Journal of Phonetics (2001) **29,** 431–450
doi:10.006/jpho.2001.0148
Available online at http://www.idealibrary.com on IDEAL®

Mechanisms of modal and nonmodal phonation

David A. Berry

Department of Speech Pathology and Audiology and Department of Biomedical Engineering, National Center for Voice and Speech, The University of Iowa, IA, U.S.A.

Modal and nonmodal phonation are contrasted from the perspective of voice production. It is shown that vocal fold vibrations from both types of phonation may be formed from just a few basic building blocks, called eigenmodes. Significantly, modal and nonmodal phonation may be distinguished based on the entrainment (i.e., synchronization) of the eigenmodes, where modal phonation corresponds to a 1 : 1 entrainment, and nonmodal corresponds to all more complex patterns. Indirect and direct investigations of eigenmode entrainment are reviewed, citing investigations from computer models, excised larynx experiments, and *in vivo* investigations on human subjects. Resonance studies of the vocal folds provide an indirect investigation of eigenmode entrainment, indicating which eigenmodes are most likely to entrain based on their natural frequencies. Challenging previous interpretations of the vocal fold resonance structure derived from the two-mass model, continuum models and *in vivo* studies on human subjects are beginning to converge on a similar description of the composite resonance. In addition, finite element models and highspeed imaging studies of the medial surface of the vocal folds provide powerful, direct evidence of eigenmode entrainment in vocal fold vibration. Applications of these techniques are suggested for exploring specific entrainment mechanisms used in language.

1. Introduction

For years, it has been known that different phonation types have a variety of linguistic uses (Ladefoged, 1983). For example, consider the parameter of glottal stricture, which is used to describe the degree of vocal fold separation, ranging from widely separated folds to tightly pressed folds. Variations in glottal stricture are employed contrastively in a number of different languages and are known to produce distinct phonation types, including both modal and nonmodal phonation (Gordon & Ladefoged, 2001). Modal phonation refers to typical or baseline phonation, and includes the range of fundamental frequencies used for the speaking voice (Hollien, 1974). It is also associated with periodic vocal fold vibrations, a well-defined glottal closure, and a rich glottal spectrum. Nonmodal phonation, on the other hand, is an all encompassing term which includes anything which deviates from modal (Kreiman & Gerratt, 2001).

E-mail: daberry@ucla.edu.

0095-4470/01/040431 + 20 $35.00/0

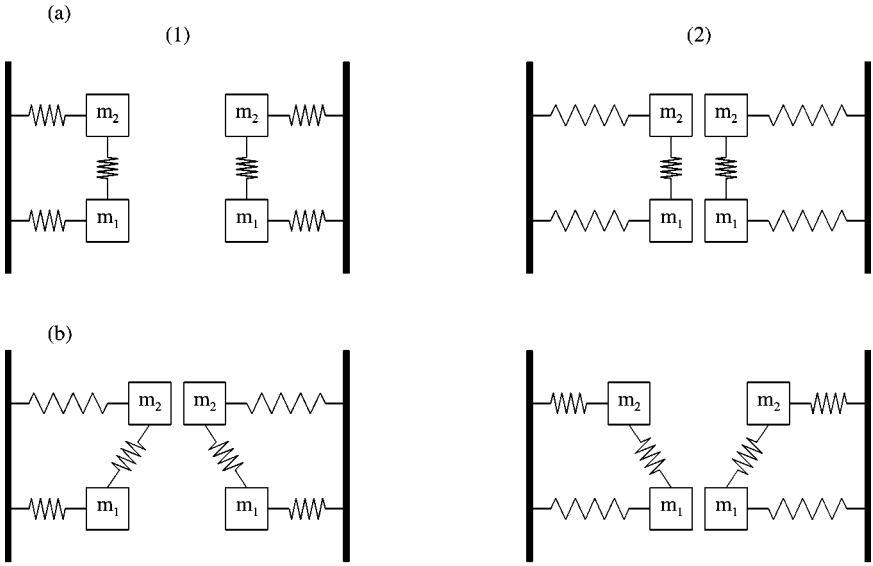

Figure 1. The two possible eigenmodes from the two-mass model, including an eigenmode with (a) the two masses vibrating in-phase, and (b) the two masses vibrating 180° out-of-phase. Columns 1 and 2 illustrate the eigenmodes at opposite extremes in the vibratory cycle. The zigzag lines represent springs.

This investigation examines modal and nonmodal phonation from the perspective of voice production. Specifically, the basic building blocks of vocal fold vibration will be examined. It will be shown that many types of vocal fold vibration can be generated from the same basic building blocks, known as eigenmodes. In fact, a basic tenet of linear vibration theory is that all vibration patterns of a system, such as the vocal folds, can be generated from the same underlying eigenmodes.

The eigenmode concept can be explained most easily using a simple, low-order model of vocal fold vibration. The most widely-known model of vocal fold vibration is the two-mass model of Ishizaka & Flanagan (1972), which contains two fundamental eigenmodes. As shown in Fig. 1, both left and right sides of the vocal folds are represented by two masses. If left–right symmetry is assumed, there are just two effective masses, which are free to vibrate horizontally. In such a model, there are two possible eigenmodes: one in which top and bottom masses vibrate in-phase with each another (Fig. 1(a)), and one in which top and bottom masses vibrate 180° out-of-phase (Fig. 1(b)). The number of eigenmodes corresponds to the number of degrees-of-freedom of the system. Any vibration pattern of the two-mass model can be expressed as a superposition of these two eigenmodes.

For a continuum of tissue like the vocal folds, an infinite number of tissue particles make up the entire mass of tissue. Thus, in the true vocal folds, there are potentially an infinite number of degrees-of-freedom, and a correspondingly infinite number of eigenmodes. However, for many types of vocal fold vibration, just a few eigenmodes are excited. Often, even very complicated vibration patterns can be explained by just a few underlying eigenmodes (Berry, Herzel, Titze & Krischer, 1994).

In a linear system, each eigenmode is associated with a fixed, characteristic frequency, called an eigenfrequency. Eigenfrequencies are identical to the resonance frequencies

of the system, a concept familiar to phoneticians and linguists. For example, in the phonetics literature, the resonance frequencies of the vocal tract are commonly referred to as formants. From lowest to highest, the first few formants of the vocal tract are usually referred to as $F1$, $F2$, $F3$, etc. For the case of a neutral vocal tract (e.g., a tract with uniform width across its entire length), the formants F_n may be computed from the following formula:

$$F_n = \frac{2n-1}{4}\frac{c}{L} \tag{1}$$

where c is the speed of sound in air, and L is the length of the vocal tract. The acoustic eigenmodes associated with the first three formants are shown in Fig. 2. All acoustic waves transmitted through the vocal tract may be expressed as a superposition of the acoustic eigenmodes.

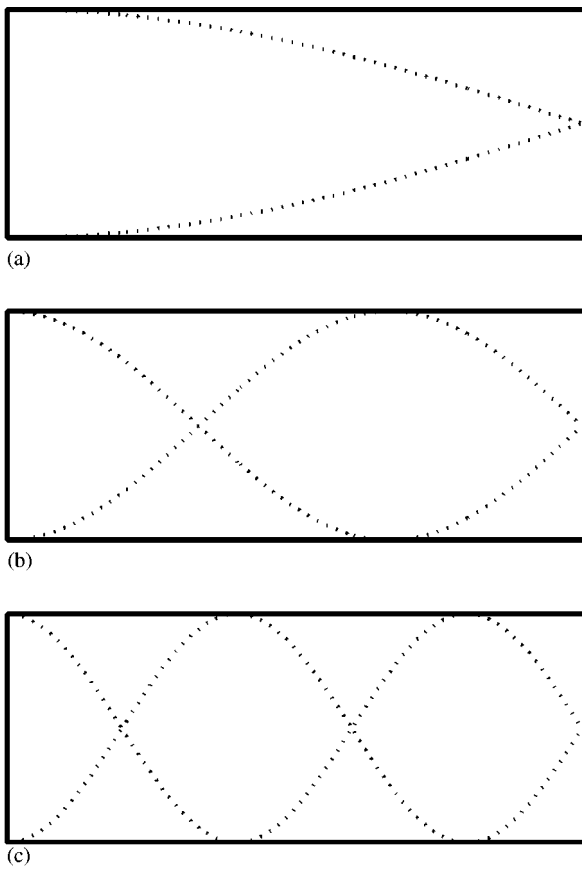

Figure 2. The acoustic eigenmodes associated with formants (a) $F1$, (b) $F2$, and (c) $F3$ for a uniform vocal tract. ····· lines represent the pressure amplitudes of the eigenmodes along the length of the vocal tract from the closed end, representing the glottis, to the open end, representing the mouth. The vocal tract of the neutral *schwa* can be considered to be a uniform closed-open tube.

When a vowel other than the neutral *schwa* is uttered, the vocal tract leaves the neutral position and becomes a nonuniform tube. This shifts the lower formants towards higher or lower frequencies, especially $F1$ and $F2$ (Hess, 1983). Each vocal tract shape has a characteristic formant structure, which is important for vowel identification. In fact, most vowels can be distinguished by where they lie in the $F1$–$F2$ parameter plane (Peterson & Barney, 1952). This description of vowel production using formants is universally appreciated by speech scientists.

Similarly, the voice source can be characterized by its eigenmodes and eigenfrequencies. However, unlike acoustic wave propagation through the vocal tract, vocal fold vibration cannot be described as a strictly linear phenomena. For example, during phonation, the eigenmodes of the vocal folds usually do not retain their characteristic frequency. In the case of modal phonation, a shifting of the eigenfrequencies occurs until the two eigenfrequencies become aligned, or tuned to the same frequency. This phenomenon is referred to by various names, including frequency locking (Bergé, Pomeau & Vidal, 1984), phase locking (Glass & Mackey, 1988), eigenmode entrainment (Berry *et al*., 1994), modal synchronization (Herzel, Berry, Titze & Saleh, 1994), and mode locking (Fletcher, 1996). Eigenmode entrainment, as it will be referred to in this article, is a nonlinear phenomenon; it cannot occur in a strictly linear system.

For example, if eigenmode entrainment were to occur in the vocal tract, $F1$ and $F2$ might shift to some common mid-range frequency. Of course, experience suggests that this does not happen. Instead, $F1$ and $F2$ retain their characteristic frequencies. However, in the case of vocal fold vibration, there are several strong nonlinearities which facilitate eigenmode entrainment. Perhaps, the most important of these is vocal fold collision, which occurs every cycle in modal phonation in connection with glottal closure. The driving force, i.e., the airflow-pressure relation in the glottis, is another important nonlinearity which facilitates eigenmode entrainment.

From a voice production point of view, one instructive way to distinguish modal and nonmodal phonation is through the entrainment of the eigenmodes. For example, if the two eigenmodes of Fig. 1 were to oscillate at the same frequency, they would be exhibiting a 1:1 (one-to-one) entrainment, the kind of entrainment usually associated with modal phonation. On the other hand, vocal fry, a common class of nonmodal phonation, is characterized by glottal pulses of alternating amplitudes or by irregular trains of pulses (Hollien and Michel, 1968). For alternating glottal pulses, entrainment patterns of 1:2 or 2:2 are most common. For a 1:2 entrainment, one eigenmode vibrates at the lower frequency, completing cyclic vibrations in connection with alternating glottal pulses. The other eigenmode vibrates at the higher frequency, completing cyclic vibrations in connection with each glottal pulse. For 2:2 entrainment, both eigenmodes vibrate at the lower frequency. For irregular trains of glottal airflow, the eigenmodes do not entrain at all, or they become disentrained.

Another type of nonmodal phonation is produced when low-frequency modulations, unrelated to the fundamental frequency, are present in the vibration pattern of the vocal folds (Herzel *et al*., 1994). In the music literature, this phenomenon has been referred to as bitonal phonation. In the clinical literature dealing with voice disorders, this phenomenon has been referred to as diplophonia, biphonation, and dicrotic dysphonia. In this phonation type, the eigenmodes may exhibit a P:Q entrainment, where P and Q are irrational or incommensurate (i.e., no simple integer relationship exists between P and Q).

Thus, from a voice production point of view, one way to distinguish modal and nonmodal phonation is as follows: (1) modal phonation is the type of phonation that results from a 1:1 entrainment of the eigenmodes of the vocal folds, (2) nonmodal phonation results from all other, more complex patterns.

Conceptually, eigenmode entrainment is important because it provides a more general description of the vibration rate of the vocal folds than the more traditional measure of fundamental frequency. Indeed, the whole concept of fundamental frequency presumes that a single frequency can adequately characterize this vibration rate. Admittedly, fundamental frequency may characterize vocal fold vibration adequately during modal phonation, because, in a 1:1 entrainment, all eigenmodes *do* entrain at a single frequency. However, for the more complex patterns associated with nonmodal phonation, more than one frequency may be needed to characterize the vibration rate of the vocal folds. Significantly, eigenmode entrainment can be used to describe not only modal phonation (1:1 entrainment), but also many types of nonmodal phonation, ranging from 1:2 entrainment, to 2:2, to *P*:*Q*, to complete disentrainment. Consequently, to describe both modal and nonmodal phonation, the concept of eigenmode entrainment is not only convenient, but necessary.

In this study, to help clarify this distinction between modal and nonmodal phonation, eigenmode entrainment will be discussed from a variety of points of view, citing both direct and indirect investigations of entrainment. First of all, entrainment will be discussed in light of the resonance structure of the vocal folds, which is an indirect investigation of eigenmode entrainment. Investigations with computer models, excised larynx experiments, and human subjects will be cited. Next, direct observations of eigenmode entrainment will be presented along the medial surface of the vocal folds. Since these observations require highly invasive measures, no data from living subjects are available. However, investigations with computer models and excised larynx experiments provide convincing evidence of entrainment/disentrainment in vocal fold vibration. Finally, future applications of these techniques will be suggested for exploring specific mechanisms of modal and nonmodal phonation in language.

2. The resonance structure of the vocal folds

All vibrating systems possess a unique resonance structure. In the case of linear dynamics, a system can be decomposed into a set of independent vibration patterns, called eigenmodes, each with a corresponding natural frequency or eigenfrequency. Any system of vibration can be described as a superposition of the eigenmodes. Expressed in another way, a specific entrainment of the eigenmodes can reproduce all possible vibrations of the system. Often, even complex vibration patterns can be described by just a few eigenmodes.

Both the eigenmodes and the eigenfrequencies are characteristic of the system, and independent of the source of excitation. The resonance structure of the system is determined by the width and location of the eigenfrequencies. Significantly, this resonance structure can either facilitate or discourage entrainment of the eigenmodes. For example, a close spacing of two eigenfrequencies will facilitate a 1:1 entrainment of the two corresponding eigenmodes. Nonlinearities in the system can also facilitate entrainment, by allowing the natural frequencies of the system to shift.

2.1. *The two-mass model*

With regard to the resonance structure of the vocal folds, first consider the two-mass model of Ishizaka & Flanagan (1972). Although a relatively simple model, it is elegant in that it helps conceptualize the interaction between airflow and tissue movement to produce self-oscillation. In particular, it possesses two eigenmodes, the entrainment of which illustrates a principal mechanism for self-oscillation of the vocal folds. However, the resonance structure of the two-mass model differs from more sophisticated continuum models of vocal fold vibration, and from *in vivo* investigations on human subjects. In particular, the two-mass model has a relatively large spacing between eigenfrequencies, which may introduce an unduly limited entrainment region for its eigenmodes, and an unduly restrictive region for self-oscillation, in general.

The eigenfrequencies (or resonances) of the model can be computed from its linearized equations of motion (Ishizaka & Flanagan, 1972; Titze, 1976):

$$\begin{aligned} m_1\ddot{x}_1 + r_1\dot{x}_1 + k_1x_1 + k_c(x_1 - x_2) &= F_1 \\ m_2\ddot{x}_2 + r_2\dot{x}_2 + k_2x_2 + k_c(x_2 - x_1) &= F_2 \end{aligned} \tag{2}$$

where m_1 and m_2 are the lower and upper masses, x_1 and x_2 are the lower and upper displacements, k_1, k_2, and k_c are the lower, upper, and coupling stiffnesses (represented by springs or zigzag lines in Fig. 1), r_1 and r_2 are the mechanical resistances associated with the lower and upper springs, and F_1 and F_2 are the aerodynamic forces on the lower and upper masses, which are set to zero for the resonance computation. To simplify the calculation, the two second-order equations of Equation (2) may be converted into one fourth-order equation (Ishizaka, 1988). This is done by solving for x_2 in the first equality of Equation (2) and substituting the result into the second equality. Assuming that x_1 can be expressed in the form of e^{st} $(0 < t < 0)$, where s is a complex number, the solution for the vocal fold resonances can be written in the form:

$$\begin{aligned} s^4 + \left(\frac{r_1}{m_1} + \frac{r_2}{m_2}\right)s^3 + \left(\frac{k_1 + k_c}{m_1} + \frac{k_2 + k_c}{m_2} + \frac{r_1r_2}{m_1m_2}\right)s^2 \\ + \left(\frac{r_1(k_2 + k_c) + r_2(k_1 + k_c)}{m_1 + m_2}\right)s + \left(\frac{k_1k_c + k_2(k_1 + k_c)}{m_1m_2}\right) = 0 \end{aligned} \tag{3}$$

Consider the resonance structure for typical Ishizaka and Flanagan parameters ($m_1 = 0.125$ g, $m_2 = 0.025$ g, $k_1 = 80\,000$ dyn/cm, $k_2 = 8000$ dyn/cm, $k_c = 25\,000$ dyn/cm, $r_1 = 0.1[m_1k_1]^{1/2}$, $r_2 = 0.6[m_2k_2]^{1/2}$). The resonance structure is obtained by solving for the variable s in Equation (3). The results are summarized in Table I and Fig. 3. Although there are four possible solutions to Equation (3), Table I lists only the two solutions which correspond to positive frequencies (the negative frequencies have the same magnitude as the two positive frequencies, and do not introduce any additional information). The real part of s is an exponential damping factor, and the imaginary part of s is the angular frequency ω. The two resonance peaks appear at 121 and 198 Hz, respectively. For comparison, the resonance structure of the model is also shown for the condition of no damping ($r_1 = r_2 = 0$). In this case, the two resonances occur at 120 and 201 Hz. As illustrated in Fig. 3, damping has the influence of both broadening and decreasing the amplitudes of the resonance peaks. However, the locations of the peaks shift very little as a function of damping.

TABLE I. The natural frequencies of the two-mass model

	$s = \beta + \mathbf{j}\omega$		$f = \omega/(2\pi)$	
	s_1 (rad/s)	s_2 (rad/s)	f_1 (Hz)	f_2 (Hz)
$r_1 = 0.1(m_1k_1)^{1/2}$ $r_2 = 0.6(m_2k_2)^{1/2}$	− 74.7 + j 757.9	− 135.0 + j 1244.4	121	198
$r_1 = r_2 = 0$	0 + j 756.6	0 + j 1260.0	120	201

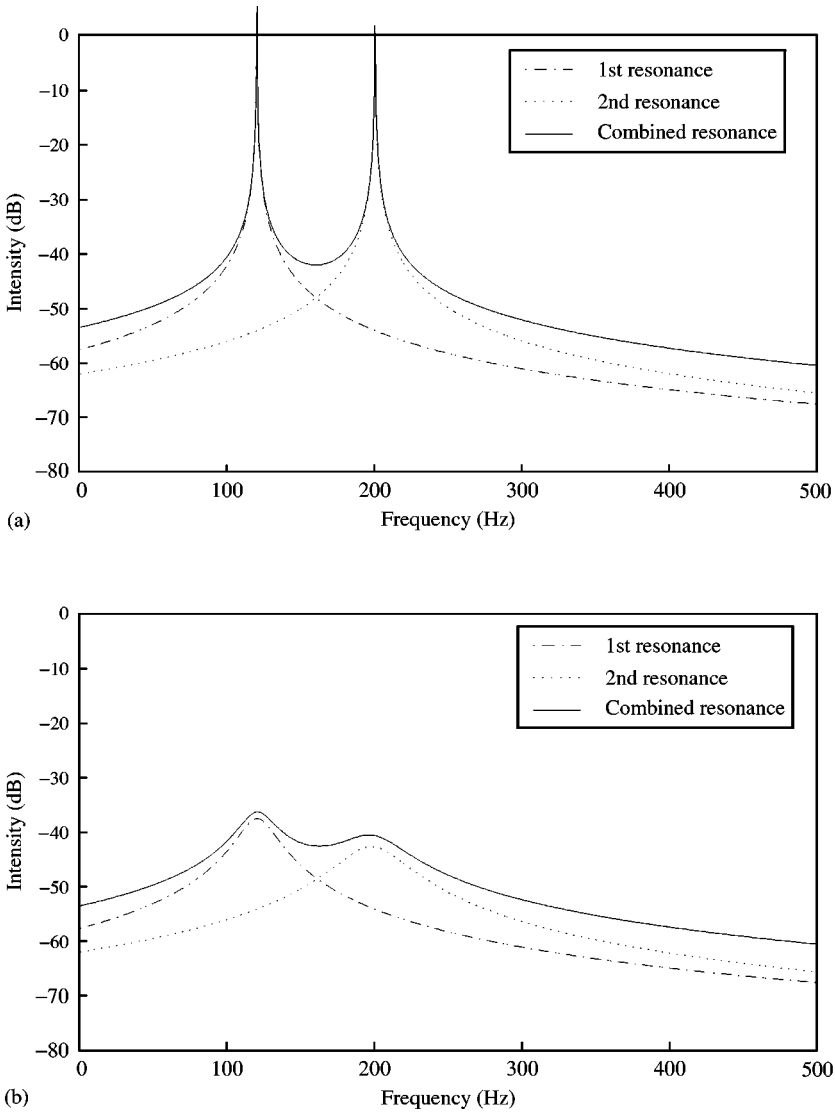

Figure 3. Resonance spectrum of the two-mass model for conditions of (a) no damping, (b) typical damping parameters of Ishizaka & Flanagan (1972).

The most significant aspect of the resonance spectrum is the relatively large spacing between the eigenfrequencies, which discourages a 1:1 entrainment of the eigenmodes. Consequently, it limits the region where modal phonation can occur. What mechanism exists to align the two natural frequencies, entraining the two eigenmodes? Aerodynamic forces (F_1 and F_2), vocal fold collision, and nonlinear spring stiffnesses can all help facilitate eigenmode entrainment. As shown previously (Titze, 1976; Ishizaka, 1988), a 1:1 eigenmode entrainment with sustained oscillations can be achieved at a fundamental frequency of approximately 130 Hz. However, for this entrainment to occur, the eigenfrequency of the second eigenmode must undergo a substantial frequency shift (e.g., from 198 to 130 Hz).

In Fig. 3, the lower resonance corresponds to an eigenmode in which the lower and upper masses vibrate approximately in-phase with each other (Fig. 1(a)), and the higher resonance corresponds to an eigenmode in which the two masses vibrate approximately 180° out-of-phase (Fig. 1(b)). The lower eigenmode is closely associated with the net lateral tissue movement. The second eigenmode captures the glottal shaping of the vocal folds, alternately shaping a convergent (Fig. 1(b), frame 1) and divergent glottis (Fig. 1(b), frame 2). A convergent glottis is one in which the superior glottal space is smaller than the inferior glottal space. That is, as one traverses the glottal airway from bottom to top, the glottal airway narrows, or converges. A divergent glottis is just the opposite, meaning that the glottal airway increases as one traverses the glottal airway from bottom to top. It is known that a divergent glottis possesses a relatively low intraglottal pressure and that a convergent glottis possesses a relatively high intraglottal pressure (Stevens, 1977; Broad, 1979; Titze, 1988; Berry *et al*., 1994). Consequently, if the two modes are properly entrained, a mechanism exists to transfer energy from the airflow to laryngeal tissues, facilitating self-sustained vocal fold oscillations. In particular, optimal energy transfer occurs when the lateral tissue velocity (controlled by the lower eigenmode) is in-phase with the intraglottal air pressure (controlled by the higher eigenmode). Consequently, a 1:1 entrainment of these two eigenmodes facilitates sustained oscillations in the two-mass model by optimizing the energy transfer from the airflow to the tissue.

2.2. *The continuum model*

Continuum models seek a more realistic representation of vocal fold tissues. While a heavy price is paid in terms of additional mathematical complexity, investigations with continuum models predict a substantially different resonance structure for the vocal folds, which corresponds more closely with resonance studies on human subjects. Berry & Titze (1996) showed a method to compute the eigenfrequencies and eigenmodes of a 3D continuum model of vocal fold tissues. As shown in Fig. 4, the model assumed a brick-shaped geometry for the tissue, with fixed boundary conditions at the anterior, posterior and lateral surfaces. The remaining surfaces (superior, inferior, and medial) were free to vibrate, with no external forces applied. Anterior–posterior vibrations were not considered because such vibrations are usually negligible (Baer, 1981). However, both medial and vertical vibrations were allowed, in contrast to the two-mass model which permits only medial vibrations. Transverse isotropy was assumed because the tissues are known to be stiffer along the direction of the fibers (roughly an anterior–posterior direction).

The resonance frequencies and eigenmodes were determined by a two-step procedure. First, a general solution was supplied which satisfied the fixed boundary conditions. In

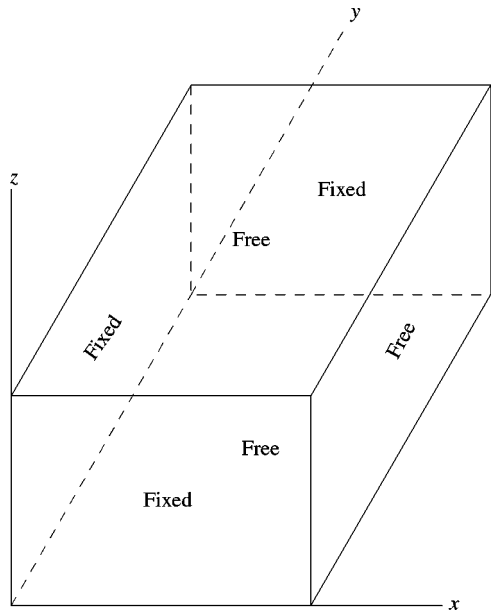

Figure 4. Boundary conditions imposed for the 3D continuum model of vocal fold tissues (after Berry & Titze, 1996).

particular, the x displacements ξ and the z displacements γ were taken to be

$$\begin{aligned}\xi(x, y, z, t) &= \sin \omega t \sin \frac{n\pi y}{L} \sum_{i=1}^{I} \sum_{j=1}^{J} A_{ij} x^i z^j \\ \gamma(x, y, z, t) &= \sin \omega t \sin \frac{n\pi y}{L} \sum_{i=1}^{I} \sum_{j=1}^{J} B_{ij} x^i z^j\end{aligned} \tag{4}$$

where i, j, n, I and J were integers and L was the anterior–posterior length of the vocal folds. Note that the fixed boundary conditions were satisfied by this general solution, i.e., both ξ and γ equal zero at $y = 0$, $y = L$, and $x = 0$.

Next, the first variation in the total potential energy Π was required to vanish. This requirement ensured that the solution satisfied the equations of motion, as well as the free boundary conditions. In particular, the first variation in Π was taken with respect to all the unknown coefficients of the general solution:

$$\frac{\partial \Pi}{\partial A_{ij}} = 0, \quad \frac{\partial \Pi}{\partial B_{ij}} = 0 \tag{5}$$

Berry & Titze (1996) noted that two of the lowest-order eigenmodes of the continuum model were qualitatively similar to the two eigenmodes of the two-mass model. As shown in Fig. 5(a), in one eigenmode, the tissue along the medial surface exhibits predominantly lateral vibrations, responsible for modulating the glottal airflow. In contrast, as shown in Fig. 6a, the other eigenmode exhibited a rotational motion along the medial surface, with lower and upper portions of the tissue vibrating 180° out-of-phase.

However, the eigenfrequency spacing in the continuum model was quite distinct from that observed in the two-mass model. In the continuum model, the two eigenfrequencies

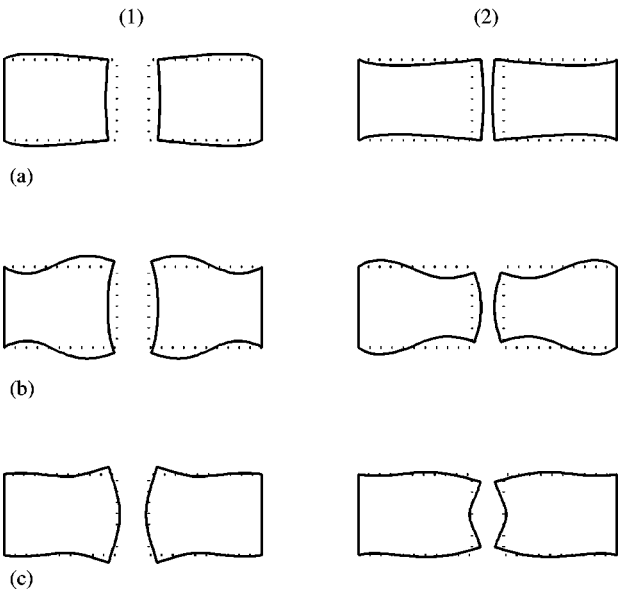

Figure 5. Low-order eigenmodes from the continuum model, shown in coronal cross-section, which describe predominantly lateral vibrations to modulate the glottal airflow. The natural frequencies of these eigenmodes are (a) 106.0, (b) 115.6, and (c) 116.7 Hz. Columns 1 and 2 illustrate the eigenmodes at opposite extremes in the vibratory cycle.

were nearly identical. Significantly, this eigenfrequency spacing encouraged a 1:1 entrainment of the eigenmodes. Thus, in the continuum model, a large shift in the eigenmodes was not necessary to initiate sustained vocal fold oscillations, yielding less restrictive oscillation conditions than the two-mass model. Furthermore, the close eigenfrequency spacing of the eigenmodes existed not only for the typical parameters, but across a wide range of vocal fold sizes and tissue stiffnesses (Berry & Titze, 1996).

Perhaps most significantly, a whole series of eigenmodes existed within the continuum model. For example, consider the composite spectrum of 16 low-order eigenmodes from the continuum model, for the parameters listed in Table II. As shown in Fig. 7, the upper solid line depicts the composite resonance spectrum, while the lower dotted lines depict the individual resonance curves. Since the individual eigenfrequencies clustered into two groups, only two broad resonances occurred in the continuum model for this parameter configuration. From a superior view, all the eigenmodes associated with the first resonance are graphically depicted in Fig. 8(a). They have an anterior–posterior index of $n = 1$ (see Equation (4)). The eigenmodes that fall within the second resonance have an anterior–posterior index of $n = 2$ and are graphically depicted in Fig. 8(b). Thus, for the parameters given in Table II, the eigenmodes naturally cluster into two groups distinguished by their anterior–posterior index.

From a superior view, the eight eigenmodes within each resonance are indistinguishable. However, they can be easily distinguished from a coronal cross-section. Descriptively, the eight eigenmodes may be classified into three distinct groups: lateral

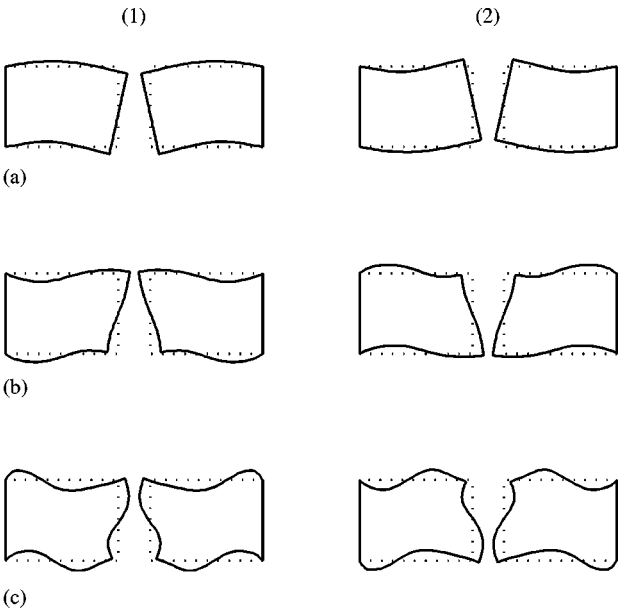

Figure 6. Low-order eigenmodes from the continuum model, shown in coronal cross-section, which are rotational in nature, alternating between convergent (frame 1) and divergent (frame 2) glottal shapes. The natural frequencies of these eigenmodes are (a) 106.7, (b) 112.0, and (c) 123.8 Hz. Columns 1 and 2 illustrate the eigenmodes at opposite extremes in the vibratory cycle.

TABLE II. Parameters of the continuum model

Transverse shear modulus μ	10^3 dyn/cm^2*
Anterior–posterior shear μ'	10^5 dyn/cm^2
Transverse Poisson's ratio v	0.99
Anterior–posterior Poisson's ratio v'	0.0
Damping ratio	0.2*
Anterior–posterior length L	1.5 cm
Lateral depth D	0.7 cm
Vertical thickness T	0.5 cm
Tissue density	1.03 g/cm^3

*Chan & Titze (1999).

eigenmodes, rotational eigenmodes, and vertical eigenmodes. These classifications are depicted in Figs 5, 6, and 9, respectively. In Fig. 5, the lateral eigenmodes exhibit a predominantly lateral vibration pattern, which modulate the glottal airflow. The first eigenmode, Fig. 5(a), is probably excited for relatively low subglottal pressures. The higher eigenmodes, Figs 5(b) and (c), have slightly higher frequencies for the same glottal configuration, and may be excited at slightly higher subglottal pressures. In Fig. 6, the rotational eigenmodes exhibit an alternating convergent/divergent glottis, and thus

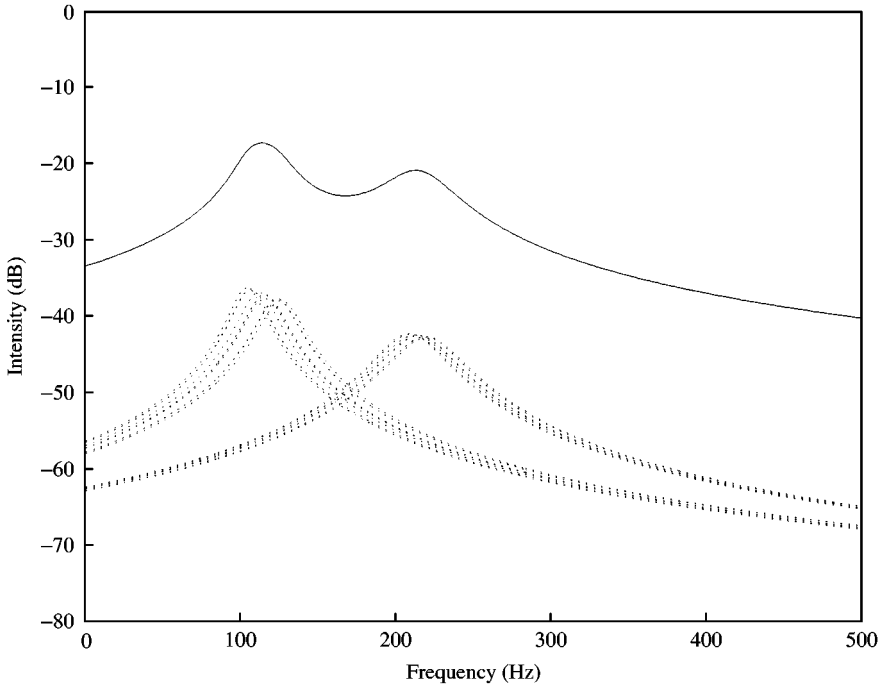

Figure 7. Resonance spectrum of the continuum model. Composite spectrum is shown as a ——, and individual resonance curves are shown as ····· lines. Two major resonances occur, corresponding to anterior–posterior indices of n equals one and two, respectively (see Equation (4)).

are closely associated with the intraglottal pressure. Note that small increases in eigenfrequency in Figs 6(a)–(c), result in eigenmodes which manifest an increasingly predominant mucosal wave. In Fig. 9, although some glottal shaping may be involved, these appear to be predominantly vertical modes of vibration, which may or may not be important in vocal fold vibration. Since some glottal shaping is involved, they could perform a function similar to the rotational eigenmodes at relatively small subglottal pressures.

Similar to the two-mass model, self-sustained oscillations of the continuum model could be facilitated if an eigenmode from Fig. 5, which controls the net lateral tissue velocity, were to entrain with an eigenmode from Fig. 6, which controls the intraglottal pressure. Properly entrained, two such eigenmodes would generate optimum energy transfer from the airflow to the tissues, facilitating self-sustained oscillations of the vocal folds. Since the first resonance peak is considerably stronger than the second, the eigenmodes of the first resonance are the most easily excited. Consequently, phonation frequency usually corresponds to the frequency of this first resonance.

In contrast to the two-mass model, the continuum model predicts that an entire series of eigenmodes exists within a single resonance of the vocal folds. The eigenmodes which most easily entrain are the eigenmodes which fall within the same resonance peak. Within the same resonance peak, there are many combinations of 1:1 entrainment which may exist between many different eigenmodes. If eigenmodes within the same resonance peak entrain, only a small shifting of eigenfrequencies is required for a 1:1 entrainment to

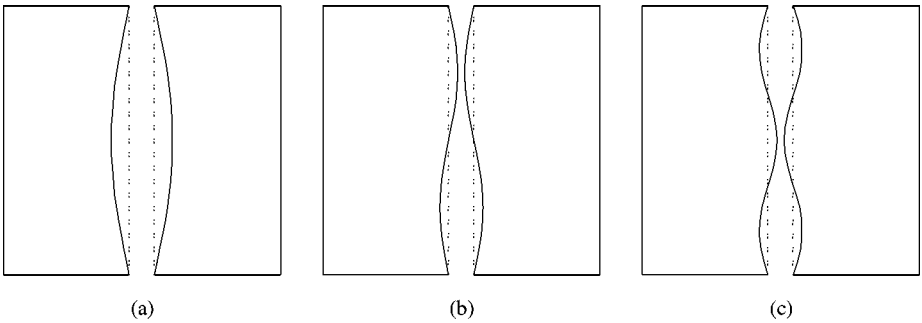

Figure 8. A superior view of eigenmodes from the continnum model (Berry & Titze, 1996) and from *in vivo* investigations (Švec *et al*., 2000), for anterior–posterior indices of (a) $n = 1$, (b) 2, and (c) 3, as defined in Equation (4).

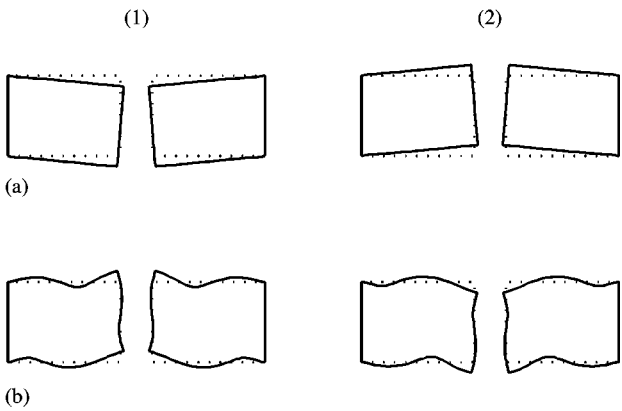

Figure 9. Low-order eigenmodes from the continuum model, shown in coronal cross-section, which are predominantly vertical modes of vibration. They may or may not be important in vocal fold vibration. The natural frequencies of these eigenmodes are (a) 104.6 Hz, (b) 115.8 Hz. Columns 1 and 2 illustrate the eigenmodes at opposite extremes in the vibratory cycle.

occur, in contrast to the relatively large frequency shift required for the two-mass model. Thus, there are many additional possibilities for 1:1 entrainment in the continuum model. This is indicative of a large, unrestricted region where modal phonation may occur in the continuum model.

2.3. *Excised larynx experiments and human subjects*

Fortunately, resonance studies of the vocal folds have not been limited to computer models. Resonance studies of the vocal folds have also been performed on excised larynxes, and on human subjects *in vivo* (Kaneko, Komatsu, Suzuki, Kanesaka, Masuda, Numata & Naito, 1983; Kaneko, Masuda, Shimada, Suzuki, Hayasaki & Komatsu, 1986; Švec, Horáček, Šram & Veselý, 2000). Summarizing the results on excised laryxnes, Kaneko *et al*. (1986) indicated that two prominent resonances appeared in the excised

larynxes, with no tension applied. The average frequencies of the lower and upper resonances were 120 and 180 Hz, respectively.

As noted before, typical Ishizaka & Flanagan (1972) parameters yield resonances of 120 and 201 Hz. Both Kaneko *et al.* (1986) and Ishizaka (1988) suggested that the resonances observed in the Kaneko *et al.* (1986) excised larynx experiments corresponded to the resonances of the two-mass model. However, no data were presented in the excised larynx experiment to indicate that just one eigenmode was present in each resonance, and that inferior–superior tissue was in-phase for the lower resonance, and out-of-phase for the higher resonance.

Furthermore, the *in vivo* investigations of Kaneko *et al.* (1986) yielded somewhat different results. Summarizing the *in vivo* investigations, Kaneko *et al.* (1986) indicated that higher resonance frequencies were observed just below the second harmonic of the lower resonance, suggesting a near-octave spacing between the resonances. This observation was significant because the *in vivo* studies examined a wider range of fundamental frequencies than the excised larynx studies. In a vibrating string, an octave spacing also exists between its first two resonances, i.e., $f_n = nc/(2L)$, where c is the wave speed, L is the length of the string, and f_n is the frequency of the nth harmonic. Note that in this expression, the integer n is an anterior–posterior index just as it was in Equation (4).

A near-octave spacing would exist in the vocal folds if the stiffness parameters in the anterior–posterior direction were substantially greater than the stiffness parameters in the tranverse plane, so that the fundamental frequency was determined principally by the anterior–posterior stiffness. This would create an overall resonance structure similar to that of a vibrating string. This is precisely the situation considered in Table II for the continuum model, in which μ', the shear modulus in the anterior–posterior direction, was 100 times greater than μ, the shear modulus in the transverse plane. Presumably, through activation of the thyroarytenoid muscle and the external tension supplied by the cricoarytenoid muscle, the *in vivo* studies of Kaneko *et al.* (1986) essentially duplicated this condition: the anterior–posterior stiffness of the folds became considerably greater than the stiffness in the transverse plane. Consequently, the resonance frequencies were governed predominantly by the anterior–posterior stiffness.

This interpretation would suggest that the second resonance of the Kaneko study would correspond to an anterior–posterior index of $n = 2$, as predicted by the continuum model. Although no imaging data were available from the Kaneko *et al.* (1986) study to indicate whether or not this was the case, Švec *et al.* (2000) recently duplicated the resonance studies of Kaneko *et al.* (1986) *in vivo*, utilizing an endoscope to image the vibration patterns. On the subject investigated in this study, Švec *et al.* observed three resonances in the folds occurring at 110, 170, and 240 Hz. Presumably, with just a $\frac{1}{2}$-octave spacing between resonances, this was a glottal configuration in which the anterior–posterior stiffness did not clearly dominate the transverse stiffness. Nevertheless, Švec reported that each resonance corresponded to a distinct anterior–posterior index: $n = 1$, 2, and 3, respectively (see Figs 8a, b, and c). Thus, even for less-than-octave spacing, the resonances of the vocal folds corresponded to eigenmodes of distinct anterior–posterior indices, not to the eigenmodes of the two-mass model. However, because imaging along the medial surface of the vocal folds was not performed, it was not possible to determine which (and how many) eigenmodes were present within each resonance peak.

Clearly, more studies are needed to provide a thorough comparison between computer models and human subjects. However, recent investigations with continuum models and human subjects appear to be converging on a similar description of the resonance

structure of the vocal folds. Significantly, the resonance structure of the vocal folds is known to be a critical feature which can either facilitate or discourage eigenmode entrainment. From theory and observation, many vocal fold vibration patterns corresponding to both modal and nonmodal phonation can be explained by the entrainment/disentrainment of just a few underlying eigenmodes. The entrainment/disentrainment of these eigenmodes is one of the distinguishing features of modal and nonmodal phonation.

3. Direct observations of entrainment

3.1. *Computer models*

Although the resonance structure of the vocal folds may provide indications as to which kinds of entrainment may be possible, direct observations of entrainment have also been made along the medial surface of the vocal folds using computer models and excised larynx experiments. In Berry *et al.* (1994), eigenmodes were extracted from a biomechanical simulation of vocal fold vibration. The simulation used a finite element approach to the solution of viscoelastic waves in a continuum (Alipour, Berry & Titze, 2000). Many nonlinearities were incorporated into the model including the nonlinear stress–strain relationship of vocal fold tissues, the nonlinearities due to collision, and nonlinear interactions with the glottal airflow. Also, a three-layer tissue morphology was implemented, including body, cover, and ligament, and more realistic vocal fold shapes were utilized than in the analytic continuum model. Although these complexities prohibited an analytical solution of the eigenmodes, empirical eigenmodes were extracted from the vibration patterns using a standard statistical technique (Berry *et al*., 1994). For a typical case of periodic oscillations, two eigenmodes explained 98% of the variance in the vibration patterns, as shown in Fig. 10. Near the top of the glottal airway, these eigenmodes were qualitatively similar to the eigenmodes shown in Figs 5(a) and 6(a).

Eigenmodes were also extracted from three segments of a bifurcation scenario (a bifurcation is an abrupt jump between distinct oscillation patterns caused by a small change in some parameter in the model). The bifurcations were induced by gradually decreasing the stiffness of the mucosal cover. One segment contained periodic vibrations with a frequency of 160 Hz, another segment contained aperiodic vibrations, and the final segment contained periodic vibrations, but with a fundamental frequency of 80 Hz, corresponding to a period doubling of the original vibration in segment one.

Although the vibration patterns in each segment were quite distinct, the three strongest eigenmodes remained essentially constant throughout the bifurcations. The essential difference between the vibration patterns was the entrainment/disentrainment of the eigenmodes. In the first segment, a 1:1 entrainment existed; in the second segment, the eigenmodes were completely disentrained; and in the final segment, a 2:2 entrainment existed. This observation illustrates that sometimes the only difference between modal and nonmodal phonation may be the entrainment of the underlying eigenmodes.

3.2. *Highspeed, excised larynx experiments*

More recently, eigenmodes have also been extracted from vibrating vocal fold tissues in the laboratory. As described in detail elsewhere (Berry, Montequin & Tayama, in press),

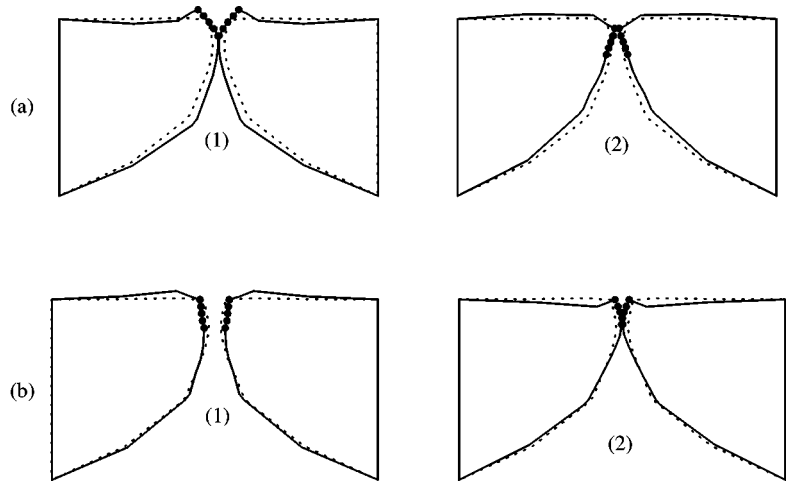

Figure 10. Empirical eigenmodes for the finite element model shown in coronal cross-section (after Berry *et al*., 1994). (a) The first eigenmode, which explains 73% of the variance, alternates between divergent (frame 1) and convergent (frame 2) glottal shapes near the top of the medial surface of the folds. (b) The second eigenmode, which explains 25% of the variance, describes predominantly lateral vibrations near the top of the medial surface of the folds. Columns 1 and 2 illustrate the eigenmodes at opposite extremes in the vibratory cycle.

this was done by modifying a hemilarynx procedure previously investigated by Jiang & Titze (1993). The left vocal fold was mounted against a glass plate. Vibrations were induced by passing airflow through the trachea, and through the glottis (i.e., the area between the glass plate and the left vocal fold). Microsutures were placed along the medial surface of one coronal plane of the tissues, mid-way along the anterior–posterior length of the folds, with approximately a 1.0 mm spacing between sutures. A prism was placed on the opposite side of the glass plate, yielding two oblique views of the medial surface of the folds. Without the oblique views from the prism, lateral vibrations could not have been quantified. Finally, a highspeed camera (a Kodak EktaPro 4540) imaged the vibrations through the two faces of the prism. Vibrations were imaged at a sampling frequency of 4500 frames/s, with a spatial resolution of 256×256 pixels.

Following image capture, a 3D calibration grid was established so that image coordinates could be mapped onto physical coordinates. This was done by removing the left vocal fold from the glass plate, and replacing it with a glass reticle containing a 2D calibration grid. The reticle was mounted against the glass plate using a micrometer. The reticle was imaged. Then, the reticle was perpendicularly displaced from the plate 1.0 mm, and imaged again. This procedure was repeated up to a 7 mm displacement. In this way, an entire 3D grid was imaged from the two views of the prism. Using the direct linear transform (Abdel-Aziz & Karara, 1971), a precise mapping was established from image coordinate to physical coordinates.

One typical cycle of a periodic oscillation is shown in Fig. 11(a). Vibrations of vocal fleshpoints were quantified with sufficient temporal and spatial resolution to compute the underlying eigenmodes. Although Baer (1981) previously tracked 1–3 markers simultaneously in excised larynx experiments, this did not yield sufficient spatial resolution to make an estimation of the eigenmodes. Also, because highspeed imaging was not available, the Baer study was not able to investigate irregular oscillations.

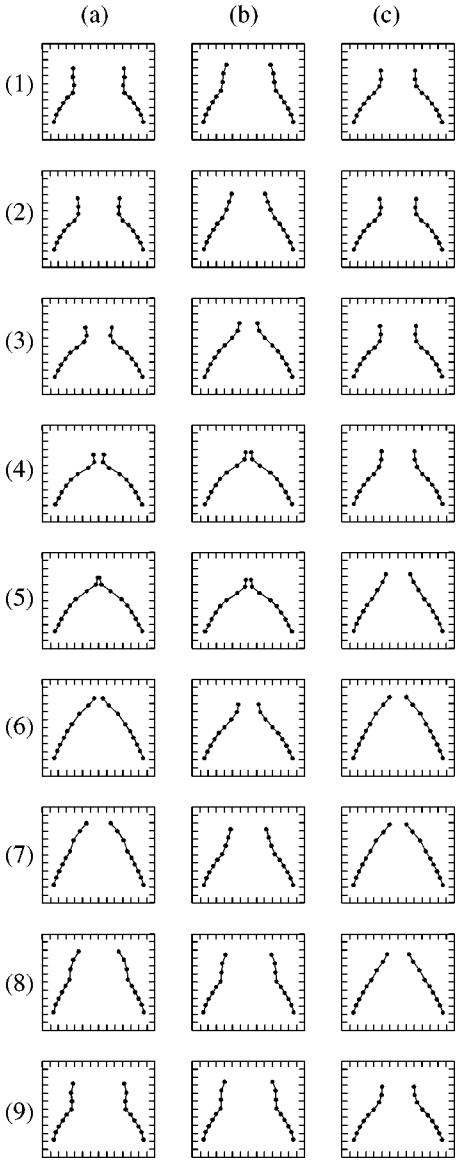

Figure 11. (a) Microsutures attached along the medial surface of a coronal cross-section of the vocal folds during excised larynx experiments, with approximately a 1 mm spacing between sutures. Frames 1–9 illustrate one full cycle of periodic vibration of the vocal fold fleshpoints. Owing to space limitations, only every fifth frame from the highspeed movie is depicted here. Only the left side is imaged; the right side is shown as a reflection to illustrate glottal shape. (b) First eigenmode, which explains 69% of the variance, describes predominantly lateral motion near the top of the glottal airway. (c) Second eigenmode, which explains 29% of the variance, describes an alternating divergent/convergent glottis near the top of the glottal airway. Ticks on the coordinate axes indicate a 1-mm spacing (after Berry *et al*., in press).

For the vibrations under consideration, two eigenfunctions explained 98% of the energy of the vocal fold vibrations. The first eigenmode captured 69% of the energy, and the second eigenmode explained 29%. This is similar to the Berry *et al.* (1994) computational study, in which the two dominant eigenmodes from a finite element model captured 73 and 25% of the energy, respectively. In both studies, two eigenmodes captured 98% of the vibrational energy.

In the hemilarynx experiment, the first eigenmode captured the net lateral vibrations of the vocal fold, as shown in Fig. 11(b). In comparison to the two-mass model (Ishizaka & Flanagan, 1972), this eigenfunction is related to the mode in which both top and bottom masses vibrate in-phase with each other. It is the mode responsible for modulating the glottal airway, and generating the acoustic signal.

The second eigenmode captured the glottal shaping of the vocal folds, alternately shaping a convergent (Fig. 11(c), frames 5–8) and divergent glottis (Fig. 11(c), frames 1–4, and 9). This is especially true for the top two or three fleshpoints. It is known that a divergent glottis possesses a relatively low intraglottal pressure and that a convergent glottis possesses a relatively high intraglottal pressure (Titze, 1988). Thus, this eigenmode is intimately related to the intraglottal pressure. To facilitate self-oscillation of the vocal folds, the energy transfer from the airflow to the vocal folds needs to be maximized (Stevens, 1977; Broad, 1979; Titze, 1988). This condition is met when the intraglottal pressure (in this case, controlled by eigenmode 2) is in-phase with the net lateral velocity of the folds (in this case, controlled by eigenfunction 1). Thus, a specific entrainment or temporal phase relationship between the eigenfunctions is known to facilitate self-oscillation. For example, as the vocal folds approach the glottal midline (Fig. 11(b), frames 1–4), a low pressure condition is created by a divergent glottis (Fig. 11(c), frames 1–4), thus helping the folds to approximate. On the other hand, as the vocal folds separate (Fig. 11(b), frames 5–8), a high-pressure condition is created by a convergent glottis (Fig. 11(c), frames 5–8), thus helping to push the folds apart. Such relationships have already been discussed in relation to eigenfunctions extracted from a finite element simulation of vocal fold movement (Berry *et al.*, 1994). However, such hypotheses concerning physical mechanisms of self-oscillation have never been studied previously on actual vibrating tissues.

Aperiodic oscillations have also been imaged using the highspeed digital imaging system. Following analysis, these data should provide a description of the disentrainment of the eigenmodes. However, these analyses of aperiodic oscillations involve tracking sutures across thousands of frames, and have not yet completed. Nevertheless, the process is being automated, and results will be reported in the future publications.

4. Discussion and conclusions

Both direct and indirect investigations of vocal fold entrainment have been reported. The experiments on the resonance structure of the vocal folds provided an indirect investigation of vocal fold entrainment, i.e., eigenmode entrainment was not explicitly observed in these studies. However, the entrainment patterns most likely to occur during sustained oscillations could be inferred from the natural frequencies of the system. A powerful advantage of the resonance method was that it could be conducted *in vivo* on human subjects (Kaneko *et al.*, 1986; Švec *et al.*, 2000), as well as in excised larynx experiments (Kaneko *et al.*, 1986) and computer models (Berry and Titze, 1996).

Direct observations of vocal fold entrainment/disentrainment have been made in computer models (Berry *et al*., 1994), and in sophisticated highspeed imaging studies of the hemilarynx (Berry & Montequin, 1998; Berry *et al*., in press). In particular, eigenmodes have been extracted from the vibration patterns, yielding explicit examples of entrainment/disentrainment. While these direct studies yield powerful, explicit evidence of vocal fold entrainment, they are too invasive to be used on human subjects. Thus, a combination of both direct and indirect methods would be useful for advancing our understanding of entrainment issues in vocal fold vibration.

How can we apply these techniques to explore specific mechanisms of modal and nonmodal phonation in language? Consider the parameter of glottal stricture mentioned earlier. Using computer models, excised larynx experiments, and human subjects, the eigenfrequencies and eigenmodes of the vocal folds could be computed systematically for various degrees of glottal stricture, ranging from widely separated folds to tightly pressed folds. Using computer models and excised larynx experiments, direct observations of entrainment/disentrainment of the eigenmodes also could be made over these same ranges of glottal stricture, explicitly illustrating examples of disentrainment in nonmodal phonation.

What implications might eigenmode entrainment have in phonetics and language? Presumably, an individual must acquire skill in the entrainment/disentrainment of eigenmodes to convey certain types of linguistic meaning. The low-dimensionality of the vocal system, as suggested by the small number of eigenmodes, may facilitate the development of this skill. As previously has been suggested with regard to music (Titze, 1994), the whole concept of vocal control may be one of learning to entrain/disentrain the vocal eigenmodes. In singing, the concept of formant tuning using the vocal tract is already understood, allowing a singer to produce maximum perceived intensity. Similarly, in phonation, a speaker must learn proper tuning of the vocal fold eigenfrequencies to convey linguistic meaning. Indeed, for nonmodal phonation, many types of complex vibration patterns must be produced.

In summary, eigenmodes and eigenfrequencies may be useful for characterizing the vibrations of the vocal folds, just as they are useful for characterizing the resonances or formants of the vocal tract. Owing to inherent nonlinearities in vocal fold dynamics, eigenmode entrainment plays an important role in vocal fold vibration. Significantly, many types of modal and nonmodal phonation are generated by the entrainment or disentrainment of just a few underlying eigenmodes. Indeed, modal and nonmodal phonation have been distinguished on the basis of this entrainment, where modal phonation corresponds to a 1:1 entrainment, and nonmodal phonation corresponds to all other more complex vibration patterns. Although the investigation of physical mechanisms of modal and nonmodal phonation may still be in its infancy, many concepts from vibration theory, linear dynamics, nonlinear dynamics, and chaos theory have provided a strong theoretical framework for investigation. Certainly, this investigation of physical mechanisms of nonmodal phonation is a rich field of study, complementary to the many perceptual, phonetic, and linguistic investigations which are being performed concurrently.

This work was supported by grant no. R29 DC03072 from the National Institute on Deafness and Other Communication Disorders. The author thanks Dr Bruce Gerratt and the organizers of the Workshop on Nonmodal Phonation for the opportunity to speak on nonmodal phonation from a unique perspective, i.e., the entrainment of the vocal folds. The author also expresses appreciation to his former postdoctoral mentor and friends: Dr Ingo R. Titze, who encouraged further exploration of the composite resonance of the continuum model, and Dr Hanspeter Herzel, for his many insights concerning the nonlinear dynamics of vocal fold vibration.

References

Abdel-Aziz, Y. I. & Karara, H. M. (1971) Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of the ASP/UI symposium on close-range photogrammetry*, American Society of Photogrammetry, Falls Church, VA, pp. 1–18.

Alipour, F., Berry, D. A. & Titze, I. R. (2000) A finite element model of vocal fold vibration, *Journal of the Acoustical Society of America*, **108,** 3003–3012.

Baer, T. (1981) Investigation of the phonatory mechanism, *ASHA Reports*, **11,** 38–46.

Bergé, P., Pomeau, Y. & Vidal, C. (1984) *Order within chaos*, New York: John Wiley & Sons.

Berry, D. A., Herzel, H., Titze, I. R. & Krischer, K. (1994) Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions, *Journal of the Acoustical Society of America*, **95,** 3595–3604.

Berry, D. A. & Montequin, D. W. (1998) Contrasting chest and falsetto-like vibration patterns of the vocal folds, *International Congress on Acoustics*, **4,** 2667–2668.

Berry, D. A., Montequin, D. W. & Tayama, N. (in press) Highspeed, digital imaging of the medial surface of the vocal folds, *Journal of the Acoustical Society of America.*

Berry, D. A. & Titze, I. R. (1996) Normal modes in a continuum model of vocal fold tissues, *Journal of the Acoustical Society of America*, **100,** 3345–3354.

Broad, D. (1979) The new theories of vocal fold vibration. In *Speech and language*: *advances in basic research and practice* (N. Lass, editor). New York: Academic Press.

Chan, R. W. & Titze, I. R. (1999) Viscoelastic shear properties of human vocal fold mucosa: measurement methodology and empirical results, *Journal of the Acoustical Society of America*, **106,** 2008–2021.

Fletcher, N. H. (1996) Nonlinearity, complexity, and control in vocal systems. In *Vocal fold physiology*: *controlling complexity and chaos* (P. J. Davis & N. H. Fletcher, editors), pp. 3–16. San Diego, CA: Singular Publishing Group.

Gerratt, B. R. & Kreiman, J. (2001) Toward a taxonomy of non-modal phonatrics, *Journal of Phoenetics*, 365–381, doi:10.1006/jpho.2001.0149.

Glass, L. & Mackey, M. C. (1988) *From clocks to chaos.* NJ: University Press.

Gordon, M. & Ladefoged, P. (2001) Phonation types: a cross-linguistic overview, *Journal of Phonetics*, **29,** 383–406, doi:10.1006/jpho.2001.0147.

Herzel, H., Berry, D. A., Titze, I. R. & Saleh, M. (1994) Analysis of vocal disorders with methods from nonlinear dynamics, *Journal of Speech Hearing Research*, **37,** 1008–1019.

Hess, W. (1983) *Pitch determination of speech signals*: *algorithms and devices*, pp. 55–56. Berlin: Springer-Verlag.

Hollien, H. (1974) On vocal registers, *Journal of Phonetics*, **2,** 125–143.

Hollien, H. & Michel, J. (1968) Vocal fry as a phonational register, *Journal of Speech Hearing Research*, **11,** 600–604.

Ishizaka, K. (1988) Significance of Kaneko's measurement of natural frequencies of the vocal folds. In *Vocal fold physiology*: *voice production, mechanisms and functions* (Osamu Fujimura, editor), pp. 181–190. New York: Raven Press, Ltd.

Ishizaka, K. & Flanagan, J. L. (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords, *Bell System Technique Journal*, **51,** 1233–1268.

Jiang, J. J. & Titze, I. R. (1993) A methodological study of hemilaryngeal phonation, *Laryngoscope*, **103,** 872–882.

Kaneko, T., Komatsu, K., Suzuki, H., Kanesaka, T., Masuda, T., Numata, T. & Naito, J. (1983) Mechanical properties of the human vocal fold—Resonance characteristics in living humans and in excised larynges. In *Vocal fold physiology*: *biomechanics, acoustics, and phonatory control* (I. R. Titze & R. C. Scherer, editors), pp. 304–317. Denver, CO: The Denver Center for the Performing Arts.

Kaneko, T., Masuda, T., Shimada, A., Suzuki, H., Hayasaki, K. & Komatsu, K. (1986) Resonance characteristics of the human vocal folds *in vivo* and *in vitro* by an impulse excitation. In *Laryngeal function in phonation and respiration* (T. Baer, C. Sasaki & K. Harris, editors), pp. 349–377. Boston: Little, Brown.

Ladefoged, P. (1983) The linguistic use of different phonation types. In *Vocal fold physiology*: *contemporary research and clinical issues* (D. M. Bless & J. H. Abbs, editors), pp. 351–360. San Diego: College-Hill Press.

Peterson, G. E. & Barney, H. L. (1952) Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, **24,** 175–184.

Stevens, K. N. (1977) Physics of laryngeal behavior and larynx modes, *Phonetica*, **34,** 264–279.

Švec, J. G., Horáček, J., Šram, F. & Veselý, J. (2000) Resonance properties of the vocal folds: *in vivo* laryngoscopic investigation of the externally excited laryngeal vibrations, *Journal of the Acoustical Society*, **108,** 1397–1407.

Titze, I. R. (1976) On the mechanics of vocal fold vibration, *Journal of the Acoustical Society of America*, **60,** 1366–1380.

Titze, I. R. (1988) The physics of small-amplitude oscillation of the vocal folds, *Journal of the Acoustical Society of America*, **83,** 1536–1552.

Titze, I. R. (1994) Singing: a story of training entrained oscillators, *Physics News*, **12,** 1–2.