig. 2d).

The structure of the vocal folds is often described by the cover-body concept [2]. It suggests that the vocal folds can be roughly divided into two tissue layers (Fig. 3) with different mechanical properties. The cover layer is comprised of pliable, non-contractile mucosal tissue that serves as a sheath around the body-layer. In contrast, the body layer consists of muscle fibers (thyroarytenoid) and some ligamentous tissue. Other more detailed schemes have been used to describe the layers of tissue in the vocal folds (see Titze [3] for a summary), but the cover-body scheme is especially convenient for purposes of studying and modeling the vibratory characteristics of the vocal folds.

Observations of the vibrating vocal folds with stroboscopic techniques or high speed film/video have shown that the cover tissue supports a surface wave that propagates from the inferior portions of the vocal fold cover to the superior (a summary of vocal fold observation techniques can be found in Moore [4]). This wave is typically referred to as the *mucosal wave* or *vertical phase difference* and is most clearly observed in the coronal plane. Vocal fold motion in this plane (along the axis of airflow) was first viewed with X-ray techniques [5] and more recently with observations of excised hemilaryngeal vibration using videostroboscopy [6] and high-speed video [7]. It is also noteworthy that Chiba and Kajiyama [8] used a stroboscopic technique to obtain motion pictures of the vocal folds during vibration. They did not specifically mention the mucosal wave but perhaps allude to it in their observation on page 24 that, ''. . . it seems that every small portion constituting the vocal chords is set in elliptical motion in vertical cross-section. . . In this way, the vocal chords as a whole come to be set in wave motion.''

Figure 4 shows an idealized cycle of vocal fold vibration in the coronal plane. In the first frame (Fig. 4a) the vocal folds on the left and right sides are initially in contact and the airway is closed. The next frame (Fig. 4b) indicates a lateral movement of the inferior portion of cover surface which continues until the left and right sides separate (Fig. 4c) thereby opening the airway. Once the maximum lateral displacement is achieved (Fig. 4d), the lower portion of the vocal folds begins to move medially with the upper portion following (Fig. 4e). Eventually, the lower portions on the left and right sides collide with each other and again close the airway (Fig. 4f). Medial displacement continues as the upper portions also collide (Fig. 4g). The entire process then repeats itself cyclically at the fundamental frequency of vibration (F_0).

Note that the lateral displacement of the upper (superior) portion of each vocal fold is not in phase with the lower portion. That is, the lower part of the vocal fold leads

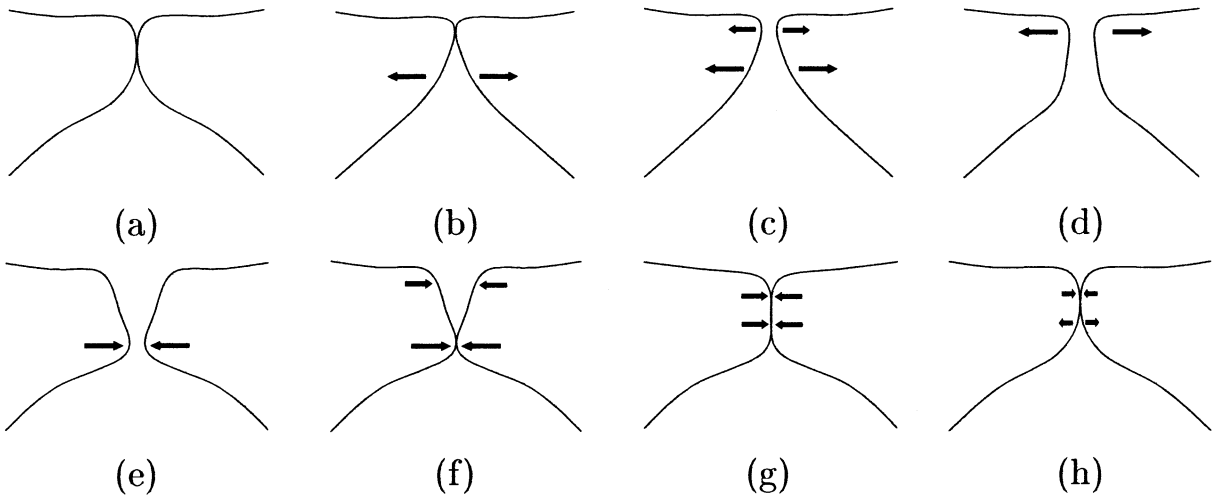


Fig. 4 Diagram showing an idealized cycle of vocal fold vibration in the coronal plane. Note that the lower portion of the vocal folds leads the upper portion creating a wave-like motion on the vocal fold surface. This is called the *mucosal wave*.

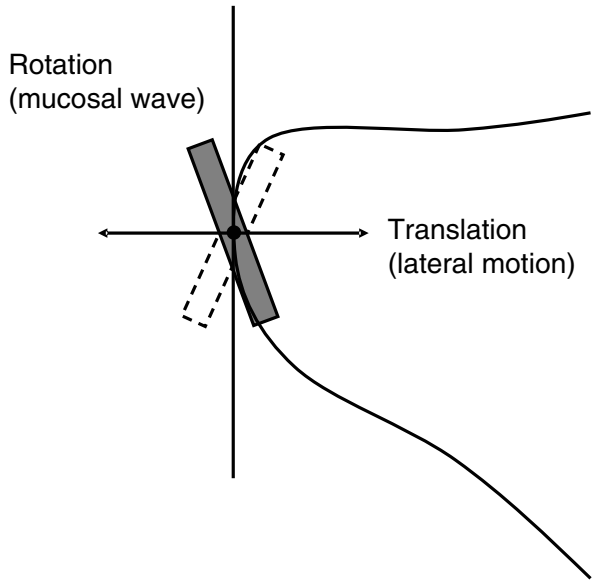


Fig. 5 Simplification of vocal fold motion. The rotation of the thin plate is intended to account for the mucosal wave while translation represents overall lateral motion.

the upper, creating wave-like motion in the cover from bottom to top. A simplified view of the vibration pattern is offered in Fig. 5 where the surface of a single vocal fold is represented as thin plate capable of both rotation and lateral displacement. Thus, the mucosal wave, indicated by the rotation of plate, ''rides'' on the overall back and forth movement of the tissue. These patterns are the two basic vibrational *eigenmodes* or natural modes of the vocal fold tissue system with regard to the motion in the coronal plane [9,10] (vibration patterns in the transverse plane can also be represented with eigenmodes [3,10]). A specific pattern of vibration can be considered to be a combination of these two *eigenmodes*.

3. OVERVIEW OF VOCAL FOLD MODELS

The kinematic patterns of tissue motion discussed in the previous section are initiated, and sustained over time, by the steady air flow and pressure supplied by the lower respiratory system. Once in vibration, the vocal folds effectively convert the steady air flow from the lungs into a series of flow pulses by periodically opening and closing the air space between the vocal folds. This airspace is called the *glottis*; hence, the stream of flow pulses is referred to as the *glottal flow* and provides the sound source for the excitation of the vocal tract resonances in vowels.

That the vocal folds are capable of converting a steady, flowing stream of air into vibratory motion puts them into a category of physical systems known as *self-oscillating* systems. Understanding the precise mechanisms responsible for self-sustained vocal fold oscillation has occupied researchers for several decades. The first formal introduction of a theory on the subject was Van den Berg's [11] ''Myoelastic-aerodynamic theory of voice production'' which empirically explored the interaction of vocal fold tissue elasticity and aerodynamic forces. This paper described the oscillatory mechanisms of vocal fold vibration in terms of tissue elasticity and the so-called Bernoulli effect. That is, high air velocity through a narrow glottis would create a negative pressure that would ''suck'' the vocal folds together, afterwhich a build-up of subglottal pressure would blow the vocal folds apart and start the process again. However, subsequent theoretical studies have assigned a secondary role to the Bernoulli effect and have formed the current view that the vocal folds may oscillate whenever an asymmetry exists between the aerodynamic driving forces produced within the glottis and the opening and closing phases of the vocal folds [9,12–14]. This asymmetry is facilitated by (1) the mucosal wave, which creates a time delay with respect to the upper and lower portions of the vocal folds, and (2) the inertial acoustic loading presented by the vocal tract which creates a time delay between the build up and release of supraglottal pressure and glottal opening/closing. Both of these factors will have the effect of decreasing the aerodynamic driving forces during the glottal closing phase and increasing them during the opening phase.

The effect of inertial loading by the vocal tract was demonstrated in an early model [15] in which the vocal fold tissue (of either the right or left side) was lumped into a single mass (Fig. 6) and allowed to have only lateral displacement. The mass was connected to a ''rigid'' lateral boundary with spring and damping elements intended to account for tissue elasticity and energy losses, respectively.

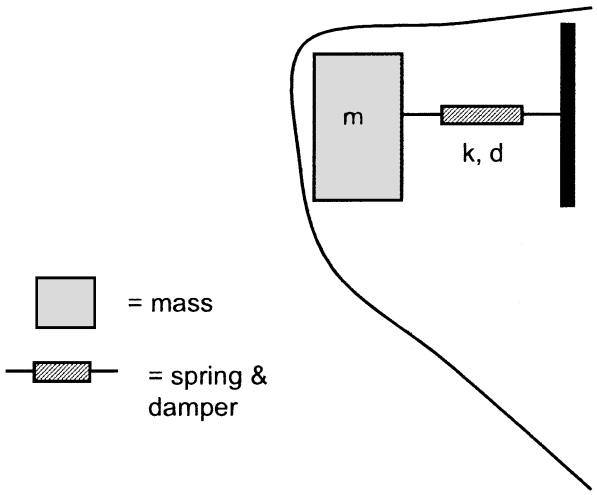


Fig. 6 The one-mass model [15] of the vocal folds. To make efficient use of space, the spring and damping elements have been lumped together in a single symbol. Note that the one-mass model accounts for lateral motion but cannot represent the mucosal wave.

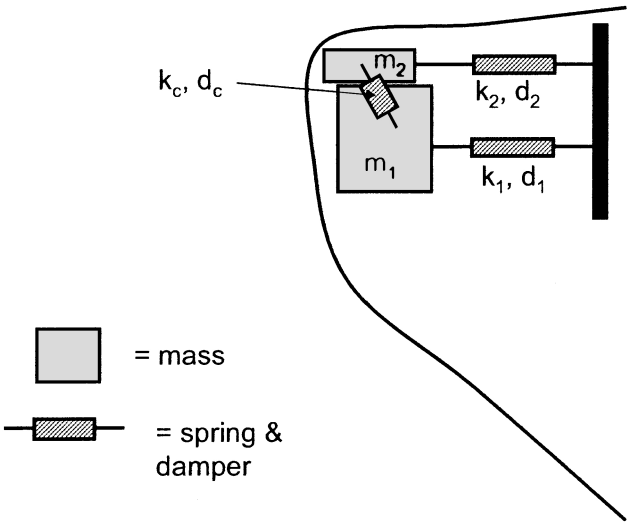


Fig. 7 A diagram of a two-mass vocal fold model [17]. Mucosal wave as well as lateral motion is represented.

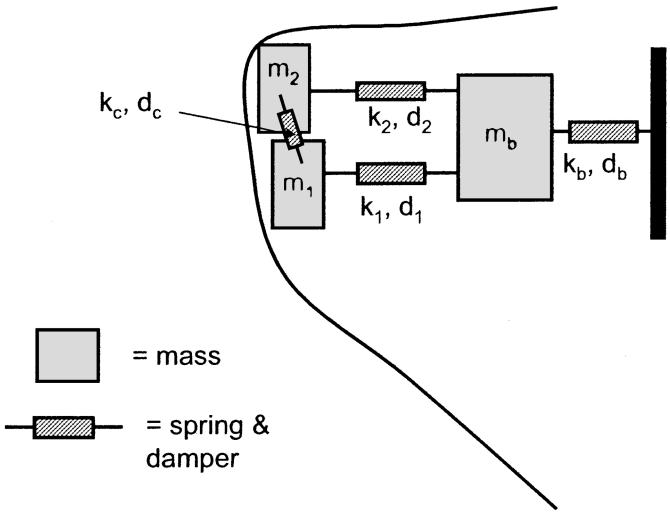


Fig. 8 The three-mass model [18] of the vocal folds. In addition to the mucosal wave and lateral motion, this model accounts for the coupling between the cover and body.

With its single-degree of freedom this one-mass model cannot represent the mucosal wave (unless it is allowed a rotational degree of freedom [16]). As a result, coupling to the acoustic inertance of the vocal tract is essential to create a condition where the tissue velocity and the intraglottal pressure are in phase and hence initiate and sustain oscillation. While the one-mass model is generally not sufficient for research purposes, it is a useful instructive tool for learning some of the fundamental principles of self-sustained oscillation [3].

Shortly after the introduction of the one-mass model [15] a slightly more complex version was proposed that represented a single vocal fold by two-masses in the coronal (vertical) dimension [12,17] (Fig. 7). Again, each mass was allowed only lateral displacement and was connected to a rigid boundary with a spring and damper system. In addition, the two masses were directly coupled to each other through another spring element to account for shear forces. The two degrees of freedom afforded by the two-mass model allows for the mucosal wave to be represented as well as overall lateral tissue displacement. While it is admittedly a crude discretization of the real tissue structure it provides appropriate conditions for oscillation to occur with or without a coupled vocal tract inertance. Because of its simplicity and reasonable agreement with physiologic data, the two-mass model has been widely used in simulation/synthesis studies of vocal fold vibration.

A limitation of both the one-mass and two-mass models is that their discretization of tissue in the coronal plane does not capture the known layered structure of the vocal folds [2]. In the two-mass model [17], the lower mass is made thicker (vertical dimension in the coronal plane) and more massive than the upper mass in an effort to simulate the effects of the body layer. However, since this arrangement does not allow for a coupled oscillation of both the cover and body layers, the two-mass model is essentially a ''cover'' model rather than a ''cover-body'' model.

It was this limitation that motivated the introduction of a three-mass model [18] that was intended to include the effect of the cover-body vocal fold structure but also maintain the simplicity of a low-dimensional system. As shown in Fig. 8, this model essentially adds a ''body'' mass that is positioned lateral to the two cover masses. The cover masses are both connected to the body mass via spring and damper systems that represent stiffness of the cover tissue as well as the effective coupling stiffness between the body and cover. The body mass, in turn, is connected to a rigid lateral boundary with another spring and damper system that account for the effective stiffness of the body which will depend on the level of contraction of the muscle tissue. Finally, to account for shear forces in the cover, the two cover masses are coupled to each other with another spring-damper element. In addition to coupled oscillation of the cover and body layers, the advantage of the three-mass model is that that physiologically realistic control parameters characterizing the cover and body tissue are more easily determined (than with the two-mass approach) when the discretization imposed by the model more closely follows anatomical boundaries. For example, a contraction of the thyroartenoid muscle (muscle in the ''body'') will increase the stiffness of the body but may not necessarily stiffen the cover.

Other more complex models have been developed to simulate the layered vocal fold structure and also account for the vibrational patterns in the transverse plane. For

example, the two-mass approach was modified such that a single vocal fold was represented by eight coupled transverse sections, each with two masses in the coronal plane that had both lateral and vertical degrees of freedom [19,20]. This model allows for simulation of the vibrational pattern of the vocal folds in both the coronal and traverse planes. A further step up in complexity came with a continuum mechanics model [21] and a more recent finite-element model [22] both of which provide a precise physiological and mechanical representation of the tissue layers of the vocal folds without lumping large anatomical regions into a few mass elements. Their large number of degrees of freedom allows for detailed study of the complex vibratory pattern observed in human vocal folds. Interestingly, however, these complex models capable of producing many modal vibration patterns, have shown that vocal fold vibration is largely dominated by only 2–3 modes of vibrations [10], similar to movement patterns shown in Fig. 5. Thus, the lumped-element models seem to capture enough of the vibratory characteristics to still serve as a useful research tool if fine detail is unnecessary.

4. INTERACTION OF THE VOICE SOURCE AND THE VOCAL TRACT

In this section a three-mass model [18] will be used to demonstrate self-oscillation of the vocal folds under the influence of several vocal tract configurations. Mathematical details concerning the mechanical aspects of the model can be found in the original publication [18] but the aerodynamic equations have been updated [23]. Table 1 gives the relevant parameter values that will be used throughout the following discussion unless otherwise stated. Note that important characteristics of phonation, such as fundamental frequency, voice quality, amplitude of vibration, may be altered by changing the parameters in Table 1. However, in keeping with the theme of this special issue concerning vowels, the following sections focus only on changing the vocal tract configuration.

Table 1 Parameter values used in the three-mass model for all simulations.

Parameter	Variable	Value	Units
lower mass	m_1	0.058	g
upper mass	m_2	0.082	g
body mass	m_b	0.094	g
lower stiffness	k_1	112,250	dyn/cm
upper stiffness	k_2	157,140	dyn/cm
body stiffness	k_b	203,660	dyn/cm
coupling stiffness	k_c	6,450	dyn/cm
start pos. - lower	x_{01}	0.0038	cm
start pos. - upper	x_{02}	0.0	cm
vocal fold length	L	1.5	cm
vocal fold thickness	T	0.3	cm
lung pressure	P_L	8,000	dyn/cm^2

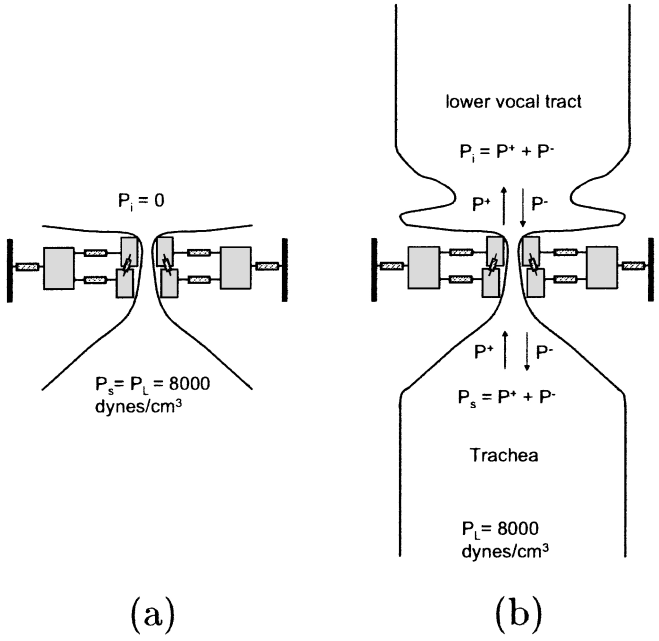


Fig. 9 Configurations of the three-mass model in ''no vocal tract'' and ''coupled vocal tract'' cases. (a) there is no vocal tract coupling so the subglottal pressure is constant and the pressure just above the vocal folds is zero, (b) this is the case with acoustic coupling of the vocal folds to both the trachea and supraglottal vocal tract.

4.1. No Vocal Tract

The first simulation was performed with a simple configuration that disregarded any influence of the acoustic pressures in the subglottal or supraglottal systems as indicated schematically in Fig. 9(a). Thus, the subglottal pressure was maintained at a constant value of 8,000 dyn/cm^2 and the supraglottal pressure was set to 0.0 (no vocal tract). The time-dependent displacement of each of the three masses is shown in Fig. 10 for the initial 25 milliseconds of simulation. To illustrate the vibration relative to the schematic at the top of the figure, a somewhat unconventional orientation was used that shows time increasing from top to bottom on the vertical axis and displacement on the horizontal axis. Zero displacement is considered to be the midline of the glottis and any time the displacement of a particular mass is negative, it would be in contact with its counterpart element of the opposing vocal fold. (Bilateral symmetry will be assumed throughout these simulations, hence the displacements of only one vocal fold are shown.) Note that the lower mass leads the upper mass thoroughout the cycles of vibration by approximately 24 degrees. This phase difference between the two masses corresponds to the mucosal wave, albeit a highly discretized wave. The displacement of the body mass is roughly in phase with the lower mass.

Figure 11 shows a set of waveforms corresponding to various quantities in the model. P_0 and P_i represent the

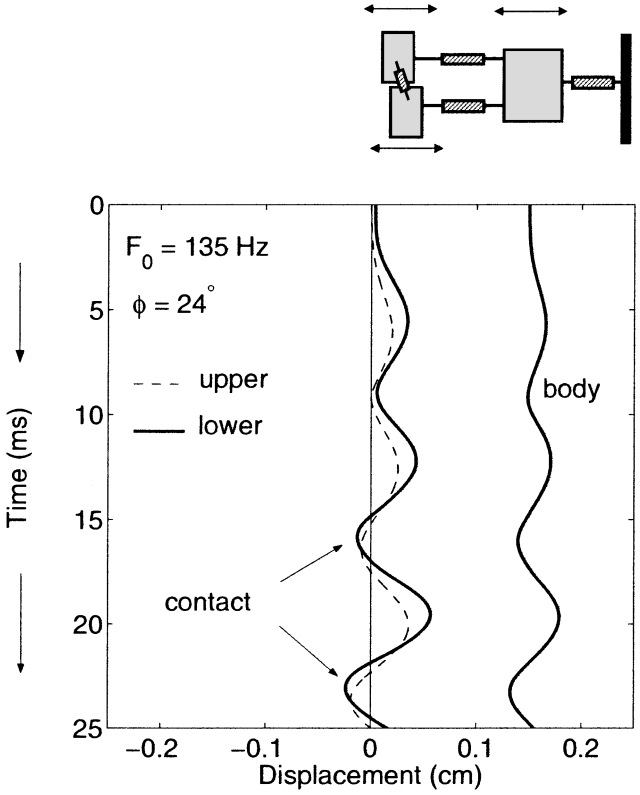


Fig. 10 Simulated displacement of the cover and body masses in the three-mass model. The orientation of the plot lines up with the schematic of the model. Time is on the vertical axis and increases from top to bottom and displacement is on the *x*-axis.

output sound pressure at the lips and the input pressure to the vocal tract, respectively. They are both zero over all time in this case because no vocal tract coupling is present. The waveform labeled P_g is the intraglottal pressure calculated as the average pressure acting on each of the cover masses. This is effectively the pressure the drives the oscillation. Note that since the interest is in waveshapes and their temporal relationships, the units on the vertical axes have been suppressed. P_s is the subglottal pressure, which is kept constant; there is, however, a short cosine ramp built into the simulation during start-up as can be seen on the left hand side of the plot. The bottom two waveforms represent glottal flow U_g and glottal area A_g, respectively. For this model, A_g is calculated as the minimum displacement of the upper or lower mass scaled by the vocal fold length and multiplied by two (to account for both left and right sides of the glottis). The flat portions in both U_g and A_g occur during the time the glottis in closed (vocal folds are in contact), thus the glottal area becomes zero and air cannot flow. Because there is no vocal tract inertance, the wave shapes of U_g and A_g are identical.

The primary mechanism of oscillation can be seen in the waveforms of Fig. 12 where intraglottal pressure P_g is again shown at the top, glottal flow A_g at the bottom, and the time derivative of A_g in the middle (denoted dA_g). The dA_g waveform effectively shows the time course of the

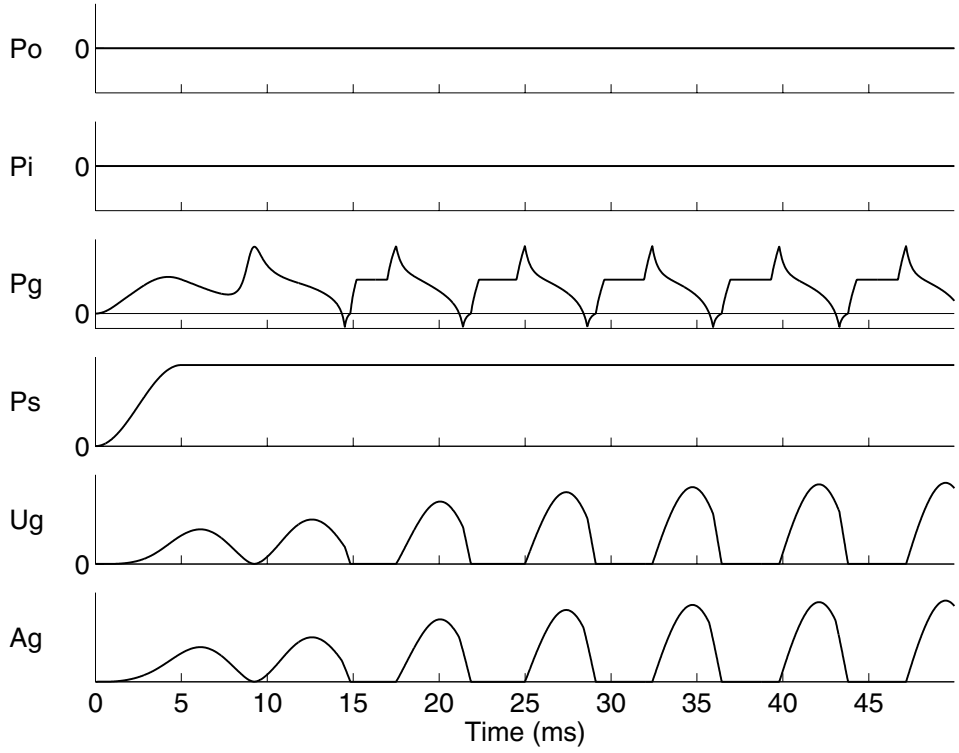


Fig. 11 Collection of waveforms for a 50 ms simulation with the three-mass model in the "no-tract" condition. P_0 = output pressure, P_i = supraglottal input pressure, P_g = intraglottal pressure (driving force), P_s = subglottal pressure, U_g = glottal flow, A_g = glottal area.

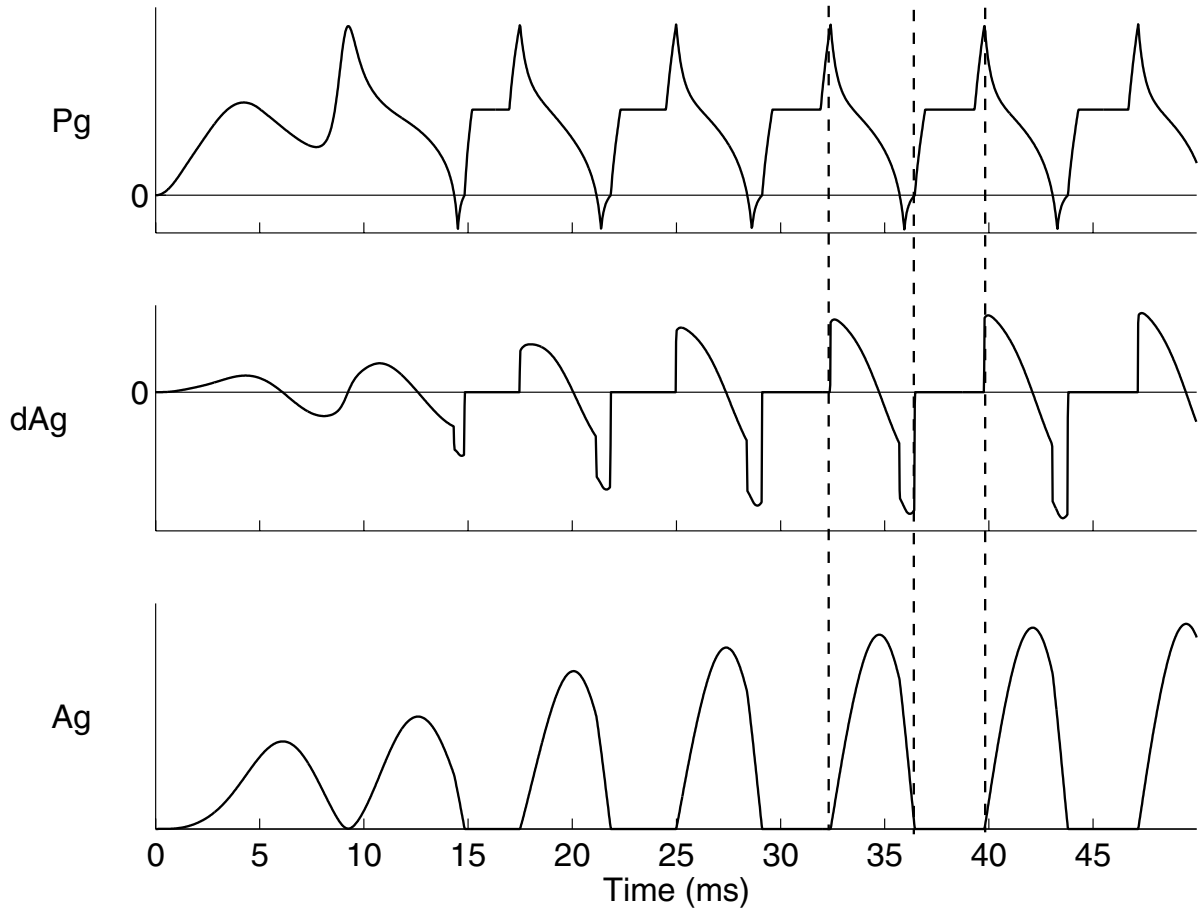


Fig. 12 Additional waveforms for a 50 ms simulation with the three-mass model in the ''no-tract'' condition. The figure shows that both the intraglottal pressure (P_g) and the velocity of the tissue (estimated by dA_g) are decreasing functions over the time course of the open phase. P_g = intraglottal pressure (driving force), dA_g = time-derivative of the glottal area (tissue velocity), A_g = glottal area.

tissue velocity. The vertical dashed lines indicate three points within one glottal cycle. The instant of glottal opening is shown by the first line where both the pressure P_g and velocity dA_g are at their maximum values. Over the course of the opening portions of the cycle, P_g and dA_g remain in phase and both are decreasing functions. Thus, the direction of driving pressure is the same as the direction of the tissue velocity. This is made possible by the two degrees of freedom in the cover portion of the model which gives rise to the mucosal wave and allows for the transfer of energy from the airstream to the tissue. Just prior to glottal closure, P_g does briefly become negative showing that the Bernoulli effect is present during vocal fold vibration. It is, however, not required for vibration to occur.

4.2. Vocal Tract Coupling With The Vowel /i/

In the second case, the vocal folds are coupled to a supraglottal vocal tract and a subglottal system, both of which allow for acoustic wave propagation as shown in Fig. 9b. The specific implementation of the vocal tract model [24] is based on the wave-reflection approach [25,26] and includes losses from yielding walls, viscosity, acoustic radiation, and heat conduction. The vocal tract shape used in this case was an /i/ vowel measured from an adult male speaker [27]. The first two formant frequencies were calculated to be $F_1 = 290\,\text{Hz}$ and $F_2 = 2,360\,\text{Hz}$. The tracheal shape was also measured for the same speaker [24]. The combined tubular system is shown in Fig. 13 as a series of ''tubelets'' representing the spatial discretization of the subglottal and supraglottal area functions (the cross-

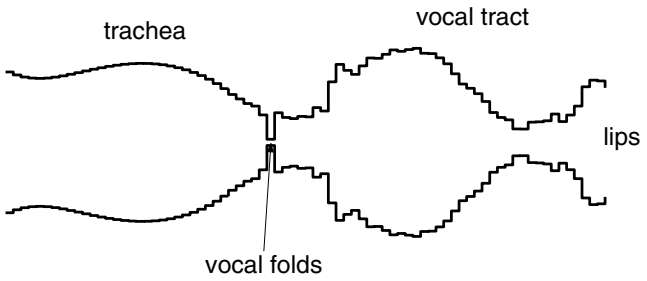


Fig. 13 Tubular representation of the subglottal and supraglottal tracts. In this case, the supraglottal tract is configured for an /i/ vowel.

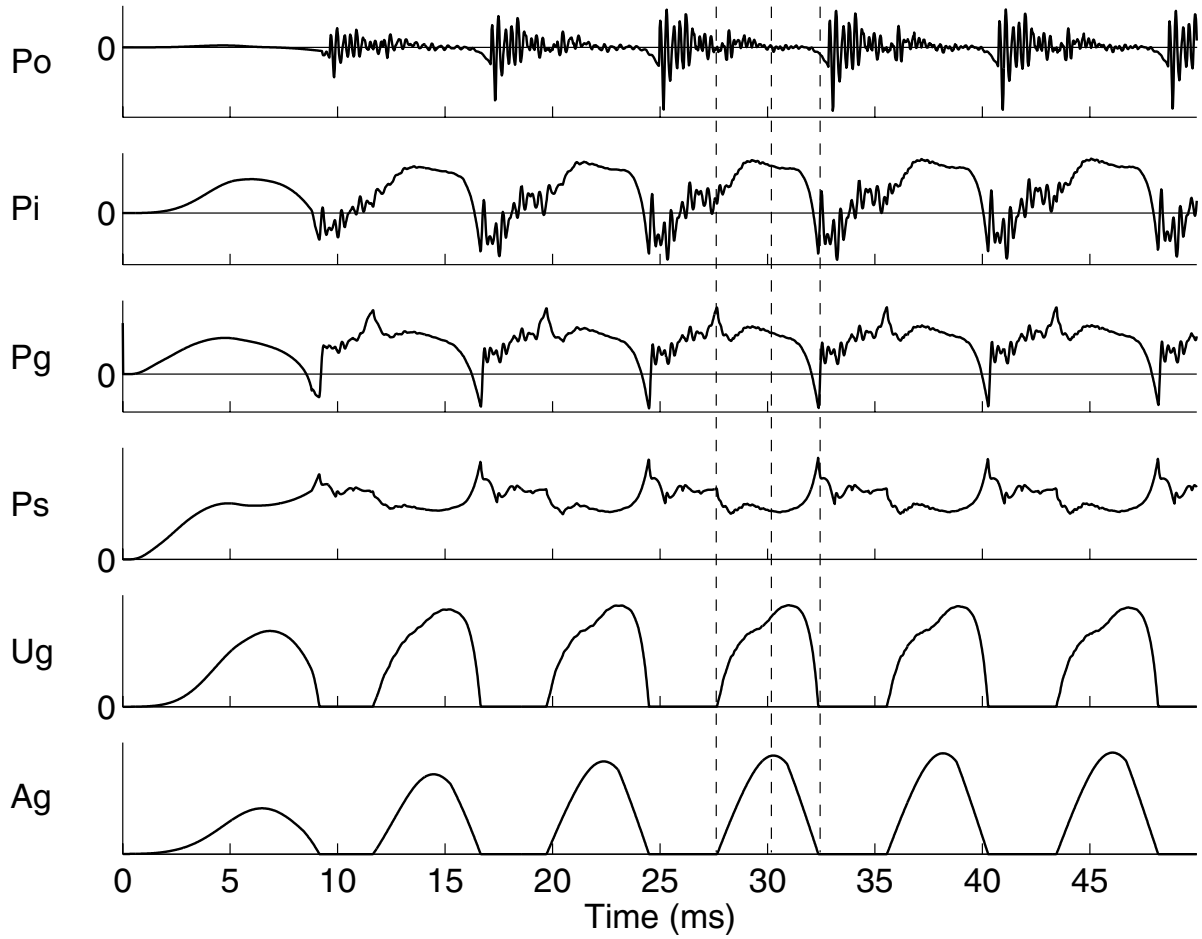


Fig. 14 Collection of waveforms for a 50 ms simulation with the three-mass model in the first ''coupled vocal tract'' condition (vowel /i/). The dashed lines represent glottal opening, maximum lateral displacement (zero velocity), and glottal closure.

sectional areas have been converted to equivalent diameters to generate the ''tube-like'' figure). Each tubelet has length of 0.396 cm giving a supraglottal vocal tract length of 17.46 cm. The three mass model would occupy the space labeled ''vocal folds.'' To maintain simplicity, the piriform sinuses have not been included in the simulation. It is recognized, however, that they may present an important modification to the input impedance of the vocal tract [28] and subsequently affect vibrational characteristics [29].

Results of the simulation are shown in Fig. 14 and all of the waveforms demonstrate the significant influence of the vocal tract resonances. The output pressure P_0 shows the result of vocal tract excitation by the source (glottal flow) as evidenced by the oscillations characteristic of the formant frequencies for an /i/ vowel. The vocal tract input pressure P_i also shows the effects of formant oscillations, especially during the closed phase of the glottal cycle (e.g. from 24–27 ms). In the open phase, P_i builds up rapidly and then maintains a nearly constant pressure until glottal closure occurs. At that point the pressure drops to a negative value, due largely to the inertia of the air column above the vocal folds. The intraglottal pressure P_g undergoes a sharp increase just prior to the opening of the glottis and then generally decreases over the course of the open phase. What is essential is that P_g begins to decrease at the point in time where the tissue maximizes its lateral displacement. This point is where the velocity becomes zero and then subsequently negative as the tissue returns toward the glottal midline; it is indicated by the middle dashed line for one of the cycles in the figure. The negative pressure just prior to closure is notably sharper than for the no-tract case; this reflects the similar negative pressure in P_i and again shows the presence of the Bernoulli effect. The pressure just below the vocal folds in the trachea P_s is roughly the mirror image of P_i, at least during the open phase. This occurs because air flows from the respiratory system through the open glottis, releasing some of the pressure that builds up in the trachea during the closed period.

Like the pressure waveforms, the shape of the glottal flow pulses are also quite different than in the no-tract condition. In particular, there is a notable skewing of the pulse rightward in time which is due to the inertia of the air column [30]. In addition, there is a slight depression in the

pulse that occurs about halfway through the open phase. This is likely created by pressure oscillations associated with the low first formant frequency of the /i/ vowel. The waveshape of the glottal area A_g may seem to be the least affected of the simulated quantities. However, the fundamental frequency of vibration in this case has been lowered to 127 Hz, down from 135 Hz in the no-tract condition. Thus, the duration of a glottal cycle has been increased about 7 percent by the presence of the vocal tract.

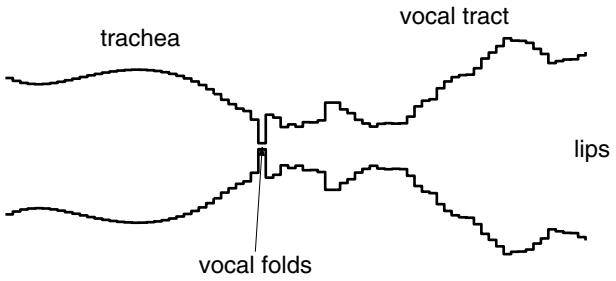


Fig. 15 Tubular representation of the subglottal and supraglottal tracts. In this case, the supraglottal tract is configured for an /ɑ/ vowel.

4.3. Vocal Tract Coupling with the Vowel /ɑ/

The supraglottal vocal tract in this case was configured with an area function for the vowel /ɑ/ measured from the same adult male speaker [27] as in the previous case of /i/. Its two lowest formant frequencies were calculated to be $F_1 = 800$ Hz and $F_2 = 1,163$ Hz. Figures 15 and 16 show the system configuration and the simulation waveforms, respectively. P_0 shows the formant oscillations of F_1 and F_2 during the closed phase. This is in contrast to the previous case of /i/ where the F_1 was so low that the duration of a single cycle was longer than the closed phase, hence the high frequency second formant was most prominent. In the open phase, the waveshapes of P_i and P_g are nearly identical suggesting a significant coupling between the source and the tract. Again it is significant for maintaining oscillation that P_g begins to decrease at the point of maximum tissue displacement (maximum glottal area). Rightward skewing of the glottal flow pulses is also a prominent feature, much as it was previously. However, the ''ripple'' on the increasing portion of the pulses is

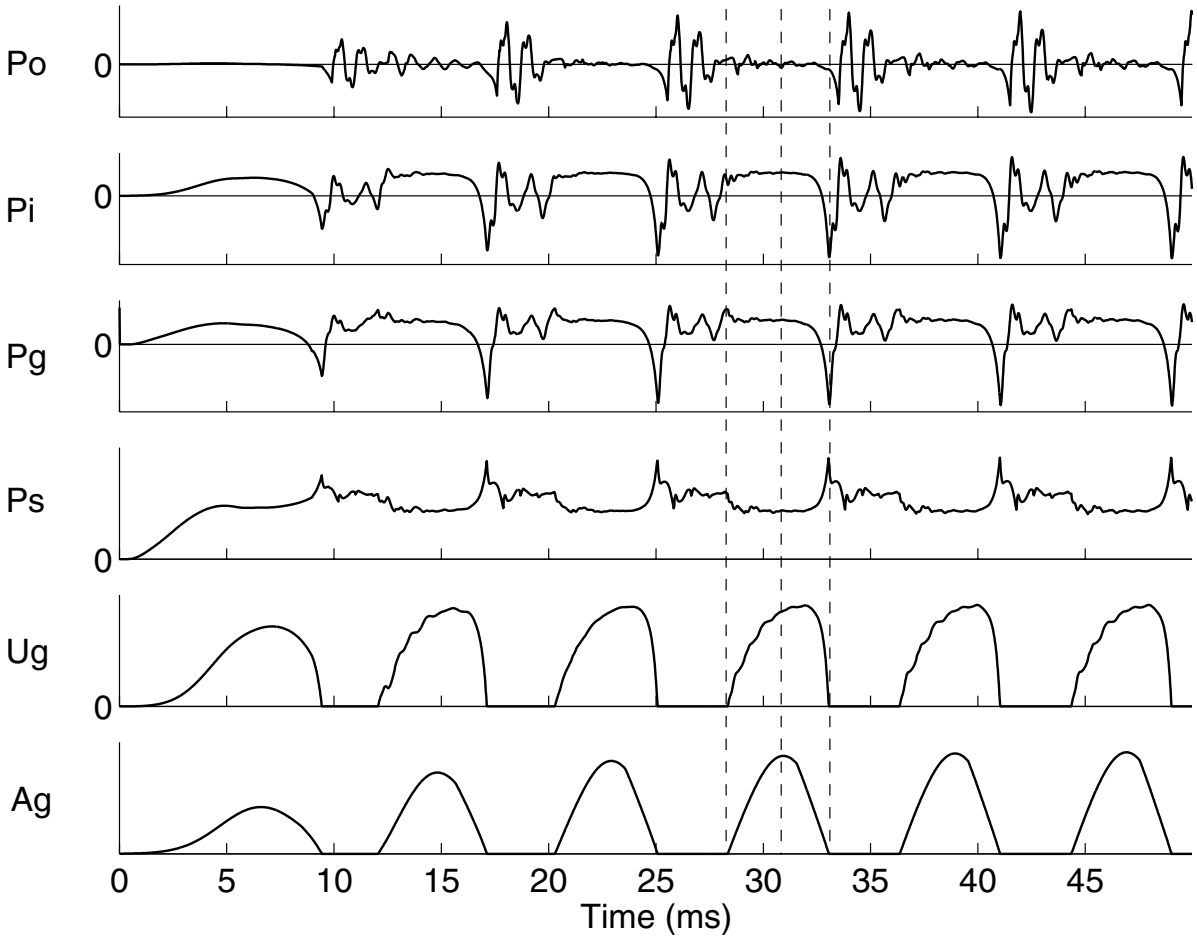


Fig. 16 Collection of waveforms for a 50 ms simulation with the three-mass model in the second ''coupled vocal tract'' condition (vowel /ɑ/). The dashed lines again represent glottal opening, maximum lateral displacement (zero velocity), and glottal closure.

indicative of a higher first formant frequency. While there is little noticeable difference in glottal area between this case and the previous, it is noted that the fundamental frequency dropped down slightly to 125 Hz suggesting the /ɑ/ vowel influences the vocal fold motion somewhat differently than the /i/.

In both cases 2 and 3, the energy losses built into the vocal tract model may be slightly excessive as illustrated by the low output pressure P_0 during the open phase. It is during this phase that losses increase because the glottis opens and allows coupling of the subglottal and supraglottal systems. In contrast, the closed phase approximates a ''closed-end'' condition for the supraglottal tract and supports the formant oscillations with less severe losses. This may be a limitation of the model, but for instructive purposes the energy losses enhance the division between the open and closed phases of the glottal cycle.

5. SUMMARY

Vocal fold vibration (and the resultant glottal flow signal) is the primary sound source for vocal tract excitation during vowel production. It is has been shown that the vocal folds are specialized anatomical structures consisting of layers of tissue with various mechanical properties. When in normal vibration, they exhibit a complex, but distinct vibrational pattern consisting of lateral motion as well as wave-like motion that travels along the surface of the vocal fold. This ''mucosal wave'' can be visualized as the lower portion of the vocal fold leading the upper in displacement over the course of a glottal cycle.

Theoretical studies of vocal fold mechanics have often generated mathematical/simulation models of the vocal folds. In general there are two classes of these models. The first is a collection of low-dimensional models that attempt to capture the essential aspects of the vibration by lumping large regions the vocal fold anatomy into discrete mass elements. Their advantage is simplicity, both in control variables and output quantities. However, their discretization of the tissue is too coarse for highly detailed simulation of tissue mechanics. Thus, a second class of models consists of finite-difference and finite-element representations of the vocal folds. These allow for many degrees of freedom and precise specification of the mechanical properties of the tissue.

Finally, as an example, a three-mass model was used to simulate vocal fold vibration under several vocal tract loading conditions. The first case was a ''no-tract'' condition in which the folds were shown to oscillate in the absence of a supraglottal vocal tract. The second and third cases were conditions with coupling to the vowels /i/ and /ɑ/, respectively. Both cases showed substantial effects of the vocal tract resonances (formants) on the output quantities of the three mass model. In all cases, the critical condition for initiating and sustaining oscillation is that the intraglottal pressure and the velocity of the tissue work in synchrony. That is, the pressure must decrease when the tissue velocity decreases.

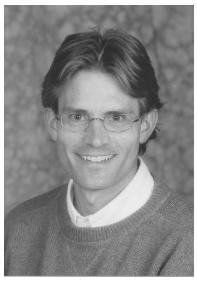
ACKNOWLEDGEMENTS

This work was supported by grant R01 DC04789 from the National Institutes on Deafness and Other Communication Disorders.

REFERENCES

[1] G. Fant, *The Acoustic Theory of Speech Production* (Mouton, The Hague, 1960).
[2] M. Hirano, ''Morphological structure of the vocal cord as a vibrator and its variations,'' *Folia Phoniatr.*, **26**, 89–94 (1974).
[3] I. R. Titze, *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[4] P. Moore, ''A short history of laryngeal investigation,'' *J. Voice*, **5**, 266–281 (1991).
[5] H. Hollien and G. P. Moore, ''Stroboscopic laminography of the larynx during phonation,'' *Acta Otolaryngol.*, **65**, 209–215 (1968).
[6] J. J. Jiang and I. R. Titze, ''A methodological study of hemilaryngeal phonation,'' *Laryngoscope*, **103**, 872–882 (1993).
[7] D. A. Berry, D. W. Montequin and N. Tayama, ''High-speed digital imaging of the medial surface of the vocal folds,'' *J. Acoust. Soc. Am.*, **110**, 2539–2547 (2001).
[8] T. Chiba and M. Kajiyama, *The Vowel: Its Nature and Structure* (Phonetic Society of Japan, Tokyo, 1958).
[9] I. R. Titze, ''On the mechanics of vocal fold vibration,'' *J. Acoust. Soc. Am.*, **60**, 1366–1380 (1976).
[10] D. A. Berry and I. R. Titze, ''Normal modes in a continuum model of vocal fold tissues,'' *J. Acoust. Soc. Am.*, **100**, 3345–3354 (1996).
[11] J. W. Van den Berg, ''Myoelastic-aerodynamic theory of voice production,'' *J. Speech Hear. Res.*, **1**, 227–244 (1958).
[12] K. Ishizaka and M. Matsudaira, ''Fluid mechanical considerations of vocal cord vibration,'' *SCRL Monograph*, **8**, 28–72 (1972).
[13] K. N. Stevens, ''Physics of laryngeal behavior and larynx modes,'' *Phonetica*, **34**, 264–279 (1977).
[14] I. R. Titze, ''The physics of small amplitude oscillation of the vocal folds,'' *J. Acoust. Soc. Am.*, **83**, 1536–1552 (1988).
[15] J. L. Flanagan and L. Landgraf, ''Self-oscillating source for vocal tract synthesizers,'' *IEEE Trans. Audio Electroacoust.*, **AU-16**, 57–64 (1968).
[16] J. Liljencrants, ''A translating and rotating mass model of the vocal folds,'' *STL Q. Progr. Stat. Rep.*, **1**, Speech Transmission Laboratory, KTH, Stockholm, Sweden (1991).
[17] K. Ishizaka and J. L. Flanagan, ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' *Bell Syst. Tech. J.*, **512**, 1233–1268 (1972).
[18] B. H. Story and I. R. Titze, ''Voice simulation with a body-cover model of the vocal folds,'' *J. Acoust. Soc. Am.*, **97**, 1249–1260 (1995).
[19] I. R. Titze, ''The human vocal cords: A mathematical model, part I,'' *Phonetica*, **28**, 129–170 (1973).
[20] I. R. Titze, ''The human vocal cords: A mathematical model, part II,'' Phonetica, **29**, 1–21 (1974).
[21] I. R. Titze and D. T. Talkin, ''A theoretical study of the effects

of various laryngeal configurations on the acoustics of phonation,'' *J. Acoust. Soc. Am.*, **66**, 60–74 (1979).
[22] F. Alipour, D. A. Berry and I. R. Titze, ''A finite-element model of vocal fold vibration,'' *J. Acoust. Soc. Am.*, **108**, 3003–3012 (2000).
[23] I. R. Titze, ''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 (2002).
[24] B. H. Story, *Speech Simulation with an Enhanced Wave-Reflection Model of the Vocal Tract*, Ph. D. Dissertation, University of Iowa (1995).
[25] J. Liljencrants, *Speech Synthesis with a Reflection-type Line Analog*, DS Dissertation, Dept. of Speech Commun. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden (1985).
[26] H. W. Strube, ''Time-varying wave filters for modeling analog systems,'' *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-30**, 864–868 (1982).
[27] B. H. Story, I. R. Titze and E. A. Hoffman, ''Vocal tract area functions from magnetic resonance imaging,'' *J. Acoust. Soc. Am.*, **100**, 537–554 (1996).
[28] J. Dang and K. Honda, ''Acoustic characteristics of the piriform fossa in models and humans,'' *J. Acoust. Soc. Am.*, **101**, 456–465 (1997).
[29] I. R. Titze and B. H. Story, ''Acoustic interactions of the voice source with the lower vocal tract,'' *J. Acoust. Soc. Am.*, **101**, 2234–2243 (1997).
[30] M. Rothenberg, ''Acoustic interaction between the glottal source and the vocal tract,'' in *Vocal Fold Physiology*, K. N. Stevens and M. Hirano, Eds. (University of Tokyo Press, Tokyo, 1981).



Brad Story is an Assistant Professor in the Department of Speech and Hearing Sciences at the University of Arizona. He received his B.S. degree from the University of Northern Iowa in 1987 and his Ph.D. from the University of Iowa in 1995. He was a Research Scientist at the Univ. of Iowa ('94–'97) and at the WJ Gould Voice Research Center in Denver, CO ('97–'00). His research focuses on the use of computer models to aid in understanding how the shapes, sizes, and movements of both the voice source components and the vocal tract contribute to the sounds of speech. He has published papers on voice source mechanisms as well as vocal tract modeling and imaging.