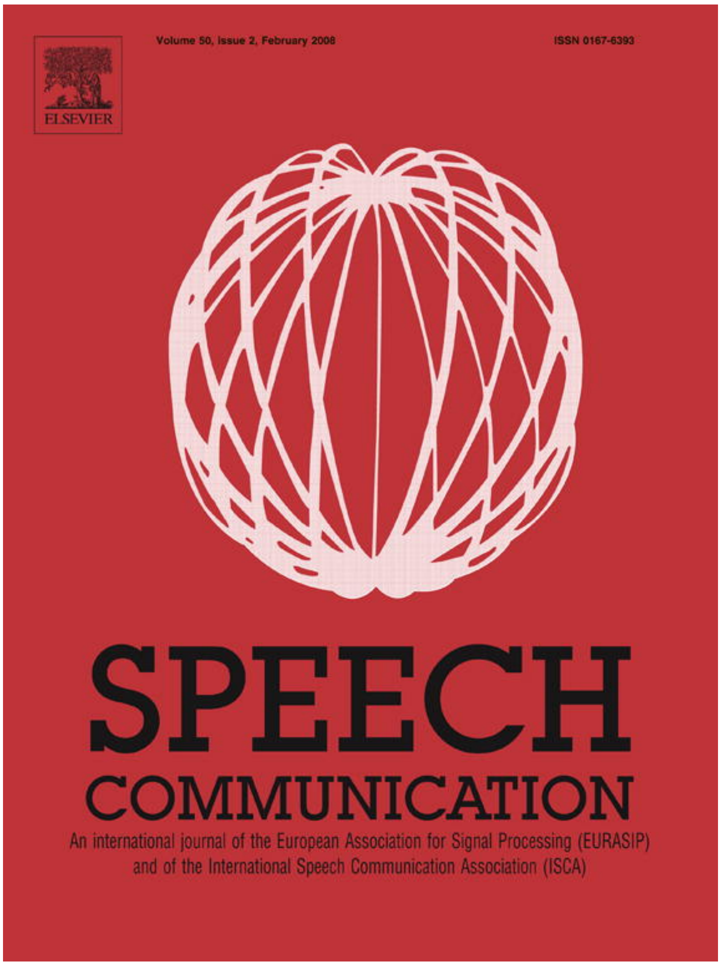
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Speech Communication 50 (2008) 95–108
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Simulation of vocal fold oscillation with a pseudo-one-mass physical model

Federico Avanzini *

Department of Information Engineering, University of Padova, Via Gradenigo 6/A, 35131 Padova, Italy

Received 18 January 2006; received in revised form 20 July 2007; accepted 23 July 2007

Abstract

This paper presents a novel "pseudo-one-mass model" of the vocal folds, which is derived from a previously proposed two-mass model. Two-mass models account for effects of vertical phase differences in fold motion by means of a pair of coupled oscillators that describe the lower and upper fold portions. Instead, the proposed model employs a single mass-spring oscillator to describe only the oscillation of the lower fold portion, while phase difference effects are simulated through an approximate phenomenological description of the upper glottal area. This approximate description is derived in the hypothesis that 1:1 modal entrainment occurs between the two masses in the large-amplitude oscillation regime, and is then exploited to derive the equations of the pseudo-one-mass model. Numerical simulations of a reference two-mass model are analyzed to show that the proposed approximation remains valid when values of the physical parameters are varied in a large region of the control space. The effects on the shape of the glottal flow pulse are also analyzed. Comparison of simulations with the reference two-mass model and the pseudo-one-mass model show that the dynamic behavior of the former is accurately approximated by the latter. The similarity of flow signals synthesized with the two models is assessed in terms of four acoustic parameters: fundamental frequency, maximum amplitude, open quotient, and speed quotient. The results confirm that the pseudo-one-mass model fit with good accuracy the behavior of the reference two-mass model, while requiring significantly lower computational resources and roughly half of the mechanical parameters.
© 2007 Elsevier B.V. All rights reserved.

PACS: 43.70.Bk; 43.72.Ja

Keywords: Voice source; Low-dimensional models; Synthesis

1. Introduction

Research on the simulation of vocal fold vibration is typically focused on two complementary goals: the study of specific aspects of the mechanics of the vibration, and the use of the simulations within a speech synthesis system. Various approaches to the modeling of vocal fold action have been proposed in the literature. In parametric signal models, the glottal flow waveform or its first-time derivative is parametrized in terms of piecewise analytical functions: the LF model proposed by Fant et al. (1985) is a well-known example of this approach, and characterizes one cycle of the flow derivative using four parameters. On the other hand physical models reproduce the oscillatory characteristics of the vocal folds through numerical simulations based on a physical description of the vocal fold system. *Distributed-element* approaches model the folds by means of systems of partial differential equations which are numerically simulated using, e.g., finite-element methods (Berry and Titze, 1996; de Vries et al., 1999; Gunter, 2003). Simpler *lumped-element* models schematize the system by means of ideal masses coupled through springs and dampers (Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972; Liljencrants, 1991; Pelorson et al., 1994; Story and Titze, 1995; Lous et al., 1998).

* Tel.: +39 049 827 7942; fax: +39 049 827 7699.
E-mail address: avanzini@dei.unipd.it

0167-6393/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2007.07.002

Author's personal copy

Oscillation of these mass-spring systems is induced by the difference in hydrodynamic force on the vocal folds during opening and closing phases. Pressure and flow within the glottis are often related through the Bernoulli equation, which implies that crude simplifications of the fluid-dynamics have to be made. Several works have investigated more refined descriptions of the glottal flow, including nonsteady behavior and behavior at collision, which is known to be a very important feature of voiced sound production (Pelorson et al., 1994; Deverge et al., 2003; de Vries et al., 2002; Vilain et al., 2004).

Physical models of the voice source have been integrated into articulatory models of the vocal apparatus (Sondhi and Schroeter, 1987), and can potentially provide realistic excitation signals, by reproducing such "natural" effects as pitch variations at the beginning and at the end of a voiced segment, or occurrence of oscillatory ripples and skewing of the glottal flow waveform due to acoustic interaction with the vocal tract (Titze, 1988; Childers and Wong, 1994). However the physical modeling approach has not been yet extensively exploited in applications such as speech synthesis and coding/compression despite its potential. One reason is to be found in the computational costs of numerical realizations, which are far more demanding than in the case of parametric signal models, and increase with the accuracy of the models. A second more substantial reason lies in the difficulty to derive the values of physical parameters required to synthesize speech, and in particular to fit the models to observed data. While signal models (e.g., the LF model) can be successfully used for fitting flow derivatives computed by inverse filtering real utterances (Childers et al., 1995; Riegelsberger and Krishnamurthy, 1993; Strik, 1998), using physical models for the same purpose is a non-trivial task, which involves inversion of non-linear dynamic systems with a large number of parameters involved. As an example, as many as 19 parameters have to be specified in the first two-mass model proposed by Ishizaka and Flanagan (1972), and later refinements such as the 3-mass model of Story and Titze (1995) involve even larger numbers of parameters.

Few studies have dealt with the problem of mapping speech signals to physical parameters of the voice source. Flanagan et al. (1980) proposed a method for parametric control of a two-mass model by an adaptive procedure, which was further improved by Schroeter and Sondhi (1992). More recently, Sciamarella and D'Alessandro (2004) have investigated the sensitivity of acoustic flow parameters to independent variation of physical parameters of a two-mass model. Avanzini et al. (2006) have used a set of rules, derived from (Titze and Story, 2002), to control a two-mass physical model through activation levels of laryngeal muscles, and explored the mapping between these muscular activations and a set of relevant acoustic flow parameters. Following a different approach, Drioli (2005) has proposed a one-mass "semi-physical" model which embeds a regressor-based non-linear mapping, whose parameters can be estimated from inverse filtered flow signals.

In this paper we present a new one-mass model, which is derived from a reference two-mass model previously proposed by Lous et al. (1998), and employs a single mass-spring system to describe the oscillation of the lower portion of the vocal fold. The main novel feature of the proposed model is that the effects of vertical phase differences in fold motion are simulated through an approximate phenomenological description of the upper glottal area. As such, the model captures the main features of a two-mass model although a single mass is used, and may therefore be defined as a "pseudo-one-mass" model.

The pseudo-one-mass model mimics closely the dynamic behavior of the reference two-mass model, while requiring significantly lower computational resources thanks to the reduction in the model dimensionality. For the same reason, the number of mechanical and geometrical parameters of the one-mass model is roughly half that of the reference two-mass model, which makes it more suitable for voice source parameter matching applications. We also show that the parameters of the one-mass model are uniquely determined given those of the reference two-mass model. Seminal ideas related to this work were presented in (Avanzini et al., 2001).

The paper is organized as follows. Section 2 reviews low-dimensional vocal fold models, and describes the two-mass model that will be used as a reference throughout the paper. Section 3 proposes an approximate description of the dynamic behavior of the two-mass model, in the hypothesis of 1:1 modal entrainment between the two masses in the large-amplitude oscillation regime. Based on this approximate description, the equations of the pseudo-one-mass model are derived. In Section 4, numerical simulations of the reference two-mass model are analyzed to show that the proposed approximation remains valid when values of the physical parameters are varied in a large region of the control space. The simulations also show that the pseudo-one-mass phenomenological parameters are uniquely determined by, and have a smooth dependence on the two-mass physical parameters. Finally, Section 5 discusses issues related to the realization of the proposed model. Comparison of simulations with the reference two-mass model and the pseudo-one-mass model show that the dynamic behavior of the former is accurately approximated by the latter.

2. Low-dimensional vocal fold models

Low-dimensional models "lump" mechanical properties of the vocal folds into coupled masses and springs. As such, these models have a limited number of degrees of freedom (typically two or three) and are accordingly able to capture a limited number of primary modes of tissue vibration. While finite-element simulations based on continuum models of the vocal folds have the advantage of embedding physiological information (e.g., muscle activations) into the

Author's personal copy

mathematical formulations, the appeal of lumped-element models lies in their conceptual simplicity and interpretability in terms of non-linear dynamics (Lucero, 1996). Because of their limited computational costs, lumped models are also often used for synthesis purposes (Sondhi and Schroeter, 1987).

2.1. Normal modes in vocal fold oscillation

Observations of vibrating vocal folds with stroboscopic techniques or high-speed film/video show that the cover tissue supports a surface wave that propagates from the lower portion of the vocal fold cover to the upper one. This is often referred to as the mucosal wave (Story, 2002), or vertical phase difference. Fig. 1 depicts a cycle of vocal fold vibration in the coronal plane. The folds are initially contacting and the glottis is closed (a), then a lateral movement of the lower portion of cover surface starts (b) and continues until the two sides separate and the glottis opens completely (c). After reaching the maximum lateral displacement (d), the lower portions begin to close (e), and eventually collide and again close the airway (f). The displacement continues until the upper portions also collide (g), and the cycle is completed (h). During this process the lateral displacement of the upper portion of each fold is not in phase with the lower one and a wave-like motion from bottom to top is created. This mucosal wave is superimposed on top of the overall back and forth tissue motion.

Numerical simulations from finite-element models of vocal fold vibration (Berry and Titze, 1996) show that the motion patterns sketched in Fig. 1 are well described by two basic vibrational eigenmodes of the vocal fold system in the coronal plane. Fig. 2 depicts the shape of these two modes: the first one is a shear mode with the lower and upper margins of the fold moving π-out of phase so that alternating divergent and convergent glottal shapes are produced during the oscillation; the second one is a compressional mode with the fold tissue moving laterally in phase.

The asymmetry in the vocal fold motion within the oscillation cycle is a key ingredient for the establishment of flow-induced sustained oscillations. Titze (1988) provides an analysis based on a forced mass-spring oscillator:

$$m\ddot{x}(t) + r\dot{x}(t) + kx(t) = f[x(t), \dot{x}(t), t], \tag{1}$$

where x is the oscillator displacement, f is the driving force, and m, r, k are the oscillator mass, damping, and stiffness, respectively. The dependence of f on the oscillator velocity $\dot{x}$ induces an asymmetry of the driving force on alternate quarter-cycles. Energy is imparted to or taken out of the system when f is in the direction of velocity or opposite to the direction of velocity, respectively. In other words, f has a component that is in phase with $\dot{x}$, and energy transfer to the fold is consequently maximized. A second important effect of the mucosal wave phenomenon is the skewing (to the right) of the glottal flow waveform. In the opening

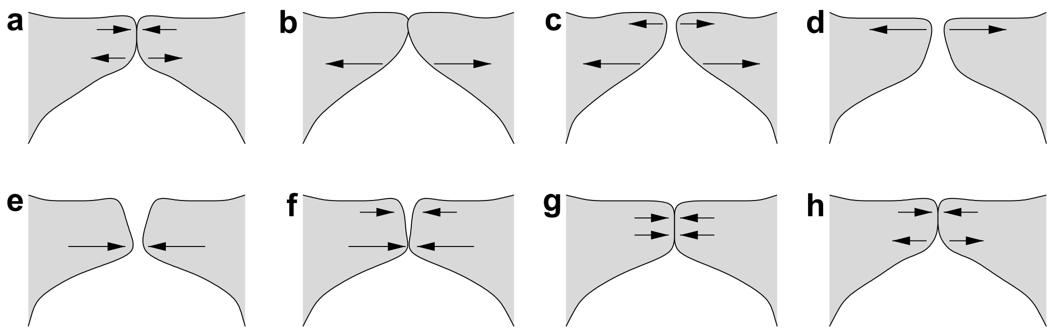


Fig. 1. Idealized cycle of vocal fold vibration in the coronal plane, showing the wave-like motion on the vocal fold surface called mucosal wave. Arrows indicate direction of motion (figure based on Story (2002)).

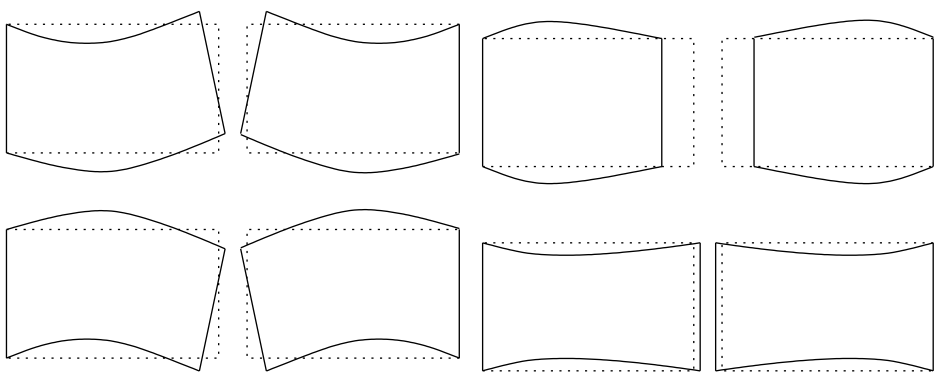


Fig. 2. The two lowest lateral eigenmodes in a finite-element model of the vocal folds (figure based on Berry and Titze (1996)).

Author's personal copy

phase, the inertia of the tissue at the top of the fold causes a slow rise of the flow pulse, which is therefore delayed with respect to the movement of the lower margin of the fold (Titze, 1988).

Phase differences in vocal fold oscillations also occur in the lateral (left–right) and longitudinal (ventral–dorsal) directions, and affect the production of different voice qualities. However the vertical mucosal wave is of particular importance because of the above described effects on the establishment of sustained oscillations. Therefore any vocal fold model that aims at capturing basic mechanisms of oscillation must provide a description of this asymmetry. One-mass models (Flanagan and Landgraf, 1968) simulate the vocal fold with a single mass-spring pair and therefore capture only one compressional mode. Two-mass models (Ishizaka and Flanagan, 1972; Pelorson et al., 1994; Lous et al., 1998) typically use two mass-spring pairs in the coronal plane, plus an additional spring that couples the masses. These models capture both a shear mode (with the two masses out of phase) and a compressional mode (with both masses in phase), which are conceptually equivalent to the two eigenmodes depicted in Fig. 2. Much of the success of two-mass models is arguably due to their ability to capture these two modes and thus to simulate the mucosal wave phenomenon.

2.2. A reference two-mass model

The analysis developed in the next sections is based on a two-mass model derived from (Lous et al., 1998) and depicted in Fig. 3. The model has been implemented in Matlab/Octave, using a numerical realization that ensures accurate simulation of the system (details about this technique have been reported in (Avanzini and Rocchesso, 2002)).

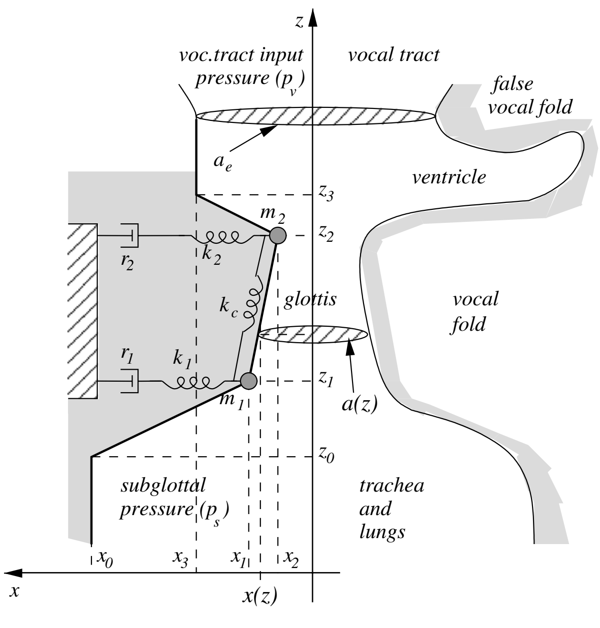


Fig. 3. Right: schematic diagram of the vocal fold, trachea, and supraglottal vocal tract; left: two-mass vocal fold model (view in the coronal plane).

The (coupled) equations of motion for the masses m_1, m_2 are

$$\begin{cases} m_1\ddot{x}_1(t) + r_1\dot{x}_1(t) + k_1[x_1(t) - x_{01}] + k_c[x_1(t) - x_2(t)] \\ \quad = f_1(t) + f_1^{(\mathrm{rest})}(t), \\ m_2\ddot{x}_2(t) + r_2\dot{x}_2(t) + k_2[x_2(t) - x_{02}] - k_c[x_1(t) - x_2(t)] \\ \quad = f_2(t) + f_2^{(\mathrm{rest})}(t), \end{cases} \tag{2}$$

where x_i are the displacements and x_{0i} are the equilibrium positions of $m_i (i = 1,\ 2)$. Collisions between folds are modeled by including restoring contact forces $f_i^{(\mathrm{rest})}$ in the equations. When the mass m_i "collides" (i.e., when the condition $x_i < 0$ holds), k_i and r_i are increased by an amount $k_i^{(\mathrm{rest})}$ and $r_i^{(\mathrm{rest})}$, respectively. The restoring forces $f_i^{(\mathrm{rest})}$ $(i = 1,\ 2)$ are then

$$f_i^{(\mathrm{rest})}(x_i(t), \dot{x}_i(t)) = \begin{cases} -k_i^{(\mathrm{rest})}x_i(t) - r_i^{(\mathrm{rest})}\dot{x}_i(t) & x_i < 0, \\ 0 & x_i \geqslant 0. \end{cases} \tag{3}$$

The dissipative component $r_i^{(\mathrm{rest})}\dot{x}_i$ has been recognized to provide a more realistic behavior during the closed phase (Sondhi and Schroeter, 1987). Expressions for the driving forces f_1, f_2 in Eq. (2) are derived following Lous et al. (1998), who employ a geometrical criterion to predict the (time-varying) point of flow separation along the glottis. More refined theoretical predictions of the flow through the glottis have been proposed. Vilain et al. (2004) have presented a model based on a quasi-steady boundary layer theory, which has been validated experimentally with a setup of oscillating rigid replicas of the vocal folds. Using a similar setup, Deverge et al. (2003) have studied flow behavior at collision. de Vries et al. (2002) have simulated a two-mass model coupled to a glottal flow model based on the incompressible Navier–Stokes equations. For the purpose of this work however we consider the model proposed by Lous et al. (1998) to be sufficiently accurate. Moreover it should be emphasized that the derivation of the pseudo-one-mass model, developed in the next sections, is independent on a specific description of the intraglottal pressure and can be applied even when more refined flow models are used.

One-dimensional, quasi-stationary, frictionless and incompressible flow is assumed from the subglottal region up to a point z_s of the glottis where flow separation and free jet formation occurs. No pressure recovery is assumed at the glottal exit. The pressure $p(z,t)$ at any point z of the glottis is then

$$p(z,t) = \begin{cases} p_s(t) - \frac{1}{2}\rho_{\mathrm{air}} \cdot \mathrm{sgn}[u(t)]u(t)^2\left(\frac{1}{a(z,t)^2} - \frac{1}{a_0^2}\right), \\ \quad \text{if} \quad z \leqslant z_s \\ p(z_s, t), \\ \quad \text{if} \quad z > z_s \end{cases} \tag{4}$$

Author's personal copy

where ρ_{air} is the air density, $u(t)$ is the glottal flow, $p_s(t)$ is the subglottal pressure, $a(z,t)$ is the glottal area at a point z, and a_0 is the area at the glottal inlet (see Fig. 3). If l_g is the vocal fold length, then $a(z,t) = 2l_g x(z,t)$ and $a_0 = 2l_g x_0$. Negative pressure drops are allowed, due to the sgn[$u(t)$] factor in the equations.

The vocal fold profile is described by three rigid plates, each of which joins the points (z_{i-1}, x_{i-1}) and (z_i, x_i) (with $i = 1, 2, 3$). Under this assumption the flow channel height is a piecewise linear function of z:

$$x(z) = \frac{x_i - x_{i-1}}{z_i - z_{i-1}} z + \frac{z_i x_{i-1} - z_{i-1} x_i}{z_i - z_{i-1}} \quad (z_{i-1} \leqslant z < z_i), \tag{5}$$

with $i = 1, 2, 3$. The time-varying separation point z_s is predicted by Lous et al. (1998) to occur at the point where the glottal diameter exceeds the minimum glottal diameter by a fixed amount (10% or 20%). By introducing a *separation constant* s (in the range 1.1–1.2), and using Eq. (5), one finds

$$x_s = \min(sx_1, x_2),$$
$$z_s = \min\left(\frac{z_2(x_s - x_1) + z_1(x_2 - x_s)}{x_1 - x_2},\ z_2\right). \tag{6}$$

The aerodynamic forces on the plates are determined by the pressure distribution $p(z,t)$. By considering force and torque balance on each plate, Lous et al. (1998) find the force $f_i(t)(=f_i[x_i(t), x_{i+1}(t), u(t)])$ to be

$$f_i(t) = \int_{z_{i-1}}^{z_i} l_g \left(\frac{z - z_i}{z_i - z_{i-1}}\right) p(z,t)\mathrm{d}z + \int_{z_i}^{z_{i+1}} l_g \left(\frac{z_{i+1} - z}{z_{i+1} - z_i}\right) p(z,t)\mathrm{d}z. \tag{7}$$

Integration of Eq. (7) for the pressure distribution of Eq. (4) provides the explicit dependence of f_i on x_i, x_{i+1}, u (Lous et al., 1998). Except for the outlet height x_3 (see Section 2.3), values for the geometrical parameters are taken from (Lous et al., 1998): $(z_1 - z_0) = 0.2$ mm, $(z_2 - z_1) = 2$ mm, $(z_3 - z_2) = 0.2$ mm, $x_0 = 9$ mm, $l_g = 14$ mm. The separation constant is $s = 1.2$.

2.3. Vocal tract load

Vertical phase difference in vocal fold motion is only one possible source of driving-force asymmetry. A second one is air inertia in the vocal tract, whose effects sum up with those of the mucosal wave, and contribute to the skewing of the glottal flow waveform (Titze, 1988, 1994).

As a first-order approximation, the vocal tract can be modeled as an inertive load. The model is derived in the limit of fundamental frequencies much lower than the first formant frequency: in this limit the air column acts approximately as a mass of air that is accelerated as a unit, and the vocal tract input pressure $p_v(t)$ can be written as

$$p_v(t) = Ru(t) + I\dot{u}(t), \tag{8}$$

where R and I are the vocal tract input resistance and inertance, respectively. Similar low-loss, low-frequency approximations of the tract load have been adopted by many authors for analysis and simulation purposes, see, e.g. (Fant, 1982; Titze, 1988; Titze, 2006). Being a first-order system, the model (8) does not account for the resonant properties of the vocal tract, however it describes with sufficient accuracy its most relevant effects on vocal fold oscillation. In particular, it has been shown (Titze, 1988; Titze and Story, 1997) that such a model predicts accurately the lowering of the oscillation threshold pressure with respect to the case where no load is present.

Values for the parameters R, I in Eq. (8) are chosen from the analysis of Titze and Story (1997), according to which in the limit of low-losses, low-frequency, and narrow epilarynx the vocal tract impedance is approximated by the epilarynx impedance. If $a_e = 0.5\ \mathrm{cm}^2$ and $l_e = 3.174$ cm are the epilarynx cross-section and length, then the values $R \sim 2.53 \cdot 10^5\ \mathrm{Ns/m}^5$ and $I \sim 724\ \mathrm{Ns}^2/\mathrm{m}^5$ can be considered to be realistic. Moreover an estimate for the outlet height is found as $x_3 = a_e/2l_g \sim 1.8$ mm. Note that this choice is in contrast with many authors, who typically use larger values ($\sim$3 cm^2) for the outlet area: such large values correspond to the initial pharynx cross-section and, as remarked by Titze and Story (1997), are physiologically unrealistic.

We neglect supraglottal physiological details (e.g., ventricles and false vocal folds), and assume that the two-mass model is ideally connected to the epilarynx. The coupling with the vocal tract is then realized by imposing continuity of pressure at the point z_3. This assumption gives

$$p_v(t) = p(z_3, t), \tag{9}$$

where the pressure value on left-hand side is computed from Eq. (8) while that on the right-hand side is computed from Eq. (4).

Following Story (2002), simulations from the two-mass model will be analyzed in the next sections using two conditions of vocal tract loading. In a first "set-up" we assume that no vocal tract load is present, so that the vocal tract input pressure is atmospheric ($p_v \equiv 0$): this roughly corresponds to the configuration of excised larynges typically used for experimental measures. In the second set-up we couple the two-mass model with the inertive vocal load of Eq. (8).

3. A pseudo-one-mass model of the vocal folds

In this section we present a novel "pseudo-one-mass" vocal fold model, which is derived from the reference two-mass model described above. The model exploits an approximation for the motion of the mass m_2 at the upper fold margin. Such an approximation is discussed in Section 3.1, while Section 3.2 introduces the equations of the pseudo-one-mass model. We argue that this model retains the behavior of the reference two-mass model, if the

Author's personal copy

hypothesis discussed in Section 3.1 are met. This conjecture will be validated in Section 4.

3.1. Approximate description of the mucosal wave

In generic large-amplitude oscillation conditions, the mass m_2 follows m_1 with a delay which is in general not constant over the oscillation cycle, and the amplitudes of the displacements x_1 and x_2 are in general different since m_1 and m_2 are subject to different restoring forces. However, due to the non-linear nature of the system (the two mechanical oscillators are non-linearly coupled through the driving forces f_1, f_2), modal entrainment occurs in the large-amplitude oscillation regime (Fletcher, 1978). "Normal" phonation is produced in the two-mass model when the two eigenfrequencies of the system are not too widely spaced and 1:1 entrainment occurs. In this case the system oscillates at a fundamental frequency located between the two eigenfrequencies, and in particular the two masses oscillate with the same frequency. This description of the effects of modal entrainment is analogous to the one given by Berry et al. (2001) with regard to the two lowest modes of a finite-element distributed model of the vocal folds (see Section 2.1).

Based on these remarks, we propose the following approximation in the large-oscillation regime:

$$a_2(t) \sim \alpha a_1(t-\tau), \tag{10}$$

where glottal area signals are hereafter defined as $a_i(t) = 2l_g x_i(t)$ for $i = 1, 2$. In this equation two new "phenomenological" parameters, τ and α, represent a delay and a scaling factor of the displacement x_2 with respect to x_1. This approximation has some resemblance with the one derived by Titze (1988) in the small-oscillation limit: assuming that the (upward) propagation of the mucosal wave obeys the 1-D wave equation $\partial^2 x/\partial t^2 = c^2 \partial^2 x/\partial z^2$ along the z direction (c is the wave velocity on the fold surface), the fold displacement $x(z,t)$ at a point z and time t is $x(z,t) = x_1(t - z/c)$. Using this solution, the rectangular glottal areas $a_1(t)$, $a_2(t)$ at the lower and upper margin of the vocal folds can be expressed as

$$a_1(t) = 2l_g x(z,t)|_{z=z_1} = 2l_g x_1(t), \tag{11}$$

and

$$a_2(t) = 2l_g x(z,t)|_{z=z_2} = a_1(t-\tau), \quad \text{with} \quad \tau = (z_2 - z_1)/c, \tag{12}$$

so that a_2 follows a_1 with a delay τ, which in turn depends on c and on the fold thickness $z_2 - z_1$. Note however that (12) holds in the small-oscillation limit, and cannot be assumed under generic conditions. Therefore (10) and (12), although similar, are derived from very different assumptions.

Given two signals a_1, a_2, the values for α and τ that best fit Eq. (10) can be determined through frequency-domain analysis as follows. Define the Fourier transforms of a_i as

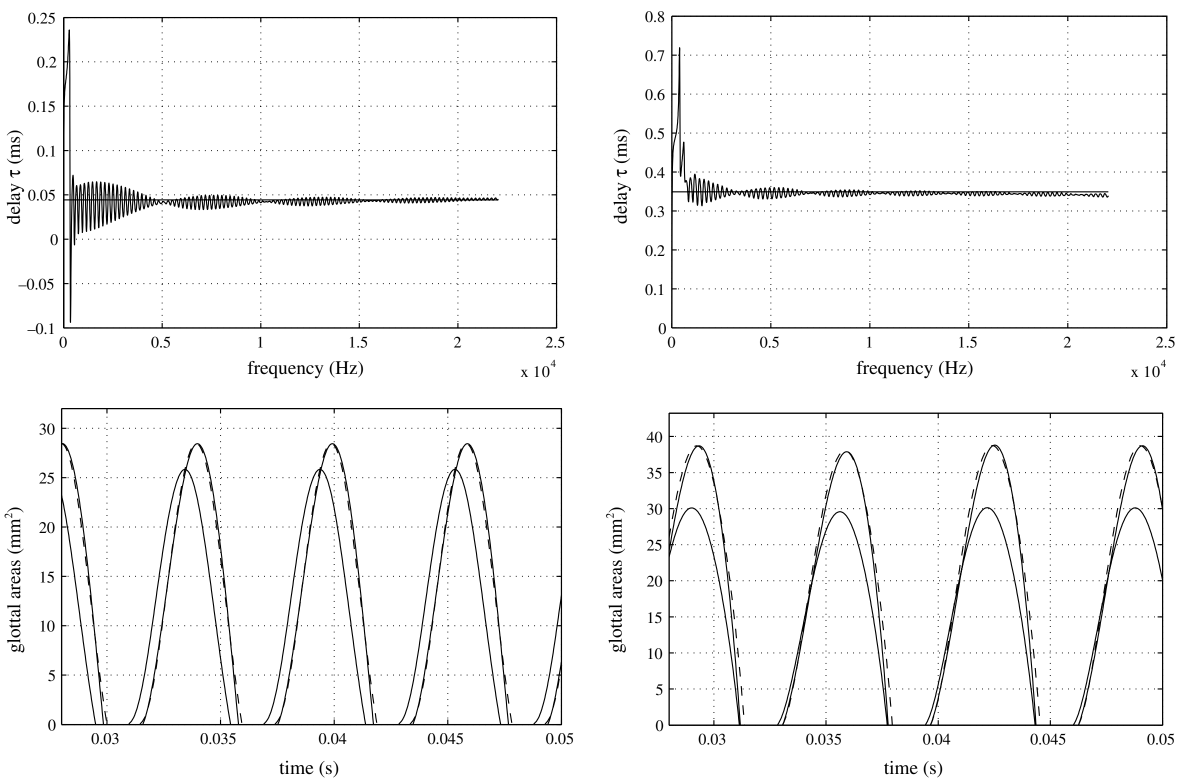


Fig. 4. Examples of fitting of Eq. (10) from two-mass model simulations without (left) and with (right) vocal tract loading. Top: determination of the fit for τ; bottom: glottal area signals $a_1(t)$ and $a_2(t)$ (solid lines) and approximated area signal $\alpha a_1(t-\tau)$ (dashed line).

Author's personal copy

$A_i(\omega) = M_i(\omega)e^{j\phi_i(\omega)}$, then Eq. (10) implies that the magnitude response M_2 and the phase response ϕ_2 are approximated as

$$M_2(\omega) \sim \alpha M_1(\omega), \quad \phi_2(\omega) \sim \phi_1(\omega) + \tau\omega, \tag{13}$$

respectively. The best fit for α is then estimated as $\alpha \sim M_2(2\pi F_0)/M_1(2\pi F_0)$, where F_0 is the fundamental frequency of oscillation. The estimate for τ is found as the constant value that best fit the function $\tau(\omega) = [\phi_1(\omega) - \phi_2(\omega)]/\omega$ in a least-squares sense. Fig. 4 provides two examples of this analysis, performed on two different simulations (without and with vocal tract loading) on the reference two-mass model. The plots at the top show that in both cases the delay function $\tau(\omega)$ is almost constant over the entire range of frequencies, thus supporting the assumption that a_2 is with good approximation a delayed version of a_1. The plots at the bottom show the analyzed glottal area signals, together with the approximation $\alpha \cdot a_1(t-\tau)$ obtained from the analysis. In both cases the fit to a_2 is very accurate. In Section 4.1 we will verify that this is true in a large region of the parameter space of the two-mass model.

3.2. Equations of the pseudo-one-mass model

Given the displacement signal $x_1(t)$ and the approximation (10), the glottal area a_s at the separation point is a function of α, τ through Eq. (6):

$$a_s(\alpha,\tau,t) = 2l_g x_s(\alpha,\tau,t) = 2l_g \min[sx_1(t),\ \alpha x_1(t-\tau)]. \tag{14}$$

Then a closed-form solution for the flow $u_{\alpha,\tau}(t)$ can be derived from Eq. (4):

$$u_{\alpha,\tau}(t) = \mathrm{sgn}[p_s(t) - p_v(t)] \cdot a_0 \cdot a_s(\alpha,\tau,t)\sqrt{\frac{2}{\rho_{\mathrm{air}}(a_0^2 - a_s(\alpha,\tau,t)^2)}}, \tag{15}$$

where $p_s(t)$ is the subglottal pressure, and $p_v(t) = p(z_3, t)$ is the vocal tract input pressure, as before. Therefore the flow is now a function of the displacement x_1 only, and is parametrized by α and τ.

We can then estimate the total force $f_{\alpha,\tau}$ acting on the mass m_1 as

$$\begin{aligned} f_{\alpha,\tau}(t) = {} & f_1[x_1(t), \alpha x_1(t-\tau), u_{\alpha,\tau}(t)] - k_c[x_1(t) - \alpha x_1(t-\tau)] \\ & + f^{(\mathrm{rest})}(x_1(t), \dot{x}_1(t)), \end{aligned} \tag{16}$$

where the three components on the right-hand side are derived from the aerodynamic force of Eq. (7), the coupling elastic force of Eq. (2), and the collision restoring force of Eq. (3), respectively. Again $f_{\alpha,\tau}$ is now a function of the displacement x_1 only, and is parametrized by α and τ.

Finally, by combining (16) with the first equation in (2) the equation of motion for the lower mass m_1 can be written as

$$m_1\ddot{x}_1(t) + r_1\dot{x}_1(t) + k_1[x_1(t) - x_{01}] = f_{\alpha,\tau}(t). \tag{17}$$

This equation describes the pseudo-one-mass model, a single second-order forced mechanical oscillator in which the effects due to phase differences in the motion of the fold margins are controlled through the parameters α, τ and incorporated into the driving force $f_{\alpha,\tau}$. In other words the constraints in vocal fold motion caused by modal entrainment and expressed by Eq. (10) have been exploited to remove one degree of freedom from the two-mass model.

The forced oscillator (17) can be compared to the general formulation given by Titze (1988) and reported in Section 2.1. As already discussed, the dependence of f on the fold velocity $\dot{x}$ in Eq. (1) introduces an asymmetry which is crucial for self-sustained oscillation. The pseudo-one-mass model (17) introduces a similar asymmetry in $f_{\alpha,\tau}$ during the open phase of the oscillation, which is caused in this case by the dependence of $f_{\alpha,\tau}$ on past values of the fold displacement x_1. Therefore the models (17) and (1) are functionally very similar.

Given the analysis developed in the previous sections, the model is expected to exhibit a dynamic behavior closely resemblant of that of the two-mass model. Moreover, given a set of parameter values for the two-mass model, the corresponding parameters for the pseudo-one-mass model are uniquely determined: the mechanical parameters in Eq. (17) and the coupling stiffness k_c in Eq. (16) take the same values of the corresponding two-mass parameters, while α, τ are found using the fitting procedure described in Section 3.1.

As a final remark, note that the above derivation of a pseudo-one-mass model can be applied in principle to any reference two-mass model, and not only to the one used in this work. One will always obtain a general equation of the form (17), in which the exact expression for the driving force $f_{\alpha,\tau}$ will vary depending on the chosen reference model.[1]

4. Validation of the glottal area approximation

In this section we assess the validity of Eq. (10), which is the main hypothesis in the derivation of the pseudo-one-mass model. The assessment utilizes results from numerical simulations of the two-mass model described in Section 2. In Section 4.1 we show that Eq. (10) holds in a large region of the parameter space. Section 4.2 studies the effects of α, τ on the shape of the glottal flow pulse.

4.1. From physical parameters to phenomenological parameters

Numerical simulations of the two-mass model were run for various values of the mechanical parameters and of the subglottal pressure, since all of these parameters are

[1] In fact the original two-mass model of Ishizaka and Flanagan (1972) was used as a reference in an earlier version of this manuscript, and an equation identical to (17) was derived in that case.

Author's personal copy

Table 1
Parameter values of the two-mass models used in (Lous et al., 1998), in (de Vries et al., 1999), and in this study

	m_1 (kg)	m_2 (kg)	k_1 (N/m)	k_2 (N/m)	k_c (N/m)	p_s (Pa)
Lous et al. (1998)	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	40	40	24	800
de Vries et al. (1999)	$2.4 \cdot 10^{-5}$	$2.0 \cdot 10^{-5}$	22	14	10	590
This study	$[1.9, 2.9] \cdot 10^{-5}$	$[1.6, 2.6] \cdot 10^{-5}$	[16, 32]	[8, 22]	[6, 16]	[400, 1400]

expected to influence values for α, τ. Table 1 lists the parameters varied in the simulations, and ranges for their values. The first row in the table reports the values used in the work of Lous et al. (1998). They assumed in particular a total mass of 0.2 g, and stiffness values were consequently chosen in order to obtain realistic values for the fundamental frequency. However such a large vocal fold mass is not confirmed by measurement reported in the literature: assuming a fold tissue density $\rho = 1030\ \text{kg/m}^3$ (Berry and Titze, 1996) and a fold cover depth $d \sim 2$ mm (an intermediate value, with reported measurements ranging from 0.75 to 3 mm (Story and Titze, 1995; Titze and Story, 2002)), an estimate of the total mass in vibration is then $m = \rho l_g (z_2 - z_1) d \sim 0.057$ g.

The order of this estimate is supported by the findings of de Vries et al. (1999), who determined parameter values of a two-mass model by requiring its behavior to match as closely as possible that of a finite-element vocal fold model: see the second row of Table 1. Although the two-mass model used here is different from the one in (de Vries et al., 1999), their findings provide at least a reference for realistic parameter values. Based on these considerations, we explored parameter ranges roughly centered around these values: the last row of Table 1 lists the chosen ranges.

For each simulation run with parameter values in the chosen ranges, synthesized glottal area signals were analyzed in order to extract best fits for α, τ using the procedure described in Section 3.1. The first row of Fig. 5 shows results in the case of no vocal tract loading. For visualization purposes, in each plot all but one of the parameters have been kept fixed around the mean values of their ranges, while the remaining one spans its entire range of variation. A missing point indicates that no phonation occurs for that point. The delay τ takes relatively high-values: this is qualitatively in accordance with the data cited by Titze (1988), who reports measures on excised

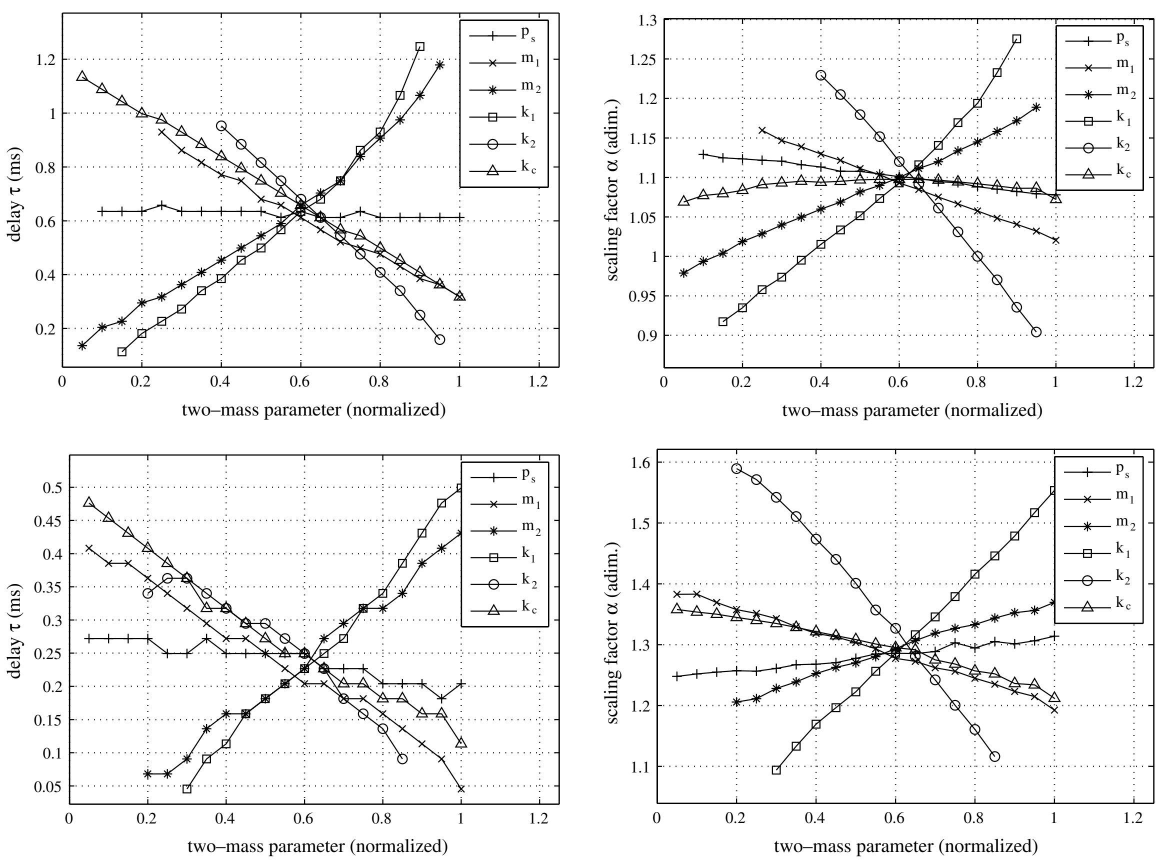


Fig. 5. Simulations on the two-mass model without (top row) and with (bottom row) vocal tract loading, showing dependence of the delay τ (left) and the scaling factor α (right) on the two-mass parameters. The ranges for each parameter are given in Table 1.

Author's personal copy

larynges that give phase differences in certain cases exceeding $\pi/3$ (which corresponds to delays up to about 1.4 ms for a fundamental frequency of 120 Hz). The second row of Fig. 5 shows results in the case of vocal tract loading. The markedly lower values of τ are consistent with the presence of a vocal tract load: air inertia in the tract causes higher positive values for the intraglottal pressure to build up, so that the delay in the motion of m_2 is reduced.

From these results we conclude that the approximation (10) remains valid over a wide region of the parameter space of the two-mass model, in which 1:1 modal entrainment is produced. Moreover, α, τ have a smooth dependence on the two-mass parameters, so that a codebook $(m_1, m_2, k_1, k_2, k_c, p_s) \mapsto (\tau, \alpha)$ can be constructed that samples the 6-to-2 mapping from the low-level parameters of the two-mass model to the pair (τ, α).

4.2. Analysis of glottal flow waveforms

We now analyze how the shape of the glottal flow waveform $u_{\alpha,\tau}(t)$ given in Eq. (15) is affected by the two phenomenological parameters α, τ. The analysis has some similarities with that of Titze (2006), who studied the dependence of the glottal flow waveform on a set of glottal area parameters, including a "phase quotient" and an "amplitude quotient" which roughly correspond to τ and α, respectively. As a displacement signal $x_1(t)$, we use a rectified sinusoid which oscillates around x_{01} with frequency F_0 and maximum amplitude $x_{\max 1}$. The glottal area signal a_1 is then perfectly symmetrical and is given as

$$a_1(t) = \max\{0, 2l_g[x_{01} + (x_{\max 1} - x_{01})\sin(2\pi F_0 t)]\}. \tag{18}$$

Fig. 6 shows the signals $u_{\alpha,\tau}(t)$ and $\mathrm{d}u_{\alpha,\tau}(t)/\mathrm{d}t$ computed from Eq. (15), in the case of no vocal tract loading, and with constant subglottal pressure $p_s = 800$ Pa. Ranges for τ and α have been chosen based on the analysis of Section 4.1, and are roughly coincident with the extremal values in Fig. 5 (first row). These plots show that τ has a direct influence on the shape of the flow waveform. Since τ regulates the opening instant of the glottal channel, it is directly proportional to the duration of the open phase. The skewness of the flow pulse is also affected by τ. The scaling factor α has a major influence on the flow amplitude, and some effect on the skewness, which decreases with increasing values of α. Note that the flow pulse becomes more and more triangular as the delay τ increases. This is a direct consequence of the flow model that is being used. In the realistic limit $a_0 \gg a_s$ Eq. (15) reduces to

$$u_{\alpha,\tau}(t) \overset{a_0 \gg a_s}{\sim} \mathrm{sgn}[p_s(t) - p_v(t)]\sqrt{\frac{2}{\rho_{\mathrm{air}}}} \cdot a_s(\alpha, \tau, t). \tag{19}$$

Therefore the flow u is quasi-proportional to the area at the separation point. For large values of τ there is a portion of the oscillation period in which $a_1(t)$ is closing while $a_1(t-\tau)$ is still opening. This results in a sharp edge of

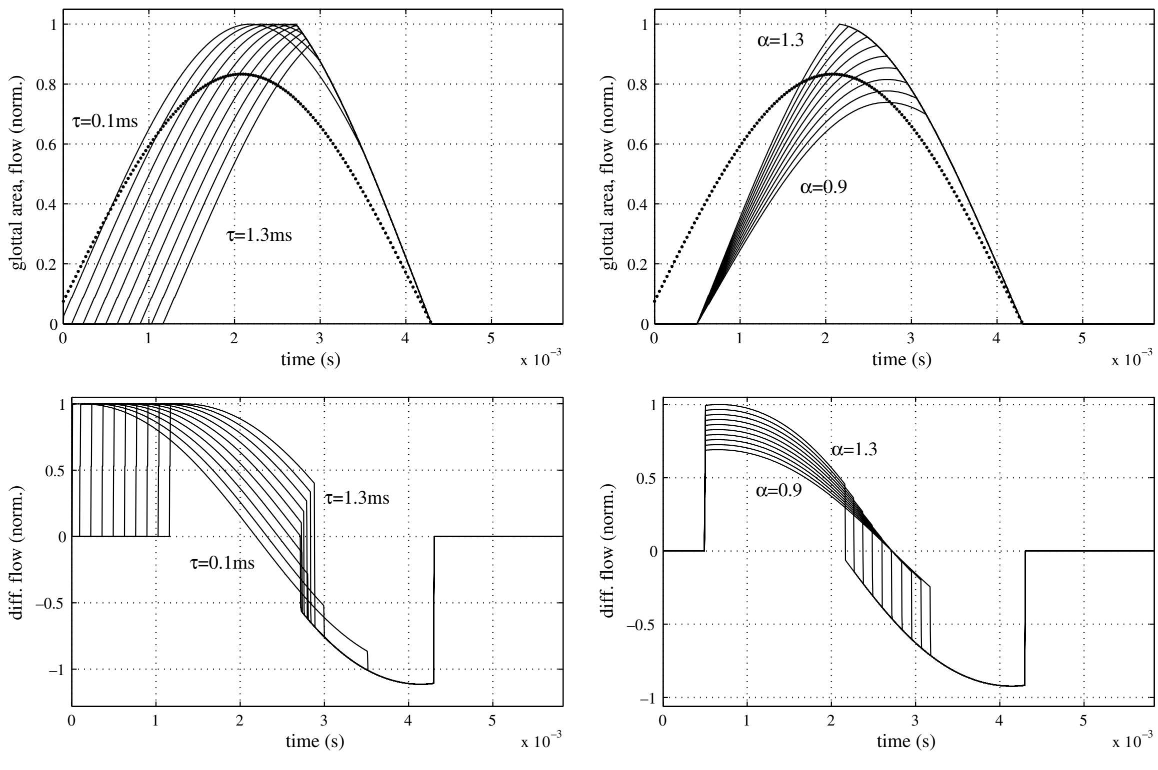


Fig. 6. Plots of $u_{\alpha,\tau}(t)$ and $\mathrm{d}u_{\alpha,\tau}(t)/\mathrm{d}t$ computed from Eq. (15) without vocal tract loading (curves are normalized with respect to maximum ranges). Left column: $\alpha = 1.1$, curves parametrized by τ. Right column: $\tau = 0.7$ ms, curves parametrized by α. Dotted lines represent the glottal area signal a_1 as given in Eq. (18).

Author's personal copy

$a_s(\alpha, \tau, t) = \min[s \cdot a_1(t), \alpha a_1(t - \tau)]$, and consequently of the flow pulse.

Fig. 7 shows the signals $u_{\alpha,\tau}(t)$ and $\mathrm{d}u_{\alpha,\tau}(t)/\mathrm{d}t$ in the more physiologically realistic case of vocal tract loading. A constant value $p_s = 800$ Pa is used as before, and the area function a_1 is defined as in Eq. (18). Again ranges for τ and α have been chosen based on the analysis of Section 4.1, and are roughly coincident with the extremal values in Fig. 5 (second row).

The overall effects of τ, α are qualitatively similar to the no-vocal-load case, with τ influencing both the open phase and the skewness of the flow pulse. In this case, and with these ranges of variation, α has almost no influence on the shape of the flow waveform, and only affects the amplitude. Note that the overall skewness of the flow pulse is more pronounced, consistently with the action of air inertia in the vocal tract. Note also that, unlike in Fig. 6, the flow pulse does not exhibit sharp edges since the delay τ is confined to a smaller range and consequently the separation point follows a smoother trajectory.

From the analysis presented in this section we conclude that τ principally affects the duration of the open phase and the skewness of the flow pulse, while α mostly affects the flow amplitude. It should be emphasized however that the two phenomenological parameters α, τ do not vary independently, as they both depend on the physical parameters of the two-mass model. Therefore the intercorrelation between these parameters and the shape of the glottal flow waveform is extremely complex. We will return on this point in Section 5.2.

5. Realization of the pseudo-one-mass model

The pseudo-one-mass model described in Section 3 has been implemented in Matlab/Octave with the same discretization techniques used for the two-mass model (Avanzini and Rocchesso, 2002). Section 5.1 provides details about the realization, while Section 5.2 discusses the ability of the model to fit the flow signals generated by the reference two-mass-model.

5.1. Properties of the model

Low-computational costs are an especially desirable property in synthesis applications, where real-time simulations are an issue. One advantage of the pseudo-one-mass model is that it requires about half of the computations with respect to the reference two-mass model, since it uses a single mechanical oscillator instead of two. This increase in efficiency comes at the expense of a small overhead in memory requirements, needed in order to store past values of the area signal a_1: a short delay line is used where a_1 is written/read at each sample computation. As a "worst case" analysis for the delay line length, consider the maximum possible value for τ to be $\tau_{max} = 2$ ms and use a sampling rate $F_s = 44.1$ kHz: then the length of the delay line is

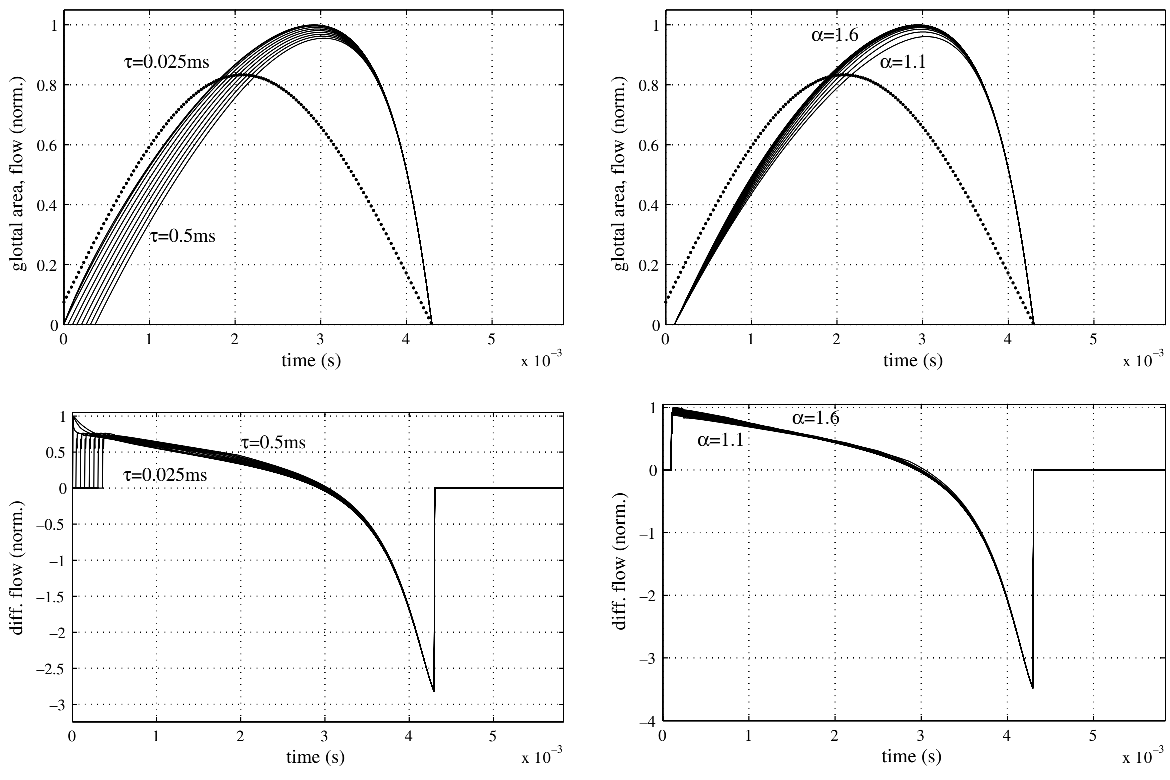


Fig. 7. Plots of $u_{\alpha,\tau}(t)$ and $\mathrm{d}u_{\alpha,\tau}(t)/\mathrm{d}t$ computed from Eq. (15) with vocal tract loading (curves are normalized with respect to maximum ranges). Left column: $\alpha = 1.35$, curves parametrized by τ. Right column: $\tau = 0.24$ ms, curves parametrized by α. Dotted lines represent the glottal area signal a_1 as given in Eq. (18).

Author's personal copy

$N = \lceil \tau_{max} F_s \rceil = 89$ samples (where the ceiling operator $\lceil \cdot \rceil$ returns the smallest integer not less than the argument). Memory occupation is therefore negligible. At present the model implementation uses an integer-delay realization, but a fractional-delay realization may be employed with little additional computational overhead.

Given a set of parameter values for the reference two-mass model the corresponding parameters of the pseudo-one-mass model as chosen as follows:

- the mechanical parameters m_1, r_1, k_1, k_c, the subglottal pressure p_s, and the geometrical parameters have the same values as in the two-mass model;
- the phenomenological parameters α, τ are derived from the physical parameters as in Section 4.1.

Fig. 8 provides an example of simulations from the two models. These confirm that the dynamic behaviors are very similar. There are slight differences at the attack, with the two-mass model having a longer transient. After a few cycles however the signals are very close. The one-mass model signals in Fig. 8 have been shifted by a small amount of time (20 samples at a 44.1 kHz sampling rate), in order to better visualize the matching of the steady-state signals. In Section 5.2 we provide more general results about the ability of the pseudo-one-mass model to fit the behavior of the reference two-mass model.

The example reported in Fig. 8 uses constant parameter values. Realizing the more general case of time-varying control parameters, which is particularly relevant for speech synthesis applications, requires a codebook to be constructed off-line and stored in a look-up table, so that the mapping $(m_1, m_2, k_1, k_2, k_c, p_s) \mapsto (\tau, \alpha)$ is sampled in a large region of the control space. Moreover, an interpolation function must be defined on the codebook, so that α, τ can be estimated even for values of the physical parameters that are not in the codebook. The results summarized in Fig. 5 have shown that the dependence of α, τ on the physical parameters is smooth and monotonic, which are desirable properties for interpolation. From these results we speculate that a coarse quantization of the parameter space and linear interpolation may already provide satisfactory results, although this issue has not been investigated at present. Alternatively more sophisticated non-linear interpolation techniques may be used, like Radial Basis Function Networks. Schroeter and Sondhi (1994) have applied similar techniques to the problem of interpolating acoustic-to-articulatory mappings in the estimation of vocal tract shapes.

5.2. Analysis of flow signals from numerical simulations

This section provides quantitative results about the ability of the pseudo-one-mass model to fit flow signals of the reference two-mass model. A set of acoustic parameters that characterize the flow pulse is chosen, and flow signals from the two models are compared in terms of the values of these acoustic parameters.

Typical voice source quantification parameters extracted from the flow and the differentiated flow waveforms include the *open quotient* OQ, defined as the ratio between the open phase and the total duration of the pulse oscillation, and the *speed quotient* SQ, defined as the ratio between the opening and closing phases of the flow pulse. These parameters are recognized to be particularly relevant for characterizing different voice qualities, such as *modal* (neutral), *breathy*, *pressed*, and so on (Childers et al., 1995; Alku and Vilkman, 1996).

We then compare flow signals from the two models by comparing values of OQ and SQ, together with the fundamental frequency F_0 and the maximum flow amplitude u_{max}. Note that this latter parameter is not particularly relevant for the characterization of voice quality: a more salient parameters is the amplitude of the negative peak of the flow derivative, whose main perceptual correlate is intensity. However the use of u_{max} is preferable in this context,

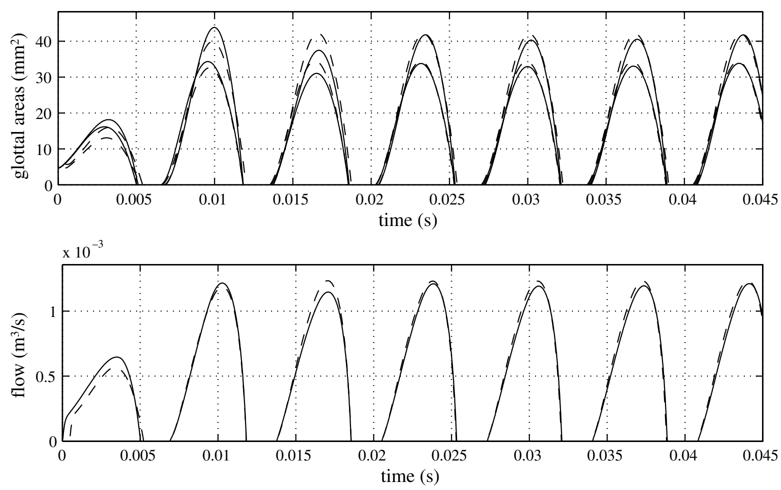


Fig. 8. Example of simulations from the two-mass model (solid lines) and from the pseudo-one-mass model (dotted lines), with vocal tract loading. Glottal area signals are shown in the top plot, flow signals are shown in the bottom plot.

Author's personal copy

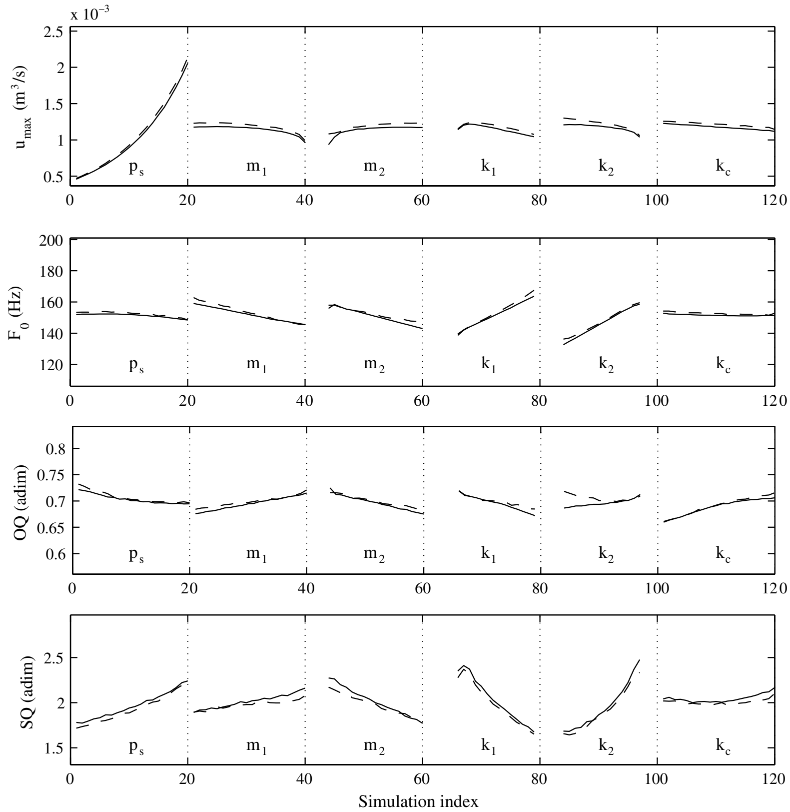


Fig. 9. Values of acoustic parameters extracted from flow signals of the reference two-mass model (solid lines) and the pseudo-one-mass model (dashed lines). Labels indicate the physical parameter that was varied during simulations. Ranges of variation for the physical parameters are as in Table 1 and Fig. 5.

as the main focus here is assessing the similarity of the flow waveforms produced by the two models. Moreover it is known that the reference two-mass model provides a poor description of the flow at small glottal apertures, which results in exceedingly high-absolute values of the negative derivative peak. For the same reason we do not include acoustical parameters correlated with the abruptness of glottal closure (e.g., the *return quotient RQ*) in our set.[2]

We restrict our analysis to the case of vocal tract loading, which provides more realistic flow signals. Fig. 9 reports the values of the four acoustic parameters u_{max}, F_0, OQ, SQ obtained from simulations of the models when one of the two-mass parameters is varied. It can be observed that the flow signals synthesized with the pseudo-one-mass model generally fit with good accuracy the ones synthesized with the two-mass model, in the considered region of the parameter space. Both F_0 and u_{max} are systematically overestimated, but the error remains very small (within 2.5% and 1.5%, respectively) everywhere. OQ values are also overestimated, consistently with the fact that our approximation (10) tends to overestimate the closing phase of the glottal area a_2 (this effect is visible in Fig. 4). The error for OQ is within 1.5% everywhere except for the lowest values of k_2, where it rises to 3–4%. Conversely, SQ values are systematically underestimated, with error within 3% everywhere except for extremal values of m_1, m_2, k_c, where it rises to 4–5%. The overall ranges for OQ and SQ are in good agreement with typical values for modal phonation: as an example, Alku and Vilkman (1996) report measured values for male speakers in the range 0.69–0.82 and 1.58–2.81 for OQ and SQ, respectively.

Although the above discussion is mainly directed at assessing the ability of the pseudo-one-mass model to fit the two-mass model, it also hints at the more general problem of studying the relationship between the physical parameters of a low-dimensional vocal fold model and the acoustic parameters that characterize different voice qualities. This is still an open challenge in voice production research, and it is often remarked that one of the main weakness of physical models lies in the difficulties in determining such a relationship. A recent study by Sciamarella and D'Alessandro (2004) has investigated the sensitivity of acoustic flow parameters to variation of physical parameters in the two-mass model of Lous et al. (1998). Their results provide indications of the "actions" that the model employs to target different voice qualities. However it has to be noted that low-level parameters like masses, spring stiffnesses, and so on, are not independently controlled by

[2] Lous et al. (1998) have shown that smoother glottal closures are obtained by including a viscosity term in the flow equations, but have also remarked that such a term is no more than an elegant way of fitting the model to experimental data.

Author's personal copy

a speaker: the use of more physiologically motivated control spaces is advisable, which requires to establish a link between physiology (muscle activations) and physics (geometrical and visco-elastic parameters of a lumped-element model). An attempt in this direction is reported in (Avanzini et al., 2006).

6. Conclusions

A new one-mass model of the vocal folds was presented, which is derived from a reference two-mass model previously proposed by Lous et al. (1998). Two-mass models account for effects of vertical phase differences in fold motion by means of two mass-spring oscillators that describe the motion of lower and upper fold portions. Instead, the proposed model employs a single oscillator to describe only the motion of the lower portion of the vocal fold, while phase difference effects are simulated through an approximate phenomenological description of the upper glottal area. Hence the name "pseudo-one-mass" model.

The approximate phenomenological description of the upper glottal area has been derived by exploiting constraints in the dynamic behavior of the reference two-mass model: in the hypothesis that 1:1 modal entrainment occurs between the two masses in the large-amplitude oscillation regime, then the displacement of the upper fold portion is approximately equal to that of the lower portion, delayed by a time τ and scaled by a factor α.

Numerical simulations of the reference two-mass model have been used to show that the proposed approximation remains valid when values of the physical parameters are varied in a large region of the control space. The same simulations have shown that the two phenomenological parameters α, τ are uniquely determined by, and have a smooth dependence on the physical parameters of the two-mass model. The effects of the phenomenological parameters α, τ on the shape of the glottal flow pulse were also analyzed.

Simulations of the reference two-mass model and the pseudo-one-mass model were compared in order to show that the dynamic behavior of the former is accurately approximated by the latter. Flow signals were synthesized with the two models in a large region of the control space, and the similarity of the signals was subsequently assessed in terms of four acoustic parameters: fundamental frequency, maximum amplitude, open quotient, and speed quotient. The results confirm that signals synthesized with the pseudo-one-mass model fit with good accuracy those synthesized with the two-mass model.

It has to be noted that the procedure used to derive the equations of the pseudo-one-mass model is independent on a specific description of the intraglottal pressure, and can be applied even when different reference two-mass models are used. The pseudo-one-mass model inherits, through the values of the parameters α, τ, the main properties of the reference two-mass model.

Applications of the proposed model are especially envisaged in the context of speech synthesis. Being a one-mass model, it requires significantly lower computational resources and about half of the control parameters with respect to the reference two-mass model, which makes it suitable for real-time implementation. Speech synthesis applications however require the model to be controllable with time-varying parameters. This is not trivially obtained, since the phenomenological parameters α, τ are themselves functions of the physical parameters. The mapping between physical and phenomenological parameters has to be stored beforehand into a look-up table, which will be interpolated at runtime in order to determine actual parameter values. This has not been implemented at present. However the results reported in this work show that the dependence of α, τ on the physical parameters is smooth and monotonic, which let us speculate that an implementation based on linear interpolation may already provide satisfactory results.

A second current limitation of the proposed model is that the results presented in this work demonstrate its applicability only for simulating modal phonation. Future work will be addressed at investigating the conditions for applying the model to non-modal phonation qualities, and to different oscillatory regimes.

Acknowledgements

The author is grateful to Paavo Alku and Carlo Drioli for their many remarks on a preliminary version of the manuscript, and to Simone Maratea for his help in developing the numerical implementation of the model. Three anonymous reviewers are acknowledged for their constructive criticism and helpful remarks, which led to a significantly improved version of the manuscript.

References

Alku, P., Vilkman, E., 1996. Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering. Speech Commun. 18 (2), 131–138.

Avanzini, F., Rocchesso, D., 2002. Efficiency, accuracy, and stability issues in discrete time simulations of single reed wind instruments. J. Acoust. Soc. Am. 111 (5), 2293–2301.

Avanzini, F., Alku, P., Karjalainen, M., 2001. One-delayed-mass model for efficient synthesis of glottal flow. In: Proc. Eurospeech Conf. Aalborg, pp. 51–54.

Avanzini, F., Maratea, S., Drioli, C., 2006. Physiological control of low-dimensional glottal models with applications to voice source parameter matching. Acta Acust. (united with Acustica) 92 (Suppl. 1), 731–740.

Berry, D.A., Titze, I.R., 1996. Normal modes in a continuum model of vocal fold tissues. J. Acoust. Soc. Am. 100 (5), 3345–3354.

Berry, D.A., Montequin, D.W., Tayama, N., 2001. High-speed digital imaging of the medial surface of the vocal folds. J. Acoust. Soc. Am. 110 (5), 2539–2547.

Childers, D.G., Ahn, C., 1995. Modeling the glottal volume–velocity waveform for three voice types. J. Acoust. Soc. Am. 97 (1), 505–519.

Childers, D.G., Wong, C.-F., 1994. Measuring and modeling vocal source-tract interaction. IEEE Trans. Biomed. Eng. 41 (7), 663–671.

Author's personal copy

Deverge, M., Pelorson, X., Vilain, C.E., Lagrée, P.-Y., Chentouf, F., Willems, J.F.H., Hirschberg, A., 2003. Influence of collision on the flow through in-vitro rigid models of the vocal folds. J. Acoust. Soc. Am. 114 (6), 3354–3362.

de Vries, M.P., Schutte, H.K., Verkerke, G.J., 1999. Determination of parameters for lumped parameter model of the vocal fold using a finite-element method approach. J. Acoust. Soc. Am. 106 (6), 3620–3628.

de Vries, M.P., Schutte, H.K., Veldman, A.E.P., Verkerke, G.J., 2002. Glottal flow through a two-mass model: Comparison of Navier–Stokes solutions with simplified models. J. Acoust. Soc. Am. 111 (4), 1847–1853.

Drioli, C., 2005. A flow waveform-matched low-dimensional glottal model based on physical knowledge. J. Acoust. Soc. Am. 117 (5), 3184–3195.

Fant, G., 1982. Preliminaries to analysis of the human voice. STL-QPSR 23 (4), 1–27.

Fant, G., Liljencrants, J., Guang Lin, Q., 1985. A four-parameter model of glottal flow. STL-QPSR 26 (4), 1–13.

Flanagan, J.L., Landgraf, L.L., 1968. Self-oscillating source for vocal-tract synthesizers. IEEE Trans. Audio Electroacoust. 16 (1), 57–64.

Flanagan, J.L., Ishizaka, K., Shipley, K.L., 1980. Signal models for low bit-rate coding of speech. J. Acoust. Soc. Am. 68 (3), 780–791.

Fletcher, N.H., 1978. Mode locking in nonlinearly excited inharmonic musical oscillators. J. Acoust. Soc. Am. 64 (6), 1566–1569.

Gunter, H.E., 2003. A mechanical model of vocal-fold collision with high spatial and temporal resolution. J. Acoust. Soc. Am. 113 (2), 994–1000.

Ishizaka, K., Flanagan, J.L., 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Syst. Technol. J. 51, 1233–1268.

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

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

Lucero, J.C., 1996. Chest- and falsetto-like oscillations in a two-mass model of the vocal folds. J. Acoust. Soc. Am. 100 (5), 3355–3359.

Pelorson, X., Hirschberg, A., van Hassel, R.R., Wijnands, A.P.J., 1994. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. J. Acoust. Soc. Am. 96 (6), 3416–3431.

Riegelsberger, E.L., Krishnamurthy, A.K., 1993. Glottal source estimation: Methods of applying the LF-model to inverse filtering. In: Proc. IEEE Int. Conf. Acoust. Speech and Signal Process. (ICASSP'93). Minneapolis, pp. 542–545.

Schroeter, J., Sondhi, M., 1992. Speech coding based on physiological models of speech production. In: Furui, S., Sondhi, M. (Eds.), Advances in Speech Signal Processing. Dekker, New York, pp. 231–263.

Schroeter, J., Sondhi, M., 1994. Techniques for estimating vocal-tract shapes from the speech signal. IEEE Trans. Speech Audio Process. 2 (I–II), 133–150.

Sciamarella, D., D'Alessandro, C., 2004. On the acoustic sensitivity of a symmetrical two-mass model of the vocal folds to the variation of control parameters. Acta Acust. (united with Acustica) 90 (4), 746–761.

Sondhi, M., Schroeter, J., 1987. A hybrid time–frequency domain articulatory speech synthesizer. IEEE Trans. Acoust., Speech, Sig. Process. 35 (7), 955–967.

Story, B.H., 2002. An overview of the physiology, physics and modeling of the sound source for vowels. Acoust. Sci. Technol. 23 (4), 195–206.

Story, B.H., Titze, I.R., 1995. Voice simulation with a body-cover model of the vocal folds. J. Acoust. Soc. Am. 97 (2), 1249–1260.

Strik, H., 1998. Automatic parametrization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses. J. Acoust. Soc. Am. 103 (5), 2659–2669.

Titze, I.R., 1988. The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. 83 (4), 1536–1552.

Titze, I.R., 2006. Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. J. Speech Lang. Hear. Res. 49 (2), 439–447.

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

Titze, I.R., Story, B.H., 2002. Rules for controlling low-dimensional vocal fold models with muscle activation. J. Acoust. Soc. Am. 112 (3), 1027–1064.

Vilain, C.E., Pelorson, X., Fraysse, C., Deverge, M., Hirschberg, A., Willems, J.F.H., 2004. Experimental validation of a quasi-steady theory for the flow through the glottis. J. Sound Vib. 276 (3–5), 475–490.