The physics of small-amplitude oscillation of the vocal folds

Ingo R. Titze

Voice Acoustics and Biomechanics Laboratory, Department of Speech Pathology and Audiology,
The University of Iowa, Iowa City, Iowa 52242

(Received 17 November 1986; accepted for publication 31 December 1987)

A theory of vocal fold oscillation is developed on the basis of the body-cover hypothesis. The cover is represented by a distributed surface layer that can propagate a mucosal surface wave. Linearization of the surface-wave displacement and velocity, and further small-amplitude approximations, yields closed-form expressions for conditions of oscillation. The theory predicts that the lung pressure required to sustain oscillation, i.e., the oscillation threshold pressure, is reduced by reducing the mucosal wave velocity, by bringing the vocal folds closer together and by reducing the convergence angle in the glottis. The effect of vocal tract acoustic loading is included. It is shown that vocal tract inertance reduces the oscillation threshold pressure, whereas vocal tract resistance increases it. The treatment, which is applicable to falsetto and breathy voice, as well as onset or release of phonation in the absence of vocal fold collision, is harmonized with former treatments based on two-mass models and collapsible tubes.

PACS numbers: 43.70.Bk, 43.75.Rs

INTRODUCTION

An analytical treatment of small-amplitude vocal fold oscillation is given by Ishizaka and Matsudaira (1972). The treatment begins with a review of pressure-flow relationships in ducts and orifices under steady flow conditions. The authors then show how these pressure-flow relationships produce instability in a soft-walled glottal duct represented by two masses and three springs. A phase lag between the lower mass and the upper mass leads to small-amplitude negative stiffness and damping, and regions of stability and instability are bracketed on the basis of average flow, average glottal opening, and spring constants.

Although the treatment serves as a milestone in quantifying vocal fold vibration, its application to general phonatory theory has been limited because: (1) the key parameters, individual masses and stiffnesses of the coupled oscillators, have been difficult to relate to the anatomical structure of the tissue (and, hence, have been difficult to measure); (2) the oscillation conditions are stated in rather complex mathematical terms because the characteristic equations are fourth order; and (3) some of the results, such as the dependence of fundamental frequency on subglottal pressure, are not in agreement with experimental observation (Baer, 1975). Oscillation conditions based on the two-mass model have been further clarified by Stevens (1977) and Broad (1979), but more work is needed to generalize these conditions to models that incorporate the layered tissue structure of the vocal folds. The more recent collapsible-tube analogy by Conrad (1980,1985) does offer an alternate conceptualization, but it describes a restricted set of vocal tract conditions.

This article is intended to serve two purposes. The first is to present a framework of basic principles by which the mechanics of vocal fold vibration can be understood in limited mathematical terms, and the second is to rework the linearized, small-amplitude analytical treatment with parameters that lend themselves to more direct experimental confirmation. The small-amplitude restriction means that the results apply only to oscillations that build up around slightly abducted (spread-apart) vocal folds, with no glottal closure.

Although this may seem rather restrictive, the small-amplitude approach is useful in establishing the oscillation threshold conditions. The oscillation threshold pressure, which is defined as the lung pressure required to initiate vocal fold vibration, will be determined as a function of vocal fold geometry and viscoelastic properties. We expect that oscillation threshold pressure may have clinical significance in determining the ease with which phonation can be achieved and sustained.

The more general large-amplitude theory, which involves limit cycles and special nonlinear effects due to vocal fold collision, will not be discussed in this article. Since the tools of analysis for large-amplitude oscillation will be primarily numerical, it is appropriate to devote a separate discussion to this follow-up topic. We begin the present small-amplitude discussion with a review of some basic principles of oscillation.

I. BASIC PRINCIPLES OF VOCAL FOLD VIBRATION

The first principle to establish is that sustained vocal fold oscillation is flow induced. The glottal airstream and the yielding duct wall (vocal folds) form a mechanical system that may demonstrate instability under specific flow conditions. If these conditions are met, a continual transfer of energy from the glottal airstream to the tissue will overcome frictional energy losses in the vocal folds. Inertial and elastic properties (mass and stiffness) combine with the geometric properties of the vocal folds to determine the range of oscillation.

A positive flow of energy from the airstream to the tissue can be realized if the net aerodynamic driving force has a component in phase with the tissue velocity. Consider the fundamental equation of motion for a mass–spring oscillator:
\[ M \ddot{x} + B \dot{x} + K x = f(x, \dot{x}, t), \]

where \( M, B, \) and \( K \) are mass, damping, and stiffness, respectively, \( x, \dot{x}, \) and \( \ddot{x} \) are displacement, velocity, and acceleration, respectively, \( f \) is the driving force, and \( t \) is time. If \( f \) is time independent, the differential equation is said to be autonomous. This is the case of interest here because it describes systems that self-oscillate rather than those that are driven into oscillation by an external source.

The dependence of \( f \) on \( \dot{x} \) is crucial for oscillation. Whenever \( f \) is in the direction of the velocity, energy is imparted to the mass (as shown by the hand in Fig. 1). Conversely, whenever \( f \) is opposite to the direction of the velocity, energy is taken out of the mass–spring–damper system. The key to whether the system exhibits self-oscillation or driven oscillation is whether or not the moving hand is considered part of the system. If it is, the system will be governed by an autonomous differential equation. Otherwise, the differential equation is nonautonomous.

It is important to understand the various ways in which the glottal airstream can supply a velocity-dependent driving force. The explanation often given in elementary textbooks (e.g., Lieberman, 1977) that a negative Bernoulli force sucks the vocal folds together prior to closure works only if we pay no attention to what happens on the return path. The same force would be in the opposite direction to the velocity just after glottal opening, thus canceling the impulse imparted prior to closure. Somehow, the system needs to change the effective driving force on alternate quarter-cycles. In Fig. 1, the force \( f \) applied by the finger needs to be reduced (or even reversed) on the return path, when the direction of \( \dot{x} \) reverses.

A driving-force asymmetry on alternate quarter-cycles can be achieved in at least two different ways: (1) by making use of oppositely phased supraglottal or subglottal acoustic pressures or (2) by varying the glottal geometry to create different intraglottal pressure distributions. In the first case, oscillation is facilitated by the inertial properties of the air column in the vocal tract, whereas, in the second case, it is facilitated by nonuniform deformation of the tissue. These two cases are, of course, not mutually exclusive, so that, in normal phonation, both mechanisms may occur simultaneously.

**A. Inertive vocal tract loading**

Consider first the vocal tract loading effect. At fundamental frequencies below the first formant frequency (the usual case for speech), the air column in the vocal tract is primarily inertive; i.e., it acts like a mass of air that is accelerated and decelerated as a unit (Rothenberg, 1981). As the glottis is opening and the glottal flow is increasing, the air column in the vocal tract is being accelerated along with the glottal air. Since a positive acoustic pressure is needed for this acceleration, the vocal tract input pressure \( P_t \) rises above atmospheric pressure [see Fig. 2(a), where the density of stipples relates to magnitude of pressure]. The average intraglottal pressure rises correspondingly, pushing the vocal folds apart. The driving force is thus in the direction of the tissue velocity, and energy is supplied to the vocal folds. When the glottis is closing, on the other hand, the supraglottal air column tends to maintain its forward momentum. This creates a reduced supraglottal pressure \( P_s \), in the wake of the air column [Fig. 2(b)]. The reduced supraglottal pressure (suction) lowers the average glottal pressure, causing the driving force on the tissue to be diminished on this return path. At some point prior to closure, the average intraglottal pressure may actually become negative, in which case the driving force is again in phase with the tissue velocity. This need not happen, however. The only requirement is that the driving force be less positive during closing than opening, and that the net driving force over the entire cycle be sufficient to overcome frictional forces.

The time sequence of events is shown in Fig. 3. Part (a) shows the vocal fold displacement and the corresponding tissue velocity. Since only the open portion of the glottis is primary inertive; i.e., it acts like a mass of air that is accelerated and decelerated as a unit (Rothenberg, 1981). As the glottis is opening and the glottal flow is increasing, the air column in the vocal tract is being accelerated along with the glottal air. Since a positive acoustic pressure is needed for this acceleration, the vocal tract input pressure \( P_t \) rises above atmospheric pressure [see Fig. 2(a), where the density of stipples relates to magnitude of pressure]. The average intraglottal pressure rises correspondingly, pushing the vocal folds apart. The driving force is thus in the direction of the tissue velocity, and energy is supplied to the vocal folds. When the glottis is closing, on the other hand, the supraglottal air column tends to maintain its forward momentum. This creates a reduced supraglottal pressure \( P_s \), in the wake of the air column [Fig. 2(b)]. The reduced supraglottal pressure (suction) lowers the average glottal pressure, causing the driving force on the tissue to be diminished on this return path. At some point prior to closure, the average intraglottal pressure may actually become negative, in which case the driving force is again in phase with the tissue velocity. This need not happen, however. The only requirement is that the driving force be less positive during closing than opening, and that the net driving force over the entire cycle be sufficient to overcome frictional forces.

The time sequence of events is shown in Fig. 3. Part (a) shows the vocal fold displacement and the corresponding tissue velocity. Since only the open portion of the glottis is primarily inertive; i.e., it acts like a mass of air that is accelerated and decelerated as a unit (Rothenberg, 1981). As the glottis is opening and the glottal flow is increasing, the air column in the vocal tract is being accelerated along with the glottal air. Since a positive acoustic pressure is needed for this acceleration, the vocal tract input pressure \( P_t \) rises above atmospheric pressure [see Fig. 2(a), where the density of stipples relates to magnitude of pressure]. The average intraglottal pressure rises correspondingly, pushing the vocal folds apart. The driving force is thus in the direction of the tissue velocity, and energy is supplied to the vocal folds. When the glottis is closing, on the other hand, the supraglottal air column tends to maintain its forward momentum. This creates a reduced supraglottal pressure \( P_s \), in the wake of the air column [Fig. 2(b)]. The reduced supraglottal pressure (suction) lowers the average glottal pressure, causing the driving force on the tissue to be diminished on this return path. At some point prior to closure, the average intraglottal pressure may actually become negative, in which case the driving force is again in phase with the tissue velocity. This need not happen, however. The only requirement is that the driving force be less positive during closing than opening, and that the net driving force over the entire cycle be sufficient to overcome frictional forces.

The time sequence of events is shown in Fig. 3. Part (a) shows the vocal fold displacement and the corresponding tissue velocity. Since only the open portion of the glottis is
considered here, the displacement and tissue velocity curves are not drawn during collision. Note that the tissue velocity is positive during outward movement and negative during inward movement. In magnitude, it is zero at the peak excursion and maximum before and after impact of the opposing vocal folds. Thus the tissue velocity is a monotonically decreasing function over the entire open portion of the cycle.

Figure 3(b) shows the airflow corresponding to the displacement and velocity in Fig. 3(a). The shape of the airflow has been discussed and modeled in detail (Rothenberg, 1981; Fant, 1983). As indicated by the solid line, the slower rise and the more abrupt fall (i.e., the skewing of the waveform to the right) are a result of the inertia of the air in the vocal tract. When the glottis opens, positive supraglottal pressure is needed to accelerate the air column above the glottis. This takes time, and, hence, the buildup of flow is delayed with respect to the movement of the folds. Note that the peak airflow comes later in time than the peak displacement. As a reference waveform, the dashed line indicates the mean airflow without air inertia, which follows the displacement exactly.

The deceleration of the flow is also delayed by the inertia of the air, but the sudden collapse of the glottal area at closure causes the flow to collapse suddenly as well. The resulting asymmetry in the flow waveform is a key factor in obtaining a velocity-dependent driving force.

Consider now the air particle velocity in the glottis. The shape of the air particle velocity waveform [Fig. 3(c)] is obtained by dividing the flow waveform in (b) by the displacement waveform in (a), since the displacement is proportional to glottal (duct) area. The important thing to note, again, is that the particle velocity is asymmetric around the peak tissue displacement, displaying a gradual increase over the open portion of the cycle. Without air inertia, the air particle velocity would be approximately constant, as shown by the dashed line in Fig. 3(c).

The waveform for the mean glottal pressure \( P_g \), and hence the driving force, can now be estimated on the basis of Bernoulli’s energy law. For negligible kinetic pressure in the supraglottal region, \( P_g = P_t - 1/2 \rho v^2 \), where \( P_t \) is the subglottal pressure. This relationship is shown in Fig. 3(d). It is seen that a monotonic downward trend in the driving pressure over the open portion of the cycle follows the tissue velocity in Fig. 3(a), suggesting a strong velocity-dependent driving force.

In differential form, \( \Delta P_g = -\rho v \Delta v \) for constant subglottal pressure. The key to instability in the duct wall (and, hence, the key to oscillation) is the negative sign in the above equation. Although it seems somehow counterintuitive that an increasing particle velocity should decrease the pressure on the wall, this is the basic characteristic in all flow-induced wall vibration phenomena. If the sign were positive, the mean driving pressure would increase with increasing particle velocity and decreasing tissue velocity. This opposite phase between driving force and tissue velocity would add a damping force similar to the frictional damping force.

Note, however, that the existence of the negative differential Bernoulli pressure alone does not guarantee negative damping. If air inertia is not present, as shown in the dashed lines of Fig. 3(d), whatever is gained during opening is lost during closing. Without the asymmetry caused by delayed response, no net energy transfer can take place.

Much has been made of the Bernoulli effect and negative glottal pressures in phonatory theory. It is not at all necessary that the pressure ever become negative. What is important is that more energy is imparted by the flow than is lost by friction in the tissue. Later quantitative analysis will reveal when this is the case. At this point, it is important to recognize that energy gain is facilitated by the combination of negative differential Bernoulli pressure and delayed vocal tract response, i.e., an inertive acoustic load.

**B. Nonuniform tissue modes**

Consider now the second, distinctly different, condition under which velocity-dependent driving forces can be achieved. If the tissue has a normal mode of vibration that results in different glottal shapes over different portions of the cycle, and if the different glottal shapes produce different glottal pressure profiles, a force asymmetry can also be established. This is illustrated in Fig. 4. In order not to complicate the explanation, assume that in this case there is no vocal tract loading. This makes the supraglottal acoustic pressure \( P_t \) approximately zero (atmospheric). Furthermore, the airstream emerges from the glottis in the form of a jet that experiences little or no aerodynamic pressure recovery upon exit (Ishizaka and Matsuda, 1972; Scherer, 1981). Within the glottis, however, the pressure varies with glottal area. If we assume Bernoulli’s energy conservation law to be applicable (at least approximately), the glottal
pressures will rise above zero from top to bottom for a convergent glottis [Fig. 4(a)], whereas the glottal pressures will fall below zero from top to bottom for a divergent glottis [Fig. 4(b)]. (Recall that, according to Bernoulli’s energy principle and the flow continuity principle, an increase in duct area is accompanied by an increase in duct pressure, and vice versa.) This establishes, at once, the desired driving pressure asymmetry. Since the average pressure in the glottis is greater for a convergent shape than for a divergent shape, a velocity-dependent driving force can be realized, and energy can flow from the airstream to the tissue.

But we need to be convinced that convergent and divergent shapes are likely to occur on alternate quarter-cycles as shown. The evidence comes from high-speed cinematography (Farnsworth, 1940) and studies on excised larynges (Baer, 1975). It has been observed repeatedly that the vocal folds vibrate with this vertical phase difference. Furthermore, normal mode analyses (Titze and Strong, 1975; Titze, 1976) have predicted an entire series of modes with horizontal and vertical phase differences if the tissue is treated as a viscoelastic medium bounded by rigid structures (e.g., cartilages). In the lowest mode (having the lowest natural frequency), the tissue moves in phase vertically, as in Fig. 2. In the second mode, top and bottom move 180° out of phase, and the center is not moving at all. The combination of these two modes, which results in the frequently observed pattern shown in Fig. 4, exhibits a phase difference less than 180°, typically about 60°–90° (Baer, 1975).

We will show in this article that the vertical phase difference can be derived by assuming a simple one-dimensional surface wave in the form of an indentation (or depression) propagating on the vocal fold surface from the bottom to the top [note upward arrows in Fig. 4(a) and (b)]. This surface wave can represent a small displacement of respiratory mucus external to the skin, or a larger displacement of moist, jellylike mucosal tissue under the skin (or both) traveling upward in the glottis at a velocity \( c \). A few millimeters into the tissue (laterally), the surface wave is attenuated sufficiently so that the remainder of the vocal fold can be treated as a single mass, or neglected completely.

The delayed action of the tissue at the top of the vocal folds in relation to the bottom produces an effect equivalent to delayed action by air inertia in the vocal tract. The fact that the upper margins of the folds take longer to open allows more pressure to build up at the lower margins for a convergent glottis, forcing the tissue apart. Conversely, air inertia can be thought of as a “delayed opening of the vocal tract,” allowing for higher intraglottal pressures on glottal opening. Similar arguments can also be made for glottal closing, with the intraglottal pressures being lowered by “delayed closing of the vocal tract” or upper margins of the tissue. Figure 3(d) is qualitatively the same, therefore, whether the oscillation mechanism involves nonuniform tissue modes or inertive vocal tract loading (or both).

C. Oscillation threshold pressure

The results of our mathematical analysis will show that a region of growing oscillation can be defined in terms of an oscillation threshold lung pressure that depends on tissue damping, the mucosal wave velocity, vocal fold thickness, prephonatory glottal width, and a prephonatory glottal convergence factor. As in the Ishizaka and Matsudaira (1972) theory, small-amplitude oscillation will be assumed around the mean prephonatory width. It will be demonstrated that lower oscillation threshold pressures occur with lower surface-wave velocities, with tighter adduction (bringing together of the folds), and with a divergent rather than a convergent prephonatory glottis. Lower surface-wave velocities are congruent with larger vertical phase differences, as we will also demonstrate.

With regard to vocal tract loading, we will show that the oscillation threshold pressure is affected by two primary variables, supraglottal inertance \( I_2 \) and supraglottal resistance \( R_2 \). Increases in \( I_2 \) will lower the threshold pressure, whereas increases in \( R_2 \) will raise it. Thus the vibration characteristics of the source are dependent on the resonator, but this dependence is likely to be strong only when inadequate tissue mobility in the mucosa limits the propagation of a surface wave.

II. ANALYTICAL TREATMENT OF VOCAL FOLD OSCILLATION

For analytical (noncomputational) treatment of vocal fold oscillation, a simplified body-cover model of the vocal folds is shown in frontal section in Fig. 5. The body of the
A. Linearized glottal area function based on surface waves

In Fig. 5, consider $A_1$ and $A_2$ to be subglottal and supra-glottal areas, and $a_1$, $a_2$, and $a_3$ to be glottal areas at entry, midglottis, and exit, respectively. We assume that $A_1$ and $A_2$ are constant, but that glottal areas are time and space dependent. In particular, we assume that the area anywhere along the glottis is

$$a(z,t) = 2L \left[ \xi_0(z) + \xi_1(t) \right],$$

where $z$ is measured from the midpoint of the glottis in the direction of flow, $L$ is the length of the vocal folds normal to the plane of the paper, $\xi_0(z)$ is the prephonatory glottal half-width, and $\xi_1(t)$ is the time-dependent displacement of the cover.

Hirano (1975) has shown that mucosal surface waves propagate along the glottis in the direction of airflow. In the simplest way, these waves can be described by a one-dimensional wave equation with wave velocity $c$, such that

$$\frac{\partial^2 \xi_1}{\partial t^2} = \frac{c^2 \partial^2 \xi_1}{\partial z^2},$$

This has the general d'Alembert solution,

$$\xi_1(z,t) = \xi_1(t - z/c),$$

which can be verified by simple substitution.

It is seen from Eq. (4) that mucosal wave propagation causes a time delay in movement from bottom to top of the vocal folds. We will demonstrate that this time delay helps to provide some of the necessary instabilities for vocal fold oscillation.

Let us assume a linear (trapezoidal) $z$ dependence of the prephonatory glottis (dashed lines in Fig. 5):

$$\xi_0(z) = (\xi_01 - \xi_02)/2 - (\xi_01 - \xi_02)z/T,$$

where $\xi_01$ and $\xi_02$ are inferior and superior glottal half-widths and $T$ is the vocal fold thickness.

With regard to the surface-wave displacement, we expand the general solution (4) in a Taylor series around the midpoint ($z = 0$) of the glottis and keep only the first-order (linear) term,

$$\xi_1(t - z/c) \approx \xi_1(t) - z / c \left( \frac{\partial \xi_1}{\partial t} \right)_{z=0} + \cdots$$

The purpose of this expansion is to obtain an explicit velocity dependence in the glottal area $a(z)$ in Eq. (2). In order to determine the validity of the expansion, consider as an expression of $\xi_1$ the harmonic wave solution $\sin \omega (t - z/c)$. By expanding the compound angle into the form $(\sin \omega t) (\cos \omega z/c) - (\cos \omega t) (\sin \omega z/c)$ and assuming small-angle approximations ($\sin \omega z/c \approx \omega z/c$ and $\cos \omega z/c \approx 1$), the solution takes on the form $\sin \omega t - (z/c) \omega \cos \omega t$. This is an explicit version of Eq. (7). The small-angle approximation is not ideal for numerical purposes because it results in errors greater than 10% for $\omega z/c > 0.75$. At a frequency of 100 Hz and a surface-wave velocity of 100 cm/s (Baer, 1975), the approximation would be crude for
vocal fold half-thickness \((z = \pm \, T/2)\) of more than about 1 mm. But the expansion is conceptually very illuminating and analytically very convenient. None of the conclusions that we draw will be affected by a 10% numerical error.

Continuing with the example, the phase delay between the upper and lower margins for \(T/2 = 1\) mm would be 80 ø (40 ø from center to top or bottom). This is in the range of predictions made by Ishizaka and Matsudaira (1972) and measurements made by Baer (1975).

The glottal areas at entry and exit can now be defined by letting \(z = \pm \, T/2\) and introducing a time delay \(\tau = T/2c\).

\[
\begin{align*}
a_1 &= 2L(\xi_{01} + \xi + \tau_D), \\
a_2 &= 2L(\xi_{02} + \xi - \tau_D).
\end{align*}
\]

B. Displacement flow, mean glottal pressure, and equation of motion

Consider now the lumped-element flow circuit shown in Fig. 6. A similar circuit was proposed by Conrad (1980) and in our preliminary study (Titze, 1985). The subglottal pressure is \(P_s\), the glottal entry pressure is \(P_1\), the glottal exit pressure is \(P_2\), and the supraglottal pressure (input pressure to vocal tract) is \(P_i\). The nonlinear flow resistances in the contraction and expansion regions are represented by \(R_c\) and \(R_e\), respectively, and the intraglottal resistances are \(R_{g1}\) and \(R_{g2}\), also nonlinear. The flow at exit is \(u\) and the flow at entry is \(u + u_d\), where \(u_d\) is the displacement flow, resulting from the yielding glottal wall. Because all displacements in Eqs. (9) and (10) are defined at the midpoint of the glottis, the mechanical properties of the tissue can also be lumped at the midpoint of the glottis by way of a net vocal fold compliance \(C_f\), and inertance \(I_f\), and a resistance \(R_f\). As will be seen momentarily, these flow circuit elements are directly related to effective stiffness, mass, and damping of the vibrating vocal fold cover.

Since vocal fold displacement has been linearized along the glottis [Eq. (7)], the mean displacement flow is at the midpoint and can be written as

\[
u_d = 2LT\dot{z}.
\]

However, the pressure variation within the glottis is not linear. A mean glottal pressure \(P_g\), which serves as the net driving pressure for the entire vocal fold tissue, can be computed if an intraglottal pressure function \(P(z)\) is known. This can then be related to the tissue response,

\[
P_g = \frac{1}{T} \int_{-T/2}^{T/2} P(z)dz = 2LT \left( I_{g1} + R_{g1} + I_{g2} + R_{g2} \right) C_f
\]

where \(M\), \(B\), and \(K\) are the effective mass, damping, and stiffness per unit area of the vibrating portion of the cover, lumped at the midpoint of the glottis. The area in question is the medial surface area \(LT\), over which the net driving pressure acts. We will show that the above equation of motion contains the necessary elements for oscillation, as proposed in Eq. (1). First, however, we focus our attention on the intraglottal pressure variation \(P(z)\). Since it is not a linear variation, it is incorrect to call the average pressure \((P + P_0)/2\), nor is it correct to divide the total glottal resistance into two equal parts above and below the midpoint, as is done in linear circuit analysis with a T section. Figure 6, therefore, serves more for conceptualization of the flow problem than for exact loop equation analysis.

C. Transglottal pressure and mean driving pressure

We assume that the results of investigations by van den Berg et al. (1957), Ishizaka and Matsudaira (1972), Scherer (1981), and Gauffin et al. (1983) on steady pressure-flow profiles in static models of the glottal airway are applicable to oscillatory flow. Their results indicate that there are two different flow regions. At glottal entry and throughout the glottis, the flow is basically nonturbulent and describable by a (modified) Bernoulli equation. At glottal exit \((z = T/2)\), the flow becomes turbulent and energy losses are substantial. Less than 20% of the kinetic pressure at glottal exit is recovered after expansion from area \(a_1\) to area \(a_2\) (Fig. 5). The transglottal pressure can be approximated by

\[
P_i - P_s = (k_e - k_r)\rho u |u|/a_2^2 = k_eP_{k_e},
\]

where \(k_e\) is a kinetic pressure loss coefficient for the preturbulent region (entry and glottis), \(k_r\) is a pressure recovery coefficient for the turbulent (expansion) region, \(\rho\) is the air density, \(u\) is the exit flow, \(k_e = k_c - k_r\) is the transglottal pressure coefficient, and \(P_{k_e}\) is the kinetic pressure at exit. Viscous resistance in the glottis is considered here to be part of the empirically determined coefficient \(k_e\). The values of the pressure coefficients have been estimated by the above investigators to be

\[
1.0 < k_e < 1.4,
\]

\[
k_e = 2(a_1/A_1)(1 - a_3/A_3) < 0.2,
\]
with the lower limit for $k_c$ applying to the highly convergent glottis and the upper limit applying to the rectangular glottis. A value of $k_c = 1.0$ implies that all the pressure drop is kinetic (no energy loss) up to glottal exit. Given the relatively small range of $k_c$, it is clear that not much energy is lost prior to exit. The equation for $k_c$ was derived by Ishizaka and Matsudaaira (1972) by application of the momentum conservation principle. Over most of the glottal cycle, $k_c$ is less than 0.1, suggesting that almost no kinetic pressure is recovered after exit.

Given that energy losses are mainly downstream of the glottis, the intraglottal pressures can (to first order) be derived from Bernoulli's energy equation

$$P(z) = P_2 + \frac{P_{g2}}{k} \left[ 1 - a^2/a^2(z) \right],$$

where $P_g$ is the exit pressure. The average of this intraglottal pressure, as indicated by integration on the left side of Eq. (12), is

$$P_g = P_2 + \frac{P_{g2}}{k} \left[ 1 - \frac{a^2}{a^2} \left( -a^{-1} \left( \frac{\partial a}{\partial z} \right)^{-1} \right) + \frac{T^2}{T^2} \right].$$

In obtaining this integration, it has been recognized that the glottal area gradient $\partial a/\partial z = (a_2 - a_1)/T$ is independent of $z$ because of the assumed linear glottis.

Upon further evaluation of the limits in Eq. (17) and substitution of the linear area gradient into the equation, we obtain

$$P_g = P_2 + \frac{P_{g2}}{k} (1 - a_2/a_1).$$

Since the exit recovery pressure is

$$P_2 - P_i = k_p P_{g2},$$

we can substitute $P_2$ from (19) into (18) and replace the exit kinetic $P_{g2}$ by $(P_2 - P_i)/k_c$ from Eq. (14). This yields an important relationship between average driving pressure of the vocal folds and transglottal pressure:

$$P_g = P_i + (P_2 - P_i) (1 - a_2/a_1 - k_c)/k_c.$$ 

It is often tacitly assumed in phonetic science that the vocal folds are driven by the transglottal pressure. This is true only if the vocal tract input pressure is zero, if the glottis is highly convergent, and if there is no exit pressure recovery. For such a rather trivial case, the transglottal pressure becomes the subglottal pressure. In general, the notion of transglottal pressure driving the vocal folds should be abolished.

Note that the effective driving pressure is a function of glottal geometry, as discussed qualitatively in Sec. I. During glottal opening, $a_2/a_1$ is smaller than during glottal closing, which sets up the asymmetry in the driving force discussed earlier. As a rectangular glottis is approached, the driving pressure can become small if the vocal tract input pressure $P_i$ is small. This can happen in the face of a large transglottal pressure. Negative driving pressures can also occur, especially if $P_i$ goes negative (as it does prior to closure when the supraglottal air inertia creates a suction above the folds) or if the glottis diverges.

D. Oscillation conditions for no vocal tract coupling

When the subglottal pressure equals the lung pressure ($P_s = P_i$) and the vocal tract input pressure is atmospheric ($P_j = 0$), the driving pressure becomes

$$P_g = P_i (1 - a_2/a_1 - k_c)/k_c.$$ 

This is an interesting case because it can be simulated with relative ease in the laboratory on excised larynges. The lung pressure is the only relevant variable because both subglottal and supraglottal pressures are fixed. We consider here the typical case where the supraglottal area is large ($a_2 \ll A_2$, or $k_c \approx 0$). Combining Eqs. (9) and (10) with (21) and substituting the result for $P_g$ into the left side of Eq. (12) yield the equation of motion,

$$P_2/k_c (\xi_{g01} - \xi_{g02} + 2\tau \xi + \tau^2 \xi')^{-1} = M \xi + B \xi + K \xi.$$ 

This is an autonomous nonlinear differential equation, i.e., a specific form of Eq. (1) without the explicit time dependence $t$. Note that the driving pressure has a component that varies directly with tissue velocity. Even though the velocity appears in both the numerator and the denominator, its presence in the numerator dominates because of the factor of 2. It is clear that the driving pressure is greater for positive $\xi$ than for negative $\xi$, which is in harmony with the more general discussion of Sec. I.

In order to get simple relationships for oscillatory conditions, we consider small-amplitude vibration by letting

$$\xi = \xi + \xi_0,$$

where $\xi_0$ is a static (mean) displacement and $\xi$ is a small oscillatory component. The static component satisfies the quadratic equation

$$\ddot{\xi}_0 = (P_2/k_c) (\xi_{g01} - \xi_{g02})/(\xi_{g01} + \xi_0) = K_0 \xi_0,$$

which is obtained by letting $\xi = 0$ in Eq. (22) and $\xi_0 \neq 0$ in Eq. (23). Solution is obtained by formula

$$\xi = - (R_0/2) + [(R_0/2)^2 + (P_2/k_c) (\xi_{g01} - \xi_{g02})/K_0]^{1/2}.$$

Since positive lung pressures must result in positive displacement for a convergent glottis, the negative sign before the bracketed term can be discarded as being nonphysical.

We assume now that the combined prephonatory and static displacements are much larger than the dynamic (oscillatory) displacements. Specifically, we expand the third factor in the left side of Eq. (22) binomially around $\xi_{g01} + \xi$, assuming that $(\xi_{g01} + \xi)^2 \gg (\xi + \tau \xi')^2$, multiply it by the second factor, and retain only first-order (linear) terms. This results in constant coefficients of $\xi$ and its derivatives in Eq. (22). Effective small-amplitude damping $B^*$ and stiffness $K^*$ per unit area can then be defined by combining all the displacement and velocity coefficients of the resulting linearized second-order differential equation. The coefficients are

$$B^* = B - (P_2/k_c) \tau (\xi_{g01} + \xi_{g02} + 2\xi)/ K^* \approx K + (P_2/k_c) (\xi_{g01} - \xi_{g02})/(\xi_{g01} + \xi_0)^2 = K + K' = K + \xi_0^2.$$
We will refer to $B'$ and $K'$ as aerodynamic damping and stiffness, respectively. The mass per unit area of the vocal fold tissue is unaltered (i.e., $M^* = M$).

1. Oscillation threshold pressure

It is clear from Eq. (26) that oscillation will grow (or be sustained) when there is negative (or zero) damping, i.e., when $B^* < 0$. Replacing $\tau$ by $T/2c$, the oscillation condition is

$$P_L > (2kT/(Bc))(\xi_{01} + \xi)^2/((\xi_{01} + \xi_{02} + 2\xi)).$$

We see that the oscillation threshold pressure, defined by the equality, is reduced by decreasing the tissue damping, by decreasing the mucosal wave velocity, by increasing the thickness of the vocal folds, or by decreasing the inferior glottal width (note the squared term involving $\xi_{01}$ in the numerator).

The product $Bc$ in Eq. (28) will be referred to as the effective damping pressure. An increase in viscous damping in the tissue can be overcome by more time delay (lower wave velocity) in the mucosa, or vice versa. In Fig. 7(a), oscillation threshold pressure $P_L$ is plotted as a function of effective damping pressure $Bc$ for four different convergence angles depicted in Fig. 7(c). The mean (midpoint) glottal half-width is held constant at 0.10 cm, and $\xi_{01}$ and $\xi_{02}$ are varied systematically from 0.04 to 0.16 cm in opposite directions. Convergence angle increases from configuration 1 to 4, the first two configurations being in fact divergent. Regions of growing oscillation in Fig. 7(a) are above the threshold (solid) lines, and regions of damped oscillation are below the threshold lines.

Note that increased convergence of the glottis reduces the region of oscillation. By way of an example, configuration 2 (slightly divergent) can sustain oscillation at 2.15-cm H$_2$O lung pressure with an effective damping pressure of 10 cm H$_2$O. Configuration 3, on the other hand (slightly convergent), requires 5.54-cm H$_2$O lung pressure at the same damping pressure [note dots in Fig. 7(a)]. The more extreme cases, configurations 1 and 4, require 0.41 and 9.94 cm H$_2$O, respectively, at the same effective damping pressures. This is a large range of values and is likely to be of major significance in phonation.
The calculation of threshold damping pressures in Fig. 7(a) required the quadratic solution (25) for the static displacement $\xi$. This solution is plotted separately in Fig. 7(b) for the same four configurations. (Sinusoidal lines will be discussed later.) Note that the static displacement (plotted on the horizontal axis) is positive for the convergent shapes and negative for the divergent shapes. This is a direct consequence of the static Bernoulli pressures that either "bulge out" or "suck in" the medial surface to establish the true equilibrium position for oscillation. The static displacement in Fig. 7(b) adds to the prephonatory displacement shown in Fig. 7(c), although they are not drawn to the same scale.

The vocal fold stiffness $K$ per unit area was chosen to be 200 kdyn/cm$^2$. This agrees with the Ishizaka and Flanagan (1972) value of 80 kdyn/cm for a medial surface area of 0.4 cm$^2$ (about 1.4-cm length x 0.3-cm thickness). In addition, the transglottal pressure coefficient $k$, in Eq. (25), was chosen to have an average value of 1.1. For these values of tissue stiffness, geometry, and pressure loss, the static displacement is less than $\pm 0.05$ cm over the entire range of lung pressures (0-20 cm H$_2$O) shown in Fig. 7(b). The curves include only stable (real) solutions of Eq. (25). For divergent shapes, the static Bernoulli pressure $P_x$ in Eq. (24) is negative and can in magnitude exceed the linear spring stiffness $K\xi$, resulting in no stable equilibrium. Complete collapse of the glottis can then occur, which is described mathematically by a complex solution of Eq. (25) and corresponds to imaginary points to the left of curves 1 and 2 in Fig. 7(b). These solutions have no physical significance here because they run counter to the basic small-oscillation assumptions, i.e., the existence of a stable equilibrium position. In large-amplitude oscillations, there is always a limiting function, such as nonlinear stiffness or vocal fold collision, that prevents the runaway problem faced by unstable equilibrium. Furthermore, in the human larynx, there is never an abrupt change in cross-sectional area from the trachea to the glottis, as shown in Fig. 7(c). Rather, there is a gradual tapering in this region; hence, the net prophonatory convergence is limited. These and other limiting functions will be discussed in the subsequent article on large-amplitude oscillation and limit cycles.

The smallest value of $\xi$ in Fig. 7(c) is 0.04 cm, which is about the largest value of $\xi$ in Fig. 7(b). One might therefore assume that for most conditions $\xi$ is negligible in comparison to $\xi_{01}$. This simplifies Eq. (28) considerably and yields the approximate oscillation threshold pressure
\[ P_L \approx (2k/\tau)(Bc)\xi_{01}^2/(\xi_{01} + \xi_{02}). \]  

The dashed lines in Fig. 7(a) represent this approximate solution, which clearly asymptotes to the exact solution for lung pressures typical in speech. The differences between the solid lines and the dashed lines represents the shift in oscillation pressure due to the small static displacement. Note that the shift is upward for convergent shapes and downward for divergent shapes. This is a type of positive feedback and accounts for the runaway (collapse) situation mentioned earlier for large oscillation. The lower the tissue stiffness, the greater this effect will be. Note the inverse relationship between $\xi$ and $K$ in Eq. (25).

An important test of the validity of the small-amplitude approximations is to solve the nonlinear differential Eq. (22) numerically and compare the solution with the predicted results of Fig. 7(a) and (b). This can be done easily with the use of conventional differential equation solvers (e.g., fourth-order Runge-Kutta techniques) if the second-order equation is replaced by two first-order equations with variables $f_1 = \xi$ and $f_2 = \dot{\xi}$. Results of this numerical solution are overlaid in Fig. 7(b) for glottal configurations 2 and 3. The total displacement $\xi = \xi + \dot{\xi}$ is shown for each of the two configurations. Arrows point to the dots from which the corresponding lung pressures and damping pressures were taken. Note, first of all, that the oscillations are neither growing nor damped. This suggests that the chosen lung pressures, 2.15 and 5.54 cm H$_2$O from Fig. 7(a) at $Bc = 10$ cm H$_2$O, were indeed the threshold values. Oscillations grew or damped out as expected for pressures above or below threshold (not shown). Note, also, that the static (offset) displacements of the sinusoids exactly match the static displacement predicted by the $\xi$ curves (see arrows). Finally, the amplitudes of the sinusoidal displacements $\xi$ are less than 0.01 cm. Since the smallest corresponding equilibrium displacement $(\xi_{01} + \xi_{02})$ is 0.08 $\pm 0.01$ cm, the binomial expansion around $(\xi_{01} + \xi_{02})$ in the linearization process of Eq. (22) is basically validated. [Recall that the criterion was $(\xi_{01} + \xi_{02})^2 \gg \xi^2$.]

Consider now the oscillation threshold pressure for the rectangular glottal configuration. It is evident from Eq. (25) and Fig. 7(b) that $\xi_{01} + \xi_{02}$ would yield no static displacement ($\xi = 0$). The oscillation threshold pressure would then simplify to
\[ P_L = (k/\tau)(Bc)\xi_{01}, \quad \xi_{02} = \xi_{01}. \]  

This clarifies the importance of the prephonatory glottal width for oscillation threshold. The closer the vocal folds are brought together, the easier it is to initiate small-amplitude oscillation. Figure 8 shows this graphically. Four values of glottal half-width are chosen for the rectangular glottis. If we again assume, for sake of an example, that the damping pressure $Bc$ is 10 cm H$_2$O for all adjustments, the oscillation threshold pressure varies from 1.5 cm H$_2$O for $\xi_{01} = \xi_{02} = 0.04$ cm to 5.9 cm H$_2$O for $\xi_{01} = \xi_{02} = 0.16$ cm (see dots).

It is interesting to note that this range of threshold pressures, although significant, is smaller than the range achieved by changing the convergence angle. Thus, with reference to Fig. 7(c), a greater range of oscillation threshold pressures is obtained by changing between configurations 1 and 4 than by moving the entire fold in a rectangular fashion between the extreme 0.04- and 0.16-cm positions.

2. Relationship to other theories

The rectangular case provides an excellent opportunity to compare our results with those of Ishizaka and Matsuda (1972). We can harmonize our theory with theirs by adopting their equation of motion (on p. 52 of their monograph) for the lower mass $m_1$. Recalling a damping coefficient $\gamma$, a stiffness coefficient $s_1$, and a driving force that is proportional to the difference in displacement between upper and lower masses, their equation of motion was

\[ m_1 \ddot{y} + s_1 \dot{y} + \gamma (y_1 - y_2) = 0. \]
One of the conditions for oscillation in the Ishizaka and Matsudaira formulation was that the coupling stiffness $s_c$ between the masses must be less than the aerodynamic stiffness $\phi$. This is equivalent, in our formulation, to saying that the mucosal wave velocity $c$ must be less than $2T\phi/\tau_1$. Otherwise, the delay time $\tau$ in Eq. (33) is too small for the aerodynamic resistance $4\phi$ to overcome the resistance $r_1$.

Conrad (1980) proposed that vocal fold oscillations are analogous to oscillations in a collapsible tube. These oscillations have been observed in arterial systems, where a downstream flow constriction is created by incomplete clamping of an artery. This condition can be approximated in our formulation by letting $a_2 = 2L_a/\xi_0 = $ constant, i.e., by clamping the upper lips of the vocal folds at a fixed value. Although this puts the downstream "constriction" very close to the vibrating tissue (immediately above it, as a movement constraint), the concept is similar to having a constriction at the false folds or further downstream, since Conrad's model is a lumped-element model without transit time delays. From Eq. (10) it is clear that $\xi = \tau_2^0$ in this case, which simplifies Eqs. (22) to (27) considerably. The driving pressure is no longer an explicit function of $\xi$, but only of $\xi_2$ and the static displacement $\xi_2$ assumes the form $(P_L/\kappa)(\xi_2 - \xi_0)/((K_s/\kappa_2)\xi_2)$.

Although the constraining (clamping) of a downstream portion of the glottal duct with yielding walls appears to produce oscillation, it is neither necessary nor a particularly appropriate condition for the vocal folds. Unless the false vocal folds were to be used to create the downstream constriction, nothing anatomical in the larynx exists that would easily and consistently maintain such a constraint. In fact, it is a small matter to demonstrate with an excised larynx that the false vocal folds are not needed to sustain in vivo-like vibratory patterns. A model that relies on downstream clamping, therefore, has limited application for general descriptions of vocal fold oscillation.

3. Aerodynamic stiffness and fundamental frequency

Consider now the effective stiffness $K^*$. Equation (27) shows a positive aerodynamic stiffness $K$ for the convergent glottis and a negative aerodynamic stiffness for the divergent glottis. For the rectangular glottis, the natural tissue stiffness is unaffected. This differs slightly from the result obtained by Ishizaka and Matsudaira (1972), who found that the effective aerodynamic stiffness for the rectangular glottis was negative. [Note that their coupling stiffness $\phi$ is subtracted from the tissue stiffness of the lower mass, which can be seen by combining the coefficients of $\xi_2$ in Eq. (31).] One oversimplification in their small-amplitude analysis was the choice of zero driving pressure on the upper mass, which might account for the lack of a positive "counterstiffness" at the top. In other words, symmetry about the glottal midpoint ($z = 0$) was distorted by their choice of zero driving pressure in the upper half of glottis.

The positive aerodynamic stiffness in Eq. (27) for a convergent glottis may, in part, be responsible for the observed increase in frequency with transglottal pressure in humans (Ladefoged, 1963) and excised larynges (Baer, 1975). Figure 9 shows fundamental frequency $[(K + K')/M]^{1/2}/2\pi$.

FIG. 8. Oscillation threshold pressure for a rectangular glottis as a function of effective damping pressure. Prephonatory half-width is the parameter.

$m_2\ddot{\xi}_1 + r_1\dot{\xi}_1 + s_c\dot{\xi}_1 = 2\phi(\xi_1 - \xi_2), \quad (31)$

with the mechanical coupling stiffness $s_c$ between the upper and lower masses set to zero. Since in our treatment the mechanical coupling between upper and lower margins of the folds is modeled through the mucosal wave velocity $c$, the Ishizaka and Matsudaira coupling stiffness $s_c$ need not be included separately in Eq. (31).

Ishizaka and Matsudaira labeled $\phi$ their aerodynamic coupling stiffness, which for equal upper and lower masses ($m_2 = m_1$) was

$$\phi = L(T/2)(1/2)pu^2/(2\xi_0)$$

$$= L(T/2)(P_L/2\xi_0)$$

$$= L(T/2)(P_L/\kappa)/(2\xi_0), \quad (32)$$

where $v$ is the average linear velocity in the square glottis. The last step in the above equation is our own derivation and is consistent with the previous assumptions of $k_s = 0$, $P_o = P_L$, and $P_i = 0$ for this particular case [recall Eq. (14)].

By letting $\xi_2 = \xi_1 - 2\tau^2_1$ as in the Taylor series expansion of the surface wave in Sec. A (but with the expansion around the lower mass instead of the glottal midpoint, Eq. (31) yields

$$m_1\ddot{\xi}_1 + (r_1 - 4\tau^2\phi)\dot{\xi}_1 + s_c\dot{\xi}_1 = 0. \quad (33)$$

In order to compare the threshold damping condition $r_1 = 4\tau\phi$ with Eq. (30), we recall that $B$ is the damping per unit surface area $LT$. Since $m_1$ assumes only half of the surface area in the Ishizaka and Matsudaira model, division of Eq. (33) by $LT/2$ gives the threshold damping condition

$$2\tau_1/LT = B = 8\tau\phi/LT. \quad (34)$$

Upon substitution of $\phi$ from Eq. (32) and recalling that $\tau = T/2c$, Eq. (34) can be shown to be identical to Eq. (30). This begins to reconcile our theory with that of Ishizaka and Matsudaira.
as a function of lung pressure, where vocal fold mass per unit area is 0.476 g/cm², tissue stiffness per unit area is 200 kdyn/cm³ (102-Hz natural resonance frequency), and geometric configurations are as previously introduced in Fig. 7(c). For glottal convergence, curves 3 and 4 predict a rise in the fundamental frequency $F_0$ of less than 0.5 Hz/cm H₂O for these values of $K$ and $M$. For glottal divergence, however, much greater changes in $F_0$ are predicted in the negative direction with increasing $P_L$. Fundamental frequency is lowered on the order of 5-10 Hz/cm H₂O when the bottom edges of the vocal folds are nearly touching (curves 1 and 2, where $\xi_{o1} = 0.04$ and 0.08 cm, respectively). For these large downward slopes, negative aerodynamic stiffnesses are evident in Eq. (27). In addition, since $\xi$ is negative and $\xi_{o1}$ is small, the potential for approaching a singularity in the denominator of $K'$ is quite strong.

The validity of such large negative $F_0$ changes with subglottal pressure must be questioned, however. In large-amplitude oscillation, vocal fold collision and nonlinear stiffness may possibly override the small-amplitude negative stiffness derived here. This was evident in the large-amplitude simulations of Ishizaka and Flanagan (1972), where a negative slope between fundamental frequency and subglottal pressure (i.e., a negative aerodynamic stiffness) did not occur as predicted by Ishizaka and Matsudaira (1972) for small amplitudes. More recently, Titze and Durham (1987) have shown that nonaerodynamic stiffness changes can account for the typical variations of $F_0$ with subglottal pressure. This is not to negate the present study, but to serve as a precautionary measure against extrapolations beyond the simplified set of conditions and the simplified model presented here.

Again, as an internal check on the validity of the small-amplitude expression for $K'$, Fig. 10 shows the frequencies of threshold oscillations obtained numerically by solution of Eq. (22). [Portions of waveforms 2 and 3 were previously shown vertically in Fig. 7(b).] In the 100-ms time interval in Fig. 10, waveforms 1-4 show 9.1, 9.9, 10.5, and 11.0 completed cycles, respectively, corresponding to frequencies of 91, 99, 105, and 110 Hz. This agrees with the frequencies indicated by the dots at the threshold pressures in Fig. 9. Damping pressures were again chosen from Fig. 7(a) for the respective configurations 1-4 at the indicated dots.

The $F_0$ dependence on lung pressure and glottal convergence may be of significance in register control in human phonation. It is known, for example, that contraction of the vocalis muscle tends to adduct the lower portions of the vocal fold toward a more rectangular (or even slightly divergent) glottis (Hirano, 1975). It is also known that the vocalis muscle is more active in chest register than in falsetto register (Hirano et al., 1970). If we assume, therefore, that configuration 4 in Fig. 7(c) is a falsettolic prephonatory adjustment and configuration 2 is a chestlike prephonatory adjustment, we can see at once why significant acoustic changes are likely to occur when the register changes. For a nominal 8-cm H₂O lung pressure in speech, the fundamental frequency in Fig. 9 drops from 108 to 70 Hz when a shift is made from configuration 4 to configuration 2. This type of a frequency change (up to an octave or so) is not uncommon in abrupt register transitions in humans or excised larynges.

More dramatic yet is likely to be the effect on intensity, although it is not quantified here. The mean vocal fold equilibrium position shifts toward the glottal midline by 0.05 cm according to Fig. 7(b) (curve 4 to 2 at $P_L = 8$ cm H₂O). Since the mean prephonatory position is 0.10 cm, the glottal width is reduced by 50%. This would create longer and more abrupt closure under large-amplitude oscillation conditions, and, hence, a higher vocal intensity. If the lung pressure were slightly higher, the glottis would, in fact, collapse whenever configuration 2 is approached, even with small amplitudes. As a qualifier to this statement, however, it should be remembered that human tissue does not exhibit linear stiffness characteristics. Excessive stretch hardens the effective spring constant $K$ (Perlman and Durham, 1987), making total collapse less likely. The fundamental frequency is significantly affected by nonlinearity in the stiffness (Titze and
E. Oscillation conditions for lumped inertive vocal tract coupling

Whenever the fundamental frequency of oscillation is below the first-formant frequency of the vocal tract, the reactive portion of the input impedance is inertive (Rothenberg, 1981). To a first-order approximation, then, the supraglottal vocal tract can be represented by a lumped inertance and a resistance \( R_2 \). Subglottal acoustic loading could be similarly represented by \( I_1 \) and \( R_1 \) but, to eliminate complexity, we restrict our discussion to supraglottal loading. In a previous article (Titze, 1985), we showed that supraglottal and subglottal loads are, to a large extent, additive. More work needs to be done, however, to understand the effect of the lungs and trachea on oscillatory conditions.

Consider, then, the vocal tract input pressure to assume the form

\[ P_i = R_2 u + I_1 \dot{u}. \]

To study the effects of vocal tract loading independent of mucosal wave motion on the vocal folds, let us eliminate vertical phase differences in tissue movement by setting \( \tau = 0 \) and \( a_1 = a_2 = a \). This is the converse of eliminating the vocal tract load in Sec. II D to study the effects of phase differences in tissue movement. It also agrees with our tutorial approach in Sec. I. Finally, it provides a direct way of studying the oscillatory behavior of a one-mass model, which is maximally dependent on vocal tract coupling for its oscillatory conditions (Ishizaka and Flanagan, 1972).

If we make the simplifying assumptions that \( k_e = 0 \) \((a_2 < A_2)\) as before, we have the important condition where the driving pressure on the vocal folds depends entirely on the vocal tract input pressure, i.e.,

\[ P_e = P_i, \]

according to Eq. (20).

Combining Eq. (35) with Eq. (13), we obtain the equation of motion

\[ R_2 u + I_1 \dot{u} = M \ddot{x} + B \dot{x} + K x. \]

By recognizing that \( u = a v \), where \( v \) is the particle velocity at glottal exit, and letting

\[ a = 2L (\xi_0 + \xi) \]

according to Eqs. (9) and (10), the equation of motion becomes

\[ M \ddot{x} + (B - 2LI_2v) \dot{x} + (K - 2LR_2v - 2LI_1 \ddot{v}) x = 2LR_2 \ddot{\xi}_0 + 2LI_1 \dot{\xi}_0. \]

This equation has two variables, \( \dot{x} \) and \( \ddot{x} \), suggesting that a second equation is needed for solution. Let us assume that the subglottal pressure drop is negligible, i.e., that \( P_s = P_L \), in which case the flow equation (14) can be written as

\[ P_L = R_2 u - I_1 \dot{u} = k_i (\rho/2) v^2. \]

Making use again of the continuity law \( u = a v \) and Eq. (38) to replace \( u \) on the left side of (40), the flow equation becomes

\[ k_i (\rho/2) v^2 + 2L [R_2 (\xi_0 + \xi) + I_2 \ddot{\xi}] u + 2LI_1 (\xi_0 + \xi) \dot{u} - P_L = 0. \]

This nonlinear differential equation, in conjunction with (39), contains the oscillation characteristics for a one-mass system that is acoustically loaded.

Exact analytic solution of the two coupled equations is difficult. Fortunately, some special insight can produce a reasonable approximate solution. First, we let \( \xi = \ddot{\xi} + \dot{\xi} \) and \( v = \ddot{v} + \dot{v} \) as in Sec. II D to separate static from oscillatory components. Second, we assume that the oscillatory component of the particle velocity varies primarily with the displacement flow and only secondarily with the driven (transglottal) flow. The small variation of \( v \) with \( u \) over a number of glottal diameters has been borne out experimentally on static models (Scherer, 1981). The displacement flow \( 2LT \xi \) is negative at glottal exit; i.e., air is pulled downward by lateral tissue movement. Making the assumption, then, that only this flow contributes to oscillations in \( v \), we can write

\[ \ddot{v} = - b \ddot{x}, \]

where \( b \) is a constant of proportionality. The approximation will be further justified by an exact numerical solution of Eqs. (39) and (41).

By substituting \( v = \ddot{\xi} + b \dot{\xi} \) and \( \xi = \ddot{\xi} + \dot{\xi} \) into Eq. (39), combining all coefficients of \( \ddot{\xi} \) and its derivatives, and by eliminating higher-order nonlinear terms (products of \( \ddot{\xi} \) and its derivatives are negligible for small-amplitude oscillation), we obtain

\[ M* = M + 2LI_2 b (\ddot{\xi}_0 + \dot{\xi}), \]

\[ B* = B - 2LI_1 \ddot{v} + 2LR_2 b (\ddot{\xi}_0 + \dot{\xi}), \]

\[ K* = K - 2LR_2 \ddot{v}. \]

These are the effective mass, damping, and stiffness of the oscillatory components. In addition, the static components satisfy the relationship

\[ (K - 2LR_2 \ddot{\xi}) \ddot{\xi} = 2LR_2 \ddot{\xi}_0 \]

from Eq. (39) and

\[ k_i (\rho/2) \ddot{v}^2 + 2LR_2 (\ddot{\xi}_0 + \dot{\xi}) \ddot{v} - P_L = 0 \]

from Eq. (41).

Before we discuss the effective mass, damping, and stiffness, we must get a solution for \( \ddot{\xi} \) and \( \dot{\xi} \). Generally speaking, the \( \ddot{v} \) term in Eq. (47) dominates over the linear \( \ddot{v} \) term. Without this domination, there is insufficient nonlinearity in the system to produce oscillation. A first approximation to the static particle velocity is, therefore, the lossless vocal tract (\( R_2 = 0 \)) case, for which

\[ \ddot{v} = (2P_L/k_i \rho)^{1/2}. \]

From Eq. (46), we then get \( \ddot{\xi} = 0 \).

Equations (46) and (47) lend themselves nicely to solution by successive approximation for more accurate values of \( \ddot{v} \) and \( \ddot{\xi} \) when \( R_2 \neq 0 \). Starting with \( \ddot{v} \) as the initial estimate, \( \ddot{\xi} \) can be computed from (46) and substituted into (47). Solution of the quadratic equation then yields a better approximation to \( \ddot{v} \), which can be used to get a better approx-
imation to \( \xi \), and so on. By way of example, consider the following tissue and air parameters: \( K = 200 \, \text{kdyn/cm}^3 \), \( M = 0.476 \, \text{g/cm}^2 \), \( B = 123 \, \text{dyn-s/cm}^3 \) (a damping ratio of 0.2), \( \rho = 0.00114 \, \text{g/cm}^3 \), \( k_r = 1.1 \), \( \xi_0 = 0.04 \, \text{cm} \), \( L = 1.4 \, \text{cm} \), \( I_2 = 0.01 \, \text{dyn-s}^2/\text{cm}^4 \), and \( R_2 = 3.1 \, \text{dyn-s/cm}^2 \). The inertance \( I_2 \) is estimated from the cylindrical tube relationship \( \rho l/A \), where \( l \) is the length of the tube (17.5 cm) and \( A \) is the cross-sectional area (2 cm²). The \( R_2 \) is estimated (very roughly) from the vocal tract resonance \( Q \) and bandwidth at the first-formant frequency. For a 500-Hz resonance frequency and a 50-Hz bandwidth, \( Q = 10 = \omega_0 I_2/R_2 \), which leads to the estimate of \( R_2 \) given above. This calculation assumes a series resonant circuit, however, with a constant resistance [Eq. (35)]. It provides, at best, a rough estimate to a more appropriate frequency-dependent \( R_2 \) that depends on heat conduction, wall vibration, lip radiation, and a number of other factors. This frequency dependence does not make a profound difference in the arguments to be presented here, but caution must be taken when more precise values are sought for vocal tract resistance and inertance.

Based on these parameters and an 18-cm H₂O threshold lung pressure (to be discussed later), the initial estimate of \( \bar{v} \) from Eq. (48) is 5304 cm/s. This leads to an initial estimate of 0.012 cm for \( \xi \) [Eq. (46)]. The second approximation of \( \bar{v} \) is 4956 cm/s from the quadratic equation (47). With one more iteration, \( \bar{v} = 4963 \, \text{cm/s} \) and \( \xi = 0.0114 \, \text{cm} \). Usually, four to five iterations are sufficient to obtain accuracy in the fourth decimal place. A family of curves relating \( \bar{v} \) and \( \xi \) to \( P_L \) is shown in Fig. 11(b). The lung pressure is on the ordinate, as in previous threshold curves. The abscissa is labeled \( B/(2L I_2) \) in cm/s and \( \xi \) at the top (dashed lines). Parameters 1–4 correspond to configurations shown in Fig. 11(c). Note that \( \xi \) increases with the prephonatory half-width \( \xi_0 \), whereas \( \bar{v} \) decreases slightly with the same parameter. Also, note that there are no negative values of \( \xi \) because of the specific choice of zero convergence of the prephonatory glottis [Fig. 11(c)]. Positive values of \( \xi \) for all the configurations are due to a static component of supraglottal

![FIG. 11. Oscillation threshold pressure \( P_L \) as a function of (a) damping velocity \( B/(2L I_2) \) and (b) static displacement \( \xi \) (solid lines) and mean particle velocity \( \bar{v} \) (dashed lines). (c) Prephonatory configurations labeled 1–4 in above curves. Dashed lines in (a) are for small-amplitude approximations (see text).](image-url)
pressure ($R_2 \bar{u}$, where $\bar{u}$ is the flow corresponding to $\bar{p}$). This static component tends to drive the vocal folds away from the midline [recall Eq. (37)] for all configurations, making oscillation more difficult. Thus we can anticipate that supraglottal resistance will raise the oscillation threshold pressure and the effective damping, which is the topic of the next discussion.

### 1. Vocal tract related damping and oscillation threshold pressures

We return now to Eqs. (43)-(45), which give the effective mass, damping, and stiffness of the oscillatory component $\xi$. In particular, we are interested in Eq. (44) in this section. Note that the vocal tract acoustic load introduces two damping terms, one negative and one positive. Negative damping, which can lead to oscillation, is provided by vocal tract inerterance $I_2$. On the other hand, vocal tract resistance adds to the tissue damping $B$. Comparing Eq. (44) to Eq. (26), we see that vocal tract inerterance has taken the place of $r$, the time delay in tissue movement between upper and lower portions of the fold. As noted in Sec. I, air inerterance produces time delay in the downstream response, making the correspondence between $I_2$ and $r$ somewhat intuitive. In effect, the air column above the folds acts like the upper mass of a two-mass vocal fold system in retarding the pressure changes above the lower mass.

The threshold condition for oscillation is $B^* = 0$, which leads to the relationship

$$\bar{u} = B / (2LI_2) + \omega_0 b (\xi_0 + \xi),$$

(49)

where $\omega_0 = R_s / I_2$ is the first-formant bandwidth of the vocal tract. (Recall the previous discussion on circuit Q and bandwidth.) For a given set of vocal tract and vocal fold parameters, $\bar{u}$ is the particle velocity needed to produce sustained oscillation. A threshold lung pressure [the ordinate of Fig. 11(b)] is directly related to $\bar{u}$. Note that there is very little spread in the velocity curves, suggesting that $\bar{u}$ is almost independent of prephonatory width (or flow). As seen from Eq. (47), a dependence of $\bar{u}$ on $\xi_0$ exists only for a sizable $R_2$. Otherwise, the simple quadratic relationship $P_L = k, (\rho/2)\bar{u}^2$ applies.

Equation (49) predicts that increased vocal fold length $L$ can lower the phonation threshold velocity, similar to the way in which vocal fold thickness lowered phonation threshold pressure for the case of nonuniform tissue movement [Eq. (28)]. In addition, increased vocal tract inerterance lowers the threshold particle velocity, whereas increased formant bandwidth raises it.

Still an unknown in Eq. (49) is the constant $b$, which relates the oscillatory components of particle velocity to tissue velocity [recall Eq. (42)]. A trial and error method has been used to determine this constant in the following way: A fourth-order Runge-Kutta numerical solution of Eqs. (39) and (41) was used to simulate displacement and particle velocity waveforms, as shown in Fig. 12. Particle velocity $v$ is at the top and displacement $\xi$ is at the bottom. For the vocal tract and tissue parameters given in the example of Sec. II D, the lung pressure $P_L$ was adjusted until no growth or decay in the amplitudes was observed. This defined the oscillation threshold pressure, which was the 18-cm H$_2$O value introduced earlier to calculate a typical value of $\bar{u}$ and $\xi$. Note that the oscillatory component $\bar{u}$ of the particle velocity is small in comparison with the static component $\bar{u}$. The amplitude of $\bar{u}$ is about 330 cm/s, whereas $\bar{u}$ is about 5000 cm/s, as determined graphically by reading the ordinate values on the right side of Fig. 12. This value of $\bar{u}$ agrees with the 4963-cm/s value computed iteratively in the previous section. Also note that $\xi$ has the value of 0.011 cm determined earlier by iterative means (ordinate values on the left).

The assumption that $\xi$, the derivative of the lower trace in Fig. 12, is 180$^\circ$ out of phase with $\bar{u}$ [Eq. (42)] is not quite validated. The observed phase shift in Fig. 12 is more like 150$^\circ$. This does not introduce some error into the predicted threshold pressures, but we will see that the error is small. An estimate of $b$ from Fig. 12 is $\omega_0 \xi_m / v_m = 30$, where $\omega_0$ is the radian frequency of oscillation ($2\pi \times 88$ Hz), $\xi_m$ is the displacement amplitude (0.02 cm), and $v_m$ is the particle velocity amplitude (330 cm/s), all determined graphically from Fig. 12.

A refinement of $b$ was obtained in the following manner. For a variety of values of $B$ and $\xi_0$, the oscillation threshold pressures were found by simulation as explained above. They were plotted as a function of $B / (2LI_2)$ as shown in Fig. 11(a) (solid lines). $B / (2LI_2)$ has the units of velocity (cm/s) and may be thought of as the effective damping velocity at threshold when the vocal tract is lossless [Eq. (49) for $\omega_0 = 0$]. This effective damping velocity corresponds to the effective damping pressure $Bc$ plotted in Fig. 7(a). The solid lines in Fig. 11(a) were then approximated by threshold curves from Eq. (49). The procedure involved choosing values of $P_L$, calculating $\bar{u}$ and $\bar{u}$ from Eqs. (46) and (47), and finding the corresponding value of $B / (2LI_2)$. The result is shown by the dashed lines in Fig. 11(a). A value of $b = 38$ gave the best match over the entire family of curves. Excepting rather low values of $B / (2LI_2)$, which are nontypical, the dashed lines are a good approximation to the oscillation threshold pressure.
threshold curves for this set of parameters. A slightly different value of $b$ may be needed, however, for a different set of $K$, $L$, $\omega_b$, and $I_2$. This needs to be investigated further.

The successful predictions of oscillation threshold conditions by the combination of Eqs. (46), (47), and (49) are merely a check of internal consistency of our mathematical model. It validates the small-amplitude approximations, but says nothing about the validity of the model itself. Such validity will be given in a follow-up study on large-amplitude oscillation that includes experimental data on excised larynges. For the present, the results given here are intended to lay the foundation for a theory that will no doubt see numerous quantitative modifications.

Before entering the final discussion on fundamental frequency of the vocal tract-coupled oscillator, a comparison of Figs. 8 and 11(a) is instructive. Here, we have the identical prephonatory configurations with a rectangular glottis, but different mechanisms of oscillation. For identical tissue damping ($B = 123$ dyn-s/cm$^3$, which corresponds to a damping ratio of 0.2), and for a typical 100-cm/s mucosal wave velocity, the damping pressure $Bc$ is 12 300 dyn/cm$^2$, or 12.6 cm H$_2$O. The oscillation threshold pressure in Fig. 8 is less than 7 cm H$_2$O for all values of $\xi_0$. In contrast, the damping velocity $B/(2LI_2)$ in Fig. 11(a) is 4393 cm/s for the typical 0.01-dyn-s$^2$/cm$^3$ vocal tract inertance. Here, the oscillation threshold pressures are all greater than 18 cm H$_2$O. For the largest prephonatory half-width ($\xi_0 = 0.16$ cm, curve 4), 48 cm H$_2$O are required in comparison with the 7-cm H$_2$O value in Fig. 8. This suggests that vocal tract inertive coupling, although theoretically a viable alternative to the mucosal wave mechanism for obtaining oscillation, seems to have a practical limitation. Of course, it must be recognized that vocal tract inertance, i.e., the acoustic mass of the airways, can vary greatly depending on vocal tract shape. In particular, vocal tract constrictions in the upper larynx (in the ventricular and epiglottal regions) can cause large inertive effects. But the primary problem is the vocal tract resistance. For the idealized $R_2 = 0$ case (zero-formant bandwidth), the combination of Eqs. (47) and (49) gives a threshold pressure

$$P_L = k_r (p/2)(B/(2LI_2))^2 \quad (R_2 = 0), \quad (50)$$

for all values of $\xi_0$. For typical parameters quoted previously, this yields the much lower threshold of 12.3 cm H$_2$O. Unfortunately, vocal tract losses can never be entirely eliminated. We conclude, therefore, that vocal tract inertive coupling may assist the primary mucosal wave mechanism for obtaining oscillation, (assuming that the effects are somewhat additive), but is unlikely to become the dominant mechanism in any normal phonatory adjustment.

2. Vocal tract-related stiffness and fundamental frequency

Equation (45) reveals that acoustic resistance in the vocal tract transforms into negative mechanical stiffness in the vocal folds. Earlier we showed that acoustic inerance in the vocal tract transforms into negative mechanical resistance (damping) in the vocal folds. This is simply a result of the fact that air movement and tissue movement are one derivative removed from each other. No tissue movement causes constant air movement in the glottis, whereas constant tissue movement causes accelerated air movement.

Numerically, the negative stiffness can be appreciable. Using the typical parameters given earlier, the vocal tract-related stiffness per unit area has the value $-43\times 100$ dyn/cm$^2$. This compares with 200 000 dyn/cm$^3$ for the tissue stiffness. The small-amplitude fundamental frequency is

$$F_0 = \left(\frac{K^*}{M^*}\right)^{1/2}/2\pi, \quad (51)$$

where $M^*$ is defined in Eq. (43). Note that the fundamental frequency can only drop with a resistive-inertive load, since $K^*$ is always smaller than $K$ and $M^*$ is always greater than $M$. For typical parameters chosen previously, $F_0$ drops from 103 Hz in the unloaded condition ($R_2 = I_2 = 0$) to 86.6 Hz in the loaded condition. The latter value agrees with the fundamental frequency of the simulated waveform in Fig. 12. This lowering, or "pulling," of the fundamental frequency with increased vocal tract loading was demonstrated by Ishizaka and Flanagan (1972) with various lengths of acoustic tubes attached to their two-mass and one-mass simulation models. The one-mass model was most susceptible to this frequency "pulling" and would oscillate only with an inertive load, as shown in the present theory.

III. CONCLUSIONS AND APPLICATION OF THE THEORY

The body-cover concept of vocal fold tissue morphology has been adopted to capture the essential features of vocal fold vibration. By assuming an upward-propagating surface wave in the cover, relatively simple differential equations were derived for oscillatory motion. Upon linearization of these equations for small-amplitude oscillations, closed-form solutions for effective damping, mass, and stiffness were obtained. Oscillation threshold pressures were computed for various glottal configurations and viscoelastic properties.

Our results are not markedly different from those obtained by Ishizaka and Matsuda (1972). The primary mechanism for obtaining velocity-dependent driving forces seems to be time-delayed action, i.e., a delayed reacting pressure on the bottom of the folds due to a retarded motion at the top. In our formulation, the delay results from a finite surface-wave velocity, whereas in the Ishizaka and Matsuda formulation, the two coupled oscillators are locked into a flow-controlled mode that keeps upper and lower masses out of phase, as cogently illustrated by their state diagram (p. 63). It is interesting to point out that delayed action is a frequently encountered mechanism for oscillation in feedback control systems.

Surface-wave propagation in the mucosa (cover) must be from bottom to top. Reversal of this direction would not produce oscillation. Rather, energy would flow from the tissue to the airstream. This is equivalent to the statement that the bottom mass in the two-mass system must always lead the top mass in phase. Negative effective damping will then be produced, which, in our study, was found to be inversely proportional to the surface-wave propagation velocity, or directly proportional to the time delay between upper and lower portions of the folds.

Ingo R. Titze: Small-amplitude oscillation of the vocal folds 1550
The delayed-action hypothesis was also used to conceptualize the effect of inertive loading by the vocal tract on the folds. Since buildup and collapse of airflow in the glottis and vocal tract are delayed with respect to tissue movement, a greater driving pressure was shown to exist during opening than during closing. This velocity-dependent driving force supplied energy to the vocal folds and helped build up oscillation. Effective damping was shown to decrease with vocal tract inertance and increase with vocal tract resistance. This was also shown by Ishizaka and Matsudaira (1972). Their results demonstrated, furthermore, that a net compliant vocal tract load would tend to squelch oscillation.

Consider an analogy between the voice and some musical instruments. Unlike most reed or brass instruments, whose source characteristics are strongly influenced by the resonator, the vocal instrument operates with varying degrees of source–resonator coupling. When coupling is weak, the vocal folds create their own oscillating conditions by virtue of mucosal surface waves (or some other nonuniform tissue movement). This situation is approached in speech, where a large variety of articulatory gestures offers many different types of acoustic loads. In the presence of these variable loads, the larynx is able to control frequency, intensity, and the glottal source spectrum rather independently. Decoupling preserves constancy of phonation and seems to be desirable. The fact that a linear source-filter theory of speech production has been successful attests to this.

When the fundamental frequency of phonation is raised above the normal speech range, as is often done in singing, two effects may take place that cause source–system independence to diminish. First, the vocal fold cover may stiffen, which would raise the mucosal wave velocity and thereby increase effective damping in the folds. Second, the reactive load of the vocal tract is greater at higher frequencies, particularly when the fundamental frequency approaches a resonant frequency. Both of these factors would make the conditions of oscillation more dependent on the vocal tract, as evidenced by voice “breaks” that occur when an upward glissando is performed into an acoustic tube of unfamiliar length. Some singers, who perhaps have a thick and mobile vocal fold cover, may be able to keep the vocal folds oscillating in a constant manner under changing vocal tract conditions, even in the face of acoustic loads that are compliant rather than inertive. Little or no vowel modification on specific pitches may then be necessary to “tune up” the larynx with the vocal tract. On the other hand, those with less ideal tissue morphology (e.g., a stiffer mucosa) may find it necessary to adjust the vocal tract in specific ways to maintain tissue damping low with inertive loading.

To speculate a bit further, vocal health and hygiene may possibly be linked to the mechanical properties of the vocal folds and, in particular, to the viscoelastic properties of the mucosa. Viscoelastic properties may change with dehydration, with infectious disease, with trauma, with toxic materials, with temperature, or any number of other internal or external conditions. We are presently investigating the effects of hydration in vocal fold tissues on the oscillation threshold pressures defined and described in this article. Exact comparisons between theory and experiment are more suitable for large-amplitude conditions (limit cycles), and will therefore be a topic of a follow-up article.

ACKNOWLEDGMENT

This research was supported by a grant from the National Institutes of Health, Grant number NS 16320-05, for which we are grateful.


