Effect of source–tract acoustical coupling on the oscillation onset of the vocal folds

Jorge C. Lucero[a)]
Department of Computer Science, University of Brasilia, Brasilia DF 70910-900, Brazil

Kélem G. Lourenço
Department of Mathematics, University of Brasilia, Brasilia DF 70910-900, Brazil

Nicolas Hermant, Annemie Van Hirtum, and Xavier Pelorson
GIPSA-lab, UMR CNRS 5216, Grenoble Universities, 961 rue de la Houille Blanche, BP 46, 38402 Saint-Martin d'Heres, France

(Received 9 November 2011; revised 22 February 2012; accepted 21 May 2012)

This paper analyzes the interaction between the vocal folds and vocal tract at phonation onset due to the acoustical coupling between both systems. Data collected from a mechanical replica of the vocal folds show that changes in vocal tract length induce fluctuations in the oscillation threshold values of both subglottal pressure and frequency. Frequency jumps and maxima of the threshold pressure occur when the oscillation frequency is slightly above a vocal tract resonance. Both the downstream and upstream vocal tracts may produce those same effects. A simple mathematical model is next proposed, based on a lumped description of tissue mechanics, quasi-steady flow and one-dimensional acoustics. The model shows that the frequency jumps are produced by saddle-node bifurcations between limit cycles forming a classical pattern of a cusp catastrophe. The transition from a low frequency oscillation to a high frequency one may be achieved through two different paths: in case of a large acoustical coupling (narrow vocal tract) or high subglottal pressure, the bifurcations are crossed, which causes a frequency jump with a hysteresis loop. By reducing the acoustical coupling (wide vocal tract) or the subglottal pressure, a path around the bifurcations may be followed with a smooth frequency variation.
 [http://dx.doi.org/10.1121/1.4728170]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt [AL] Pages: 403–411

I. INTRODUCTION

According to source–filter theory of phonation, the vocal tract acts as a linear filter excited by the sound produced at the larynx (Fant, 1970). The configuration of the vocal tract has no effect on the dynamics of the vocal fold oscillation, and therefore the latter may be studied as an isolated system. This conception has been a usual simplifying assumption to facilitate the analysis of the oscillation dynamics, and has permitted to gain an understanding of the main underlying mechanisms of phonation (see, e.g., Jiang and Tao, 2007; Lucero, 1999; Steinecke and Herzel, 1995; Titze, 1988; Zhang *et al.*, 2005). In fact, the separation between the glottal source and the vocal tract is a reasonable approximation as long as the fundamental frequency of the oscillation is well below the first resonance frequency (first formant) of the vocal tract, as in the case of normal male adult speech. However, at higher oscillation frequencies, as in female and child speech and in singing, source–tract interactions due to the acoustical coupling between both systems become significant. The coupling influences the oscillation onset and may cause frequency jumps, subharmonics, and other instabilities (Titze, 2008b). Similar phenomena may be produced also by lengthening artificially the vocal tract and, consequently, lowering the formants (Hatzikirou *et al.*, 2006).

The purpose of this paper is to investigate and model the acoustical coupling between the vocal fold oscillation and vocal tract. In recent experimental works, the oscillation has been simulated and studied by means of mechanical replicas (Lucero *et al.*, 2011; Ruty *et al.*, 2008; Zhang *et al.*, 2006, 2009). The replicas permit to gather data easily of different parameters (e.g., subglottal pressure, glottal aperture, oscillation frequency) under a variety of controlled configurations, and the data may be used to build and test theories of the oscillation. This approach has proven useful to, e.g., validate theoretical relations of phonation threshold pressure vs fundamental frequency (Lucero *et al.*, 2009) and hysteresis effects at phonation onset vs offset (Lucero *et al.*, 2011), assess the influence of glottal aperture and vocal fold stiffness on oscillation conditions (Cisonni *et al.*, 2011), and model nonlinear phenomena observed during singing (Bailly *et al.*, 2010). Particularly, our recent measures of the oscillation threshold subglottal pressure and fundamental frequency at the oscillation onset have shown fluctuations of those parameters as a function of the vocal tract length. Even drastic jumps of oscillation frequency may appear when crossing a vocal tract resonance. Here, the data will be reported and characterized with a simple physical model.

Onset pressure and frequency fluctuations have also been found in the experiments by Zhang *et al.* (2006, 2009)

a)Author to whom correspondence should be addressed. Electronic mail: lucero@unb.br

Their work focused on the production of aerodynamically vs acoustically driven modes of oscillation as a function of the subglottal and supraglottal acoustics loads. The present study follows a similar direction; however, a simple theoretical model in terms of the theory of dynamical systems is sought.

The proposed model comes from studies of birdsong production. In fact, source-tract interactions and frequency jumps are also relevant in the context of bird-song production (Zollinger *et al.*, 2008). In songbirds, sound is produced at the syrinx, located at the junction between the trachea and bronchi. Unlike the larynx, the syrinx contains two valves, each one similar to the glottis, located at the exit of each bronchus. However, the membranes at each valve oscillate by the same aeroelastic mechanism of the vocal folds. Frequency jumps are a characteristic signature of some songbirds, and may be produced by either alternate action of both sides of the syrinx (Laje and Mindlin, 2005) or source–tract interaction (Arneodo and Mindlin, 2009). The case of source–tract interaction was analyzed using a version of Titze's (1988) mucosal wave model coupled to waveguide analogy of the vocal tract. It was shown that the reflected acoustical wave from the bird's beak back to the syrinx interacts with the oscillation of the syrinx's membranes, in a way that depends on the relative phase between them.

The subject of source–tract interaction on human phonation has been recently analyzed by Titze (2008a,b), in a extensive work using models and data from subjects. The present paper has a similar objective, however, a much simpler representation of the interaction will be proposed. As in the work of Zhang *et al.* (2006), Titze's model of the vocal tract relies on concepts from the theory of electrical circuits. The model of Arneodo and Mindlin (2009), on the other hand, is based the propagation of plane acoustical waves and is therefore closer to the represented phenomenon. Further, it is elegant in its simplicity, which facilitates the analysis. A similar approach was previously used by Mergell and Herzel (1997) to simulate biphonation in asymmetrical vocal folds.

The subject of source–tract interaction is also relevant for its implications to vocal registers. Registers are distinct types of voice qualities over specific ranges of fundamental frequency and loudness, such as the chest and falsetto (Titze, 1994). Transitions between registers during speech and singing typically occur with voice discontinuities and a hysteresis effect (Svec *et al.*, 1999). A few theories have been proposed to explain the mechanism of the chest-falsetto transition, based on subglottal resonances and specific actions of laryngeal muscles (Titze, 1994). Although numerical simulations of the phenomenon using multimass representations of the vocal folds have been reported (Tokuda *et al.*, 2007, 2010), a clearer model of its dynamics is still desirable.

II. MODEL

A. Glottal source

The vocal fold model is based on an original characterization of the oscillation as a mucosal wave movement by Titze (1988), and is schematically shown in Fig. 1. It assumes complete right–left symmetry, and allows for motion of tissues only in the horizontal direction. A wave

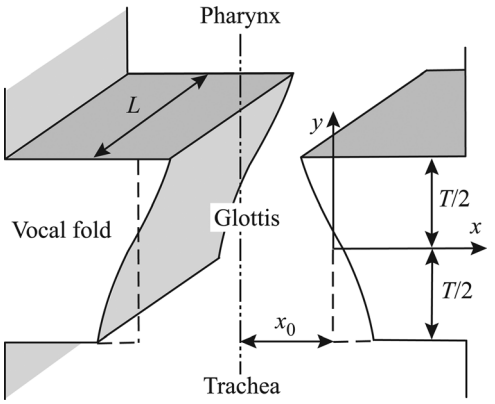

FIG. 1. Mucosal wave model of the vocal folds (Titze, 1988).

propagates through the superficial tissues, in the direction of the airflow (upward).

Its governing equations have been extensively discussed in the literature (see, e.g., Laje *et al.*, 2001, 2002; Lucero *et al.*, 2011; Titze, 1988), and may be summarized as follows. The equation of motion for the tissues is

$$M\ddot{x} + B[1 + \eta x^2(t)]\dot{x} + Kx = P_g, \tag{1}$$

where x is the tissue displacement at the midpoint of the glottis, M, B, and K are the mass, damping, and stiffness, respectively, per unit area of the vocal fold medial surface, η is a nonlinear dissipation coefficient, and P_g is the glottal mean air pressure. The glottal aerodynamics is characterized by

$$P_g = P_i + \frac{P_s - P_i}{k_t}\left(1 - \frac{a_2}{a_1} - k_e\right) \qquad (a_1 > 0), \tag{2}$$

where P_i is the supraglottal pressure (at the epilarynx), P_s is the subglottal pressure, a_1 and a_2 are the cross-sectional glottal areas at the lower and upper edges of the vocal folds, respectively, $k_t = k_c - k_e$ is a transglottal pressure coefficient, $1.0 < k_c \leq 1.4$ is a pressure loss coefficient for the glottal entry and $0 \leq k_e \leq 0.2$ is a pressure recovery coefficient for the glottal exit.

Further, the glottal areas are given by

$$a_1(t) = 2L_v[x_0 + x(t + \tau)], \tag{3}$$

$$a_2(t) = 2L_v[x_0 + x(t - \tau)], \tag{4}$$

where L_v is the vocal fold length, x_0 is the vocal fold displacement at rest (prephonatory position), and τ is the time delay for the surface wave to travel half the glottal height T.

B. Vocal tract

Following Arneodo and Mindlin (2009), the vocal tract downstream the glottis is represented as a hard-walled tube of area A_i and length L_i.

The total input pressure P_i is the sum of two components: A static pressure $\bar{P}_i$ and an acoustic pressure $\tilde{P}_i$,

$$P_i = \bar{P}_i + \tilde{P}_i. \tag{5}$$

Further, the acoustic pressure at the tube entry is the resultant of two components: an incident pressure wave injected into the tube from the glottis, denoted by p_i^+, and a backward propagating wave after partial reflection at the other end of the tube (the open mouth), denoted by p_i^-,

$$\tilde{P}_i = p_i^+ + p_i^-.$$

Doing similarly for the vocal tract upstream the glottis,

$$P_s = \bar{P}_s + \tilde{P}_s = \bar{P}_s + p_s^+ + p_s^-.$$

The incident pressure wave at the downstream tube is produced by variations of the airflow coming from the glottis

$$p_i^+ = \frac{\rho_0 c \tilde{u}}{A_i}, \tag{6}$$

where ρ_0 is the unperturbed air density, c is the sound speed, and $\tilde{u}$ is the varying component of the glottal air volume velocity.

The volume velocity at the glottis is $u = va$, where v is the particle velocity and $a = 2L_v[x_0 + x(t)]$ is the glottal area (for simplicity, variations along the glottal channel are disregarded here). According to Titze (1988), variations in the volume velocity are produced mainly by variations in the glottal area, and therefore the approximation $\tilde{u} = \bar{v} \cdot \tilde{a}$ may be adopted, where $\bar{v}$ is the mean particle velocity and $\tilde{a}$ is the glottal area variation. The vocal fold displacement is $x(t) = \bar{x} + \xi(t)$, where $\bar{x}$ is an equilibrium position (to be computed later) and $\xi(t)$ is the displacement from that position. Then, $\tilde{a}(t) = 2L_v\xi(t)$ and

$$p_i^+(t) = \alpha_i \bar{v} \xi(t), \tag{7}$$

where

$$\alpha_i = \frac{2L_v\rho_0 c}{A_i} \tag{8}$$

is a coupling coefficient.

The reflected wave may be simply expressed as $p_i^- = r_i p_i^+(t - v_i)$, where $-1 \le r_i \le 1$ is a reflection coefficient, and $v_i = 2L_i/c$ is the time delay for the acoustic wave to travel forth and back the vocal tract tube. Any attenuation suffered by the wave along the vocal tract is neglected. Therefore, the acoustic pressure at the downstream vocal tract entry is

$$\tilde{P}_i(t) = \alpha_i \bar{v}[\xi(t) + r_i \xi(t - v_i)]. \tag{9}$$

Similarly, for the upstream vocal tract

$$\tilde{P}_s(t) = -\alpha_s \bar{v}[\xi(t) + r_s \xi(t - v_s)], \tag{10}$$

where α_s, r_s, and v_s are the upstream coupling coefficient, reflection coefficient (at the lungs end) and acoustic wave delay to travel back and forth the upstream tube. The sign of the acoustical wave is negative, because the positive direction of the glottal flow is opposite to the direction of p_s^+ (from the glottis to the lungs).

Finally, let us consider the equilibrium position $\bar{x}$. Letting $x(t) = \bar{x}$ (constant), then the above-presented equations produce

$$K\bar{x} = \bar{P}_i - \frac{k_e}{k_t}\Delta\bar{P}, \tag{11}$$

where $\Delta\bar{P} = \bar{P}_s - \bar{P}_i$. Applying next the change of variable $x = \xi + \bar{x}$, the model becomes

$$M\ddot{\xi} + B[1 + \eta(\xi + \bar{x})^2]\dot{\xi} + K\xi = \frac{1}{k_t}[\Delta\bar{P}F_a + \tilde{P}_i(k_c - F_a) - \tilde{P}_s k_t(k_e - F_a)], \tag{12}$$

where F_a is the glottal area function

$$F_a = 1 - \frac{a_2}{a_1}. \tag{13}$$

In the previous equations, the mean particle velocity $\bar{v}$ may be approximated by

$$\bar{v} = \sqrt{2\Delta\bar{P}/(k_t\rho_0)} \tag{14}$$

(Alipour *et al.*, 2001; Fulcher *et al.*, 2006; Titze, 1988).

III. DATA

The source–tract interaction was studied experimentally in a mechanical replica of the vocal folds (Cisonni *et al.*, 2011; Ruty, 2007; Ruty *et al.*, 2005; Ruty *et al.*, 2007). The replica consists of two pieces of metal covered with latex, which mimics the vocal fold structure in a 3:1 scale and a 2 cm length. The space between each metal piece and the latex cover is filled with water at a controlled internal pressure. The downstream vocal tract is simulated with a cylindrical tube, with an internal diameter of 2.5 cm and variable lengths. An upstream cylindrical tube of the same diameter connects the replica to a large pressure reservoir fed by a constant pressure. The reservoir is filled with acoustical foam to avoid internal resonances. The pressure upstream of the replica (subglottal pressure) is measured by a pressure sensor (Kulite XCS 093) with a typical accuracy of ± 5 Pa, and the separation between the opposite latex folds (aperture) is measured using a laser setup with a typical accuracy of 0.01 mm. Details of the replica and several illustrations may be found in the previously cited references.

First, the mechanical response of the replica was measured in order to obtain its natural frequencies and respective Q (quality) factors, following a method by Gilbert *et al.* (1998). The latex folds were excited by an acoustical signal from a compression chamber located in its vicinity, without any physical contact. The signal's frequency was increased from 50 to 400 Hz at 1 Hz step, and at each step, the replica's aperture was measured. From the data, the amplitude and phase of the oscillatory motion of the latex folds were computed. The natural frequencies (peaks in the amplitude

response) and respective Q factors (bandwidth relative to the natural frequency) were obtained by fitting a quadratic polynomial in the vicinity of each peak (Robinson and Clegg, 2005). Figure 2 shows an example of the measured response.

Next, measures of the upstream pressure and the replica's aperture were obtained while increasing the air pressure in the reservoir from zero until an oscillation of the latex was detected. Then, the pressure was decreased until the oscillation was interrupted. Figure 3 shows an example of the measured data. The time instants of oscillation onset and offset were determined by spectral analysis on upstream pressure signal (Ruty *et al.*, 2005), and the mean upstream pressure and oscillation frequency at those times were computed. In the present study, only the onset values will be reported and used.

Several measures were taken at different lengths of the upstream and downstream tubes, simultaneously keeping all other parameters constant. In a first set of measures, the length L_i of the downstream tube was varied from 1.6 to 245.6 cm, at 55 different lengths, with a fixed upstream tube of $L_s = 15$ cm. In a second set, the downstream tube was removed and the upstream tube's length was varied from 8.5 to 176.8 cm, at 8 different lengths. The sets of measures were done with an interval of several days apart, and different latex covers in the replica. Therefore, the mechanical response of the replica was repeated before each set. The natural frequencies and Q factors for the first set are those reported in Fig. 2. For the second set, the mean values of the first three resonance frequencies of the replica are 130, 172, and 264 Hz, and their Q factors are 12.7, 17.5, and 13.6, respectively. The initial aperture of the replica (measured without airflow) was $h_0 = 1.33$ mm for the first set, and $h_0 = 1.50$ mm in the second set.

IV. COMPARISON OF THE MODEL WITH THE DATA

A. Equations of oscillation threshold pressure

The linearization of Eq. (12) is

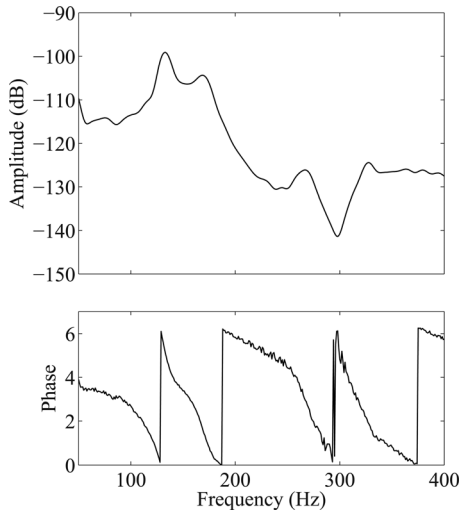

FIG. 2. Example of the mechanical response of the replica. The first three resonance frequencies are 132, 168, and 266 Hz, and their Q factors are 10.5, 8.5 and 12.6, respectively.

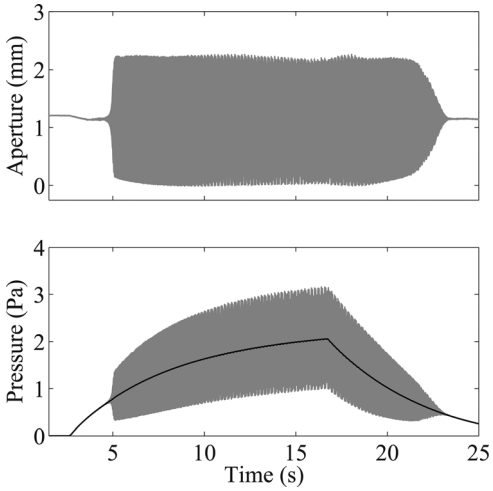

FIG. 3. Example of the data collected from the replica when varying the upstream pressure. (Upper plot) Aperture (separation) between the two latex folds. (Lower plot) Pressure immediately upstream the replica. The black curve in the lower plot is the mean pressure, computed by low-pass filtering the measured pressure.

$$M\ddot{\xi} + B\dot{\xi} + K\xi = \frac{\Delta\bar{P}}{k_t(x_0+\bar{x})}[\xi(t+\tau) - \xi(t-\tau)] + \frac{k_c\alpha_i\bar{v}}{k_t}[\xi(t) + r_i\xi(t-v_i)] + \frac{k_e\alpha_s\bar{v}}{k_t}[\xi(t) + r_s\xi(t-v_s)]. \quad (15)$$

Proposing a solution of the form $\xi(t) = Ce^{\lambda t}$, where C and λ are complex constants, and seeking nonzero solutions, produces the associated characteristic equation

$$M\lambda^2 + B\lambda + K = \frac{2\Delta\bar{P}}{k_t(x_0+\bar{x})}\sinh(\lambda\tau) + \frac{k_c\alpha_i\bar{v}}{k_t}(1 + r_i e^{-\lambda v_i}) + \frac{k_e\alpha_s\bar{v}}{k_t}(1 + r_s e^{-\lambda v_s}). \quad (16)$$

The conditions to start an oscillation may be found by applying Hopf's bifurcation theorem for functional differential equations (Hale and Lunel, 1993). Assuming a pair of imaginary roots ($\lambda = \pm i\omega$) of Eq. (16), and separating real and imaginary parts, we find

$$-M\omega^2 + K = \frac{\bar{v}}{k_t}[k_c\alpha_i(1 + r_i\cos\omega v_i) + k_e\alpha_s(1 + r_i\cos\omega v_s)] \quad (17)$$

and

$$\omega B = \frac{2\Delta\bar{P}}{k_t(x_0+\bar{x})}\sin(\omega\tau) - \frac{\bar{v}}{k_t}[k_c\alpha_i r_i\sin\omega v_i + k_e\alpha_s r_s\sin\omega v_s]. \quad (18)$$

It may be shown that at the value of $\Delta\bar{P}$ defined by the previous equations, the roots cross the imaginary axes of the

complex plane from left to right, and that all other roots of the characteristic equation have negative real parts (Lucero and Koenig, 2007). Therefore, the equations define the oscillation threshold value of the transglottal pressure $\Delta\bar{P}$ and the frequency ω at the oscillation onset.

The threshold equations clearly show that the vocal tract introduces fluctuations in both the oscillation frequency and the transglottal pressure at oscillation onset. The effects of the upstream and downstream tracts are additive, in agreement with the findings of Titze (2008a). Note that the acoustical terms related to the upstream and downstream vocal tracts are scaled by coefficients k_e and k_c, respectively. As $k_c \gg k_e$, then the downstream vocal tract has a much larger effect than the upstream tract. If the upstream vocal tract is neglected, with $k_e = 0$, and the downstream vocal tract is removed ($\alpha_i = 0$ or $v_i = 0$), the equations reduce to $\omega^2 = K/M$ and $\bar{P}_s = k_t x_0 B\omega/[2\sin(\omega\tau)]$, which were already determined by Lucero and Koenig (2007).

B. Results when varying the length of the downstream tube

From the mechanical response of the replica, the second resonance was adopted as its natural frequency $f_n = \sqrt{K/M}/(2\pi) = 168$ Hz. This frequency was the closest resonance to the measured oscillation frequencies (which were in the range 150–180 Hz), and provided the best match of the theory with the data. Its Q factor is $Q = \sqrt{MK}/B = 8.5$.

The oscillating mass of the replica is mainly determined by the body of water, because the mass of the latex cover may be neglected in comparison. From the geometrical dimensions of the replica, the volume of water contained in the replica was estimated as 0.63 cm^3 (Ruty, 2007). The medial area of the latex surface, exposed to the airflow, is 1.57 cm^2. These values would produce a mass per unit area $M = 0.4$ g/cm^2. Now, probably not all the mass of water and latex surface are involved in the oscillatory motion. However, we keep the above-mentioned value as a crude estimate. With that value for M, we obtain $K = 446$ kdyn/cm^3 and $B = 49.7$ dyn s/cm^3. The prephonatory rest position is $x_0 = 0.67$ mm, computed as half the measured aperture between the opposite latex folds. An estimate of the time delay τ may be determined assuming a phase delay (δ) between the oscillation at the entry and exit of the latex channel of 60° (Titze, 1988), which, at an oscillation of 168 Hz (natural frequency), corresponds to $\tau = \delta/(4\pi f) = 0.5$ ms. We assumed further $k_c = 1.2$.

The time delay for the acoustical wave to travel back and forth the downstream tube is $v_i = 2L_i/c$, where $c = 343.2$ m/s is the speed of sound in dry air and 20 °C. The reflection coefficient at the mouth end is $r_i = (z_0 - z_i)/(z_0 + z_i)$, were z_0 is the radiation impedance and $z_i = \rho_0 c/A_i$ is the acoustic impedance of the tube. The radiation impedance was computed using the low-frequency approximation for a flanged pipe in free space

$$z_0 \approx \frac{1}{2}\rho_0 c A_i \frac{jka}{1+jka}, \tag{19}$$

where $k = 2\pi f/c$ and $a = \sqrt{A_i/\pi}$ (Lous *et al.*, 1998). For frequency values below 200 Hz, $\rho_0 = 1.204$ kg/m^3 and other parameters as indicated previously, the above-presented equations produce $-1 \le r_i \le -0.997$ (with an imaginary part smaller than 0.04), and therefore, a value of $r_i = -1$ was adopted. Due to the large area of the tube, the static pressure at its entrance is $\bar{P}_i \approx 0$. In a first approximation, the effect of the upstream tube was neglected by assuming $k_e \approx 0$.

Figure 4 shows the collected data and theoretical results with the above-mentioned values (curve 1). We may note large fluctuations in the threshold pressure and frequency caused by the acoustical coupling with the downstream tube. There are even frequency jumps when the oscillation frequency is slightly above a tube resonance. At the jumps, the threshold pressure is at a maximum. The theory matches the data in good qualitative approximation, considering the extreme simplifying assumptions of the mathematical model, and the crude estimates of some of its parameters. The theoretical threshold pressure is lower than the data, and the frequency shows steeper decreases. However, both the pressure and the frequency follow the same pattern as the data.

To investigate if a better match of the theory with the data could be obtained, a least squares optimization procedure was run with τ and M as parameters (these two parameters were selected because the proposed values mentioned previously were crude estimates). Their optimal values were determined by minimizing the sum of squared differences between measured and computed values of oscillation frequency $f = \omega/(2\pi)$ and static subglottal pressure $\bar{P}_s$. The result of the optimization was $\tau = 1.57$ ms and $M = 1.38$ g, which produce curve 2 in Fig. 4. The match with the data is now very good, although the fluctuations of the onset frequency are smoother and do not show drastic jumps as the data. The large differences between the optimal values of M and τ and their respective initial estimates may be attributed to the uncertainty of the vibration portion of the mass of water and latex surface, and of the actual value of the time

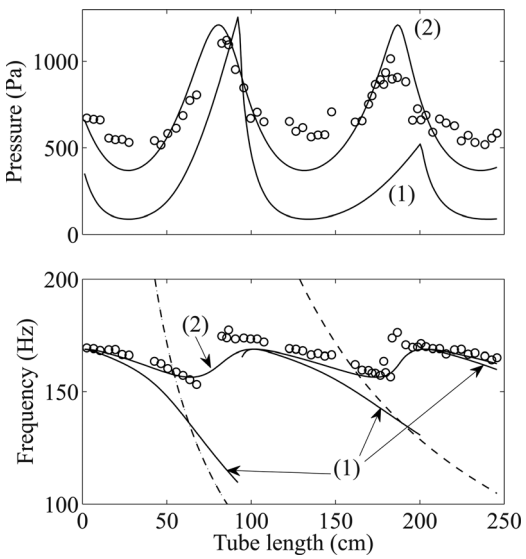

FIG. 4. Subglottal pressure (top) and oscillation frequency (bottom) at oscillation onset vs length of the downstream tube, for a fixed upstream tube. Data (circle); theoretical result (solid curve); resonance frequencies of the downstream tube (broken curve in the bottom plot). Curve 1 is the theoretical result with $\tau = 0.5$ ms and $M = 0.4$ g/cm^2. Curve 2 is the theoretical result with $\tau = 1.75$ ms and $M = 1.38$ g/cm^2.

delay of the latex wave motion. For example, the same mass of water on a latex surface of one-third the initial estimate produces a value of M close to the optimal value. Further, the optimal time delay is in the order of standard values for the vocal fold oscillation. For example, a mucosal wave propagating at 100 cm/s on a vocal fold height of 3 mm (Titze, 1988) produces a value of $\tau = 1.5$ ms.

C. Results when varying the length of the upstream tube

Prior to the second set of measures, the mechanical response of the replica was measured again. This time, its second resonance was 172 Hz and its corresponding Q factor was 17.5. Assuming, again, $M = 0.4\ \mathrm{g/cm^2}$, we obtain $K = 467\ \mathrm{kdyn/cm^3}$ and $B = 24.7\ \mathrm{dyn\ s/cm^3}$. The prephonatory rest position was $x_0 = 0.75$ mm.

For the upstream tube, the time delay for the acoustical wave to travel back and forth is $v_s = 2L_s/c$. Its reflection coefficient was estimated by considering that it opens into the reservoir, where no acoustical wave may propagate due to the acoustical foam. Therefore, the radiation impedance is zero and $r_s = -1$. In case of the actual phonatory system, the lungs may be modeled as an exponential horn with a cutoff frequency around 500 Hz (Lous *et al.*, 1998). All acoustical waves below that frequency are reflected, and so the reflexion coefficient is also $r_s = -1$.

The pressure recovery coefficient has been estimated as $k_e = 2(a_2/A_i)(1 - a_2/A_i)$ (Ishizaka and Matsudaira, 1972). Letting $a_2 = 2Lx_0 = 0.3\ \mathrm{cm^2}$ and using $A_i = 4.9\ \mathrm{cm^2}$, we obtain $k_e = 0.11$.

Figure 5 shows the collected data and the theoretical results. Again, we note the large fluctuations in transglottal pressure and oscillation frequency, and a frequency jump slightly above the tube's first resonance. Curve 1 shows the theoretical prediction with the above-mentioned parameters, which do not reproduce the observed fluctuations and frequency jump. For a better match, a least squares optimization procedure was again run with τ and M as parameters. The result of the optimization was $\tau = 0.025$ ms and $M = 0.064$ g, which produce curve 2 in Fig. 5. The match with the data is now reasonably good, although the predicted transglottal pressure has values much lower than the measured ones. In this case, the optimal values found for M and τ are much lower than those found for the first set of measures (in the previous section), and suggest a small mass moving laterally with almost no wave-like motion of the latex cover ($\delta \approx 3°$). It might be possible that the use of a new latex cover for the second set of measures, under slight different adjustments, or the absence of a downstream vocal tract had triggered a different mode of oscillation of the latex cover.

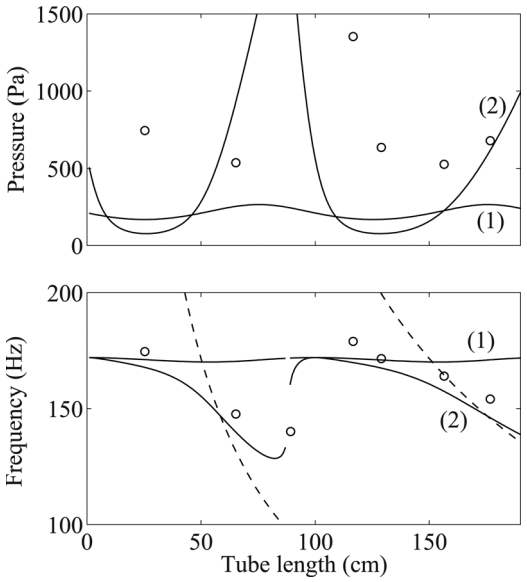

FIG. 5. Subglottal pressure (top) and oscillation frequency (bottom) at oscillation onset vs length of the upstream tube, without downstream tube. Data (circle); theoretical result (solid curve); resonance frequencies of the upstream tube (broken curve in the bottom plot). Curve 1 is the theoretical result with the same parameters used for curve 1 in Fig. 4 and $k_e = 0.11$. Curve 2 is the theoretical result with $\tau = 0.025$ ms and $M = 0.064$ g, and other parameters as presented earlier. There is one measure not shown in the top plot, with $L = 89.3$ cm and $\bar{P}_s = 2600$ Pa, which is outside the vertical range. Also, for $L = 8.5$ cm no oscillation could be detected.

V. DYNAMICS OF THE FREQUENCY JUMPS

Let us consider next the production of frequency jumps. To isolate its main mechanism from any influence of the pressure variations shown in the plots of the previous sections, we consider a constant static subglottal pressure above the onset threshold. Figure 6 presents a simulation of the oscillation with a static subglottal pressure of 2500 Pa and a

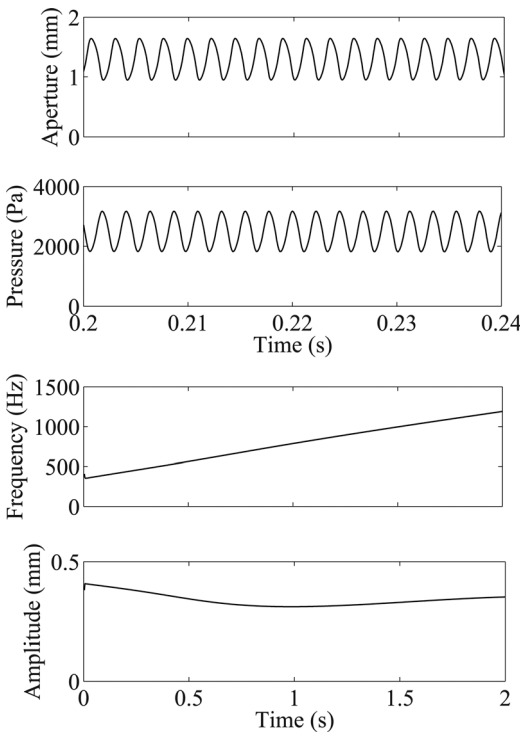

FIG. 6. Simulation of the latex oscillation when varying f_0 from 400 to 1200 Hz in a time period of 2 s, with $P_s = 2500$ Pa, $L_i = 17.16$ cm (first resonance at 500 Hz), $B = 10\ \mathrm{dyn\ s/cm^3}$, $\tau = 1$ ms, $\eta = 10^6\ \mathrm{cm^{-2}}$, $k_e = 0.11$ and other parameters as in curve 1 of Fig. 4. The two upper panels show a portion of the vocal fold aperture and the subglottal pressure. The third panel is the oscillation frequency, computed as the inverse of the period of each cycle, and the bottom panel is the amplitude.

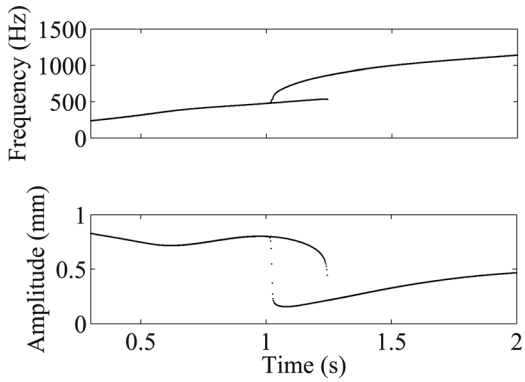

FIG. 7. Simulation of the latex oscillation when the cross-sectional area of the downstream tube is one-tenth of the value used for Fig. 6. The top panel shows the oscillation frequency, computed as the inverse of the period of each cycle, and the bottom panel is the amplitude.

varying natural frequency from 400 to 1200 Hz. As shown, the model is able to produce a stable oscillation, although its amplitude must be small so that the glottal closure is never achieved. For the simulations, a first resonance of the downstream tube of 500 Hz was adopted, which corresponds to a length of 17.2 cm. The subglottal pressure varies in synchrony with the vocal fold oscillation: Maxima of the subglottal pressure occur at the minima of the glottal aperture. The variation is consequence of the acoustical wave in the upstream tube, and may take a large value, as revealed by Fig. 3.

Figure 7 shows results when the downstream acoustical coupling is increased 10 times by reducing the area of the tube. Each plot was constructed by increasing first the natural frequency from 400 to 1200 Hz, and next decreasing the frequency in the same interval. In this case, we observe frequency and amplitude jumps when the oscillation frequency crosses the first resonance of the tube, with a hysteresis loop.

The frequency jumps may be analyzed using Eq. (17). The equation was derived for the oscillation onset, but it also holds as an approximation when the oscillation amplitude is

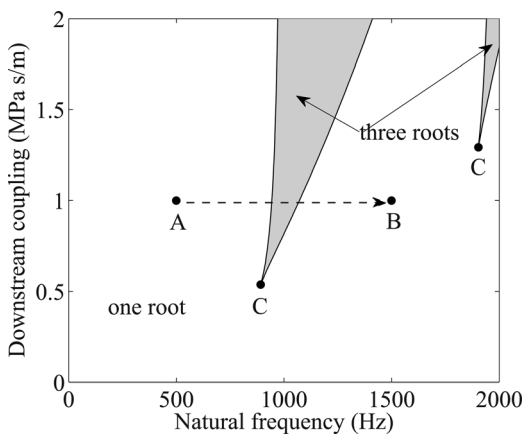

FIG. 8. Number of roots of Eq. (20). On each of the curves that separate the one- and three-root regions, the equation has two roots and a saddle-node bifurcation between limit cycles occur, in which a stable and an unstable limit cycle coalesce and cancel each other. The trajectory from point *A* to *B* represents the simulation plotted in Fig. 10. Point *C* indicates a cusp catastrophe.

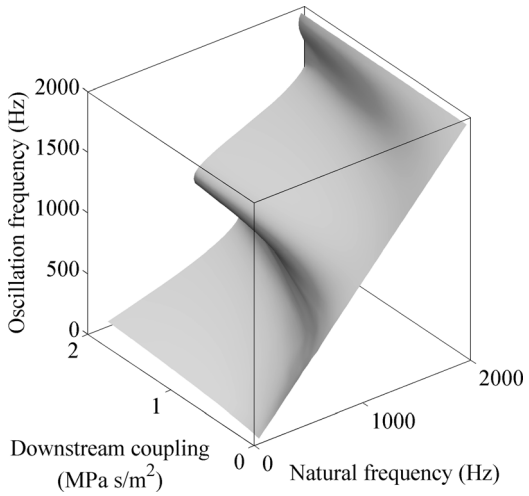

FIG. 9. 3D view of the results in Fig. 8.

small (i.e., much smaller than the rest position x_0), as in the case of Figs. 6 and 7.

Letting $f_0 = \sqrt{K/M}/(2\pi)$ be the natural frequency, and introducing the first resonance frequencies of the downstream and upstream tubes, $f_{i1} = 1/(2v_i)$ and $f_{s1} = 1/(2v_s)$, respectively, Eq. (17) becomes

$$f_0^2 - f^2 = \frac{\bar{v}}{4\pi^2 k_t M}\left[k_c\alpha_i\left(1-\cos\frac{\pi f}{f_{i1}}\right)+k_e\alpha_s\left(1-\cos\frac{\pi f}{f_{s1}}\right)\right]. \tag{20}$$

When the coupling with the vocal tracts is $\alpha_i = \alpha_s = 0$, Eq. (20) has the unique solution $f = f_0$; i.e., the vocal folds oscillate at their natural frequency. As the coupling increases (e.g., by reducing the area of the downstream vocal tract), multiple solutions might appear. Figure 8 shows a diagram of the regions with different numbers of roots, for $f_{i1} = 500$ Hz and other parameters as in Figs. 6 and 7. The existence of n roots in a particular region indicates n oscillatory solutions (limit cycles).

Figure 9 shows a 3D view of the same results. The 3D surface folds on itself and forms the characteristic shape of a

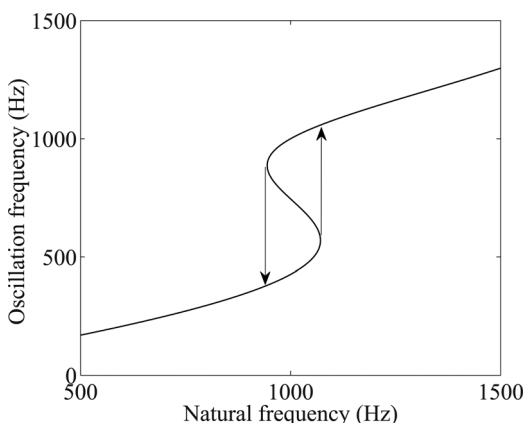

FIG. 10. Oscillation frequency when moving from point *A* to *B* in Fig. 8 across the fold. The arrows indicate frequency jumps, forming a hysteresis loop.

cusp catastrophe. The folded surface causes a frequency jump with hysteresis loop, when a point moves from the lower part of the surface (point *A* in Fig. 8) to the upper part (point *B* in Fig. 8) crossing the fold. Figure 10 illustrates the frequency jumps. If, instead of crossing the fold, the point moves from *A* to *B* following a path around the cusp (point *C* in Fig. 8), then the jump will not occur.

Equation (20) also reveals that the occurrence of frequency jumps also depends on the static value of the transglottal pressure, through the mean air particle velocity $\bar{v}$ [see also Eq. (14)]. Increasing $\bar{v}$ has the same effect as increasing the acoustical coupling of the vocal tract (α_i and α_s).

VI. CONCLUSIONS

The threshold values of both the subglottal pressure and frequency of the vocal fold oscillation are sensitive to the acoustical coupling between glottal source and the upstream and downstream vocal tracts. Fluctuations of both values occur as the resonance frequencies of the upstream and downstream vocal tracts are varied by changing their length, and may include drastic jumps of the oscillation frequency. The acoustical coupling with the upstream and downstream vocal tracts have additive effects; however, the influence of the downstream coupling is much larger than the upstream one.

The oscillation frequency jumps occur when the frequency is slightly above a vocal tract resonance, in coincidence with maxima of the threshold pressure. From the point of view of theory of dynamical systems, the jumps are produced by saddle-node bifurcations between limit cycles forming a classical pattern of a cusp catastrophe. Moving from a low frequency oscillation to a high frequency one may be achieved through two different paths: Crossing the saddle-node bifurcations, with a frequency jump and a hysteresis loop, or moving around the cusp, with a smooth transition from the low to the high frequency. The selection of one or the other path depends on the level of acoustical coupling, which is inversely proportional to the cross-sectional area of the vocal tract, and also on the level of transglottal pressure. Thus, a narrow vocal tract results in a large coupling, which will produce the frequency jumps, whereas a wide vocal tract results in a low coupling and a smooth frequency transition. The same effects may be achieved by keeping the coupling constant, and setting a large transglottal pressure vs a smaller one, respectively.

The above-mentioned model might also describe voice discontinuities observed at chest-falsetto register transitions. The transitions might be caused by the acoustical coupling with the supraglottal tract, in agreement with computer simulations by Tokuda *et al.* (2010), and also with the subglottal tract, as proposed by Titze (1994). Yet, it has been claimed that the vocal fold oscillation may produce chest-falsetto transitions also in the absence of a vocal tract, as shown in experiments with excised larynges (Berry *et al.*, 1996; Svec *et al.*, 1999). Although theoretical models based on multimass representations of the vocal folds are capable of reproducing the data without any vocal tract acoustical load (Tokuda *et al.*, 2007), the origin of the observed transitions is not clear. In fact, Zhang *et al.* (2006) pointed out that the experimental setups contained a subglottal tube to pass air to the larynx, whose length was unspecified. Therefore, it may be possible that the transitions were caused by acoustical interaction with the subglottal system.

The results may be used to interpret recent observations by Koenig *et al.* (2011) in subjects producing speech. They found that the intraoral pressure at voicing onset following an obstruent consonant is always larger for a high vowel than a low vowel. High vowels have a lower first formant than low vowels; e.g., typical values for /i/ vs /a/ are 270 Hz vs 730 Hz (male values) and 310 Hz vs 850 Hz (female values) (Stevens, 1998). Assuming an approximately constant subglottal pressure in each subject, then the data implies a lower transglottal pressure for a larger formant. In the case of the replica used in the present study, the above-mentioned formant values would correspond to downstream tube lengths around 30 cm for /i/ and 11 cm for /a/, which fall in the left-hand side of the plots in Fig. 4. Further, as the static downstream pressure is approximately atmospheric, then the vertical axis may be also interpreted as transglottal pressure. Therefore, both the measures and theoretical prediction show a higher transglottal threshold at an 11 cm tube length (/a/) vs 30 cm (/i/), in agreement with the speech data.

It is interesting to note also that both the experiments and theory show that, except for the regions where the oscillation frequency is close to a tube resonance, the oscillation frequency always decreases when the tube's length increases. This pattern is in the opposite direction to the well-known "intrinsic F0 effect" in speech: Voice's fundamental frequency is higher in high vowels than low vowels (Whalen and Levitt, 1995). In terms of equivalent downstream tube lengths, the effect would mean that the oscillation frequency should increase with the tube length. Its origin is still not clear and under debate. However, our results suggest that the effect is not caused by acoustical interaction with the vocal tract.

The results also agree with several past modeling studies, which have found that the inertia of the column of air downstream the glottis has the effect of enhancing the vocal fold oscillation, when the oscillation frequency is below the first formant (see, e.g., Titze, 1988). However, Fig. 4 shows that, below the first formant, the vocal tract effect has a "U" shape curve and so the enhancing effect only occurs in the left-hand side of that curve. The theoretical minimum occurs at 27.1 cm (both curves 1 and 2), which corresponds to a first resonance frequency of 316.6 Hz. At that length, the theoretical oscillation frequency is 161 Hz (curve 1) and 164 Hz (curve 2), which is about half the resonance frequency. In fact, it is easy to show that in the case of a constant oscillation frequency, a minimum occurs when the resonance frequency is exactly twice the oscillation frequency.

Our study is also in general agreement with the theory of nonlinear source-filter coupling developed by Titze (2008a,b) based on multimass model simulations and vocal exercises from human subjects. We stress again the helpful resource provided by the mechanical replica, which allows us to test theories by easily collecting relevant data in a controlled experimental setup. Further, the simplicity of the

proposed theoretical model permits to isolate and analyze main mechanisms of the oscillation. On the other hand, better qualitative agreement with the data demands a more sophisticated model, with better flow and acoustic models.

ACKNOWLEDGMENTS

This work was done while J.C.L. was an Associated Researcher of the Centre Nacional de la Recherche Scientifique (Poste Rouge CNRS) at GIPSA-lab, in Grenoble (France), and K.G.L. was a Ph.D. exchange student at the same laboratory under the Erasmus Mundus EU-Brazil Startup Program. J.C.L. was also supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) of Brazil.

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

Arneodo, E. M., and Mindlin, G. B. (**2009**). "Source-tract coupling in birdsong production," Phys. Rev. E **79**, 061921.

Bailly, L., Henrich, H., and Pelorson, X. (**2010**). "Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling," J. Acoust. Soc. Am. **127**, 3212–3222.

Berry, D., Herzel, H., Titze, I. R., and Story, B. H. (**1996**). "Bifurcations in excised larynx experiments," J. Voice **10**, 129–138.

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

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague), pp. 1–26.

Fulcher, L. P., Scherer, R. C., Melnykov, A., Gateva, V., and Limes, M. E. (**2006**). "Negative coulomb damping, limit cycles, and self-oscillation of the vocal folds," Am. J. Phys. **74**, 386–393.

Gilbert, J., Ponthus, S., and Petiot, J. F. (**1998**). "Artificial buzzing lips and brass instruments: Experimental results," J. Acoust. Soc. Am. **104**, 1627–1632.

Hale, J. K., and Lunel, S. M. V. (**1993**). *Introduction to Functional Differential Equations* (Springer, New York), pp. 331–335.

Hatzikirou, H., Fitch, W. T., and Herzel, H. (**2006**). "Voice instabilities due to source-tract interactions," Acta Acust. United Acust. **92**, 468–475.

Ishizaka, K., and Matsudaira, M. (**1972**). "Theory of vocal cord vibration," Rep. Univ. Electro-Comm. **23**, 107–136.

Jiang, J. J., and Tao, C. (**2007**). "The minimum glottal airflow to initiate vocal fold oscillation," J. Acoust. Soc. Am. **121**, 2873–2881.

Koenig, L. L., Fuchs, S., and Lucero, J. C. (**2011**). "Effects of consonant manner and vowel height on intraoral pressure and articulatory contact at voicing offset and onset for voiceless obstruents," J. Acoust. Soc. Am. **129**, 3233–3244.

Laje, R., Gardner, T. J., and Mindlin, G. B. (**2001**). "Continuous model for vocal fold oscillations to study the effect of feedback," Phys. Rev. E **64**, 056201.

Laje, R., Gardner, T. J., and Mindlin, G. B. (**2002**). "Neuromuscular control of vocalizations in birdsong: A model," Phys. Rev. E **65**, 051921.

Laje, R., and Mindlin, G. B. (**2005**). "Modeling source-source and source-filter acoustic interaction in birdsong," Phys. Rev. E **72**, 036218.

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

Lucero, J. C. (**1999**). "Theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. **105**, 423–431.

Lucero, J. C., and Koenig, L. L. (**2007**). "On the relation between the phonation threshold lung pressure and the oscillation frequency of the vocal folds," J. Acoust. Soc. Am. **121**, 3280–3283.

Lucero, J. C., Koenig, L. L., Lourenço, K. G., Ruty, N., and Pelorson, X. (**2011**). "A lumped mucosal wave model of the vocal folds revisited: Recent extensions and oscillation hysteresis," J. Acoust. Soc. Am. **129**, 1568–1579.

Lucero, J. C., Van Hirtum, A., Ruty, N., Cisonni, J., and Pelorson, X. (**2009**). "Validation of theoretical models of phonation threshold pressure with data from a vocal fold mechanical replica," J. Acoust. Soc. Am. **125**, 632–635.

Mergell, P., and Herzel, H. (**1997**). "Modelling biphonation—The role of the vocal tract," Speech Commun. **22**, 141–154.

Robinson, M. P., and Clegg, J. (**2005**). "Improved determination of q-factor and resonant frequency by a quadratic curve-fitting method," IEEE Trans. Electromagn. Compat. **47**, 399–402.

Ruty, N. (**2007**). "Modèles d'interactions fluide parois dans le conduit vocal. applications aux voix et aux pathologies (Models of fluid-wall interactions in the vocal tract. Applications to voice and pathologies)," Ph.D. thesis, Institut National Polytechnique de Grenoble, Grenoble, France, pp. 1–189.

Ruty, N., Pelorson, X., and Van Hirtum, A. (**2008**). "Influence of acoustic waveguides lengths on self-sustained oscillations: Theoretical prediction and experimental validation," J. Acoust. Soc. Am. **123**, 3121.

Ruty, N., Pelorson, X., Van Hirtum, A., Lopez-Arteaga, I., and Hirschberg, A. (**2007**). "An in vitro setup to test the relevance and the accuracy of low-order vocal folds models," J. Acoust. Soc, Am. **121**, 479–490.

Ruty, N., Van Hirtum, A., Pelorson, X., Lopez, I., and Hirschberg, A. (**2005**). "A mechanical experimental setup to simulate vocal fold vibrations. preliminary results," ZAS Papers in Linguistics **40**, 161–175; also available as arXiv:0710.

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

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA), p. 288.

Svec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**, 1523–1531.

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

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ), pp. 252–271.

Titze, I. R. (**2008a**). "Nonlinear-source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.

Titze, I. R. (**2008b**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.

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

Tokuda, I. T., Zemke, M., Kob, M., and Herzel, H. (**2010**). "Biomechanical modeling of register transitions and the role of vocal tract resonators," J. Acoust. Soc. Am. **127**, 1528–1536.

Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic F_0 of vowels," J. Phonetics **23**, 349–366.

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

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

Zhang, Z., Neubauer, J., and Berry, D. A. (**2009**). "Influence of vocal fold stiffness and acoustic loading on flow-induced vibration of a single-layer vocal fold model," J. Sound Vib. **322**, 299–313.

Zollinger, S. A., Riede, T., and Suthers, R. A. (**2008**). "Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations," J. Exp. Biol. **211**, 1978–1991.