Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control

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

Laura L. Koenig[b]
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut, 06511 and Long Island University, Brooklyn, New York 11201-8423

In this study we use a low-dimensional laryngeal model to reproduce temporal variations in oral airflow produced by speakers in the vicinity of an abduction gesture. It attempts to characterize these temporal patterns in terms of biomechanical parameters such as glottal area, vocal fold stiffness, subglottal pressure, and gender differences in laryngeal dimensions. A two-mass model of the vocal folds coupled to a two-tube approximation of the vocal tract is fitted to oral airflow records measured in men and women during the production of /aha/ utterances, using the subglottal pressure, glottal width, and Q factor as control parameters. The results show that the model is capable of reproducing the airflow records with good approximation. A nonlinear damping characteristics is needed, to reproduce the flow variation at glottal abduction. Devoicing is achieved by the combined action of vocal fold abduction, the decrease of subglottal pressure, and the increase of vocal fold tension. In general, the female larynx has a more restricted region of vocal fold oscillation than the male one. This would explain the more frequent devoicing in glottal abduction–adduction gestures for /h/ in running speech by women, compared to men. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1853235]

I. INTRODUCTION

Our purpose in this paper is to explore the capability of low-dimensional vocal fold models to reproduce vocal fold vibration onset–offset patterns observed experimentally during speech. At the same time, it analyzes how the vibratory behavior of the vocal folds is controlled during speech by individual speakers. The analysis follows an inverse dynamic approach, based on previous modeling studies by McGowan *et al.* (1995). Basically, this approach consists in fitting a model of the glottal source and vocal tract to collected speech records. The fit is performed by computing the time-varying laryngeal parameters that best reproduce the given speech signal at the model output. It must be done in a dynamic fashion, since the characteristics of the targeted speech signal, such as frequency and amplitude, vary in time. In this way, it represents a more demanding test for the vocal fold model than usual simulations of sustained voice, in which the laryngeal configuration is static. The model must be capable of simulating the pattern of variation of the speech signal, and within restrictions of smoothness and physiologically realistic ranges of the control parameters.

We consider airflow signals collected during productions of the utterance /aha/. This utterance is convenient for studying laryngeal control during speech, since it involves an open vocal tract with little upper articulatory movement (Koenig, 2000). Glottal abduction and subsequent adduction, com-bined with variations of other laryngeal biomechanical parameters such as subglottal pressure and vocal fold stiffness, yield elevated airflow, breathy voicing, and, in some cases, a cessation of voicing during /h/. By fitting a model of the larynx to the airflow data, we seek to infer the temporal control patterns of those parameters. Our purpose is to determine control strategies of voicing onset and offset used by speakers, and detect possible differences between female and male speakers. In addition, we intend to provide modeling support to our work on the development of speech motor control in children (e.g., Koenig, 2000; Koenig and Lucero, 2002).

For the vocal fold model, we adopt a version of the popular two-mass model of Ishizaka and Flanagan (1972). This is a simple model that has been widely adapted to yield realistic simulations of glottal source behavior (e.g., Koizumi *et al.*, 1987; Lous *et al.*, 1998; Pelorson *et al.*, 1994; de Vries *et al.*, 2002).

We will first describe the collected airflow data, followed by the laryngeal and vocal tract models and the fitting algorithm. Our results present the best fits of the model to the recorded data.

II. DATA

We collected data from eight normal English speaking subjects: four men and four women. All the subjects were between 20 and 40 years old, with mean ages of 27 and 28 years for men and women, respectively. The data consisted of oral airflow signals, recorded using a Rothenberg mask

[a])Electronic mail: lucero@mat.unb.br
[b])Electronic mail: koenig@haskins.yale.edu
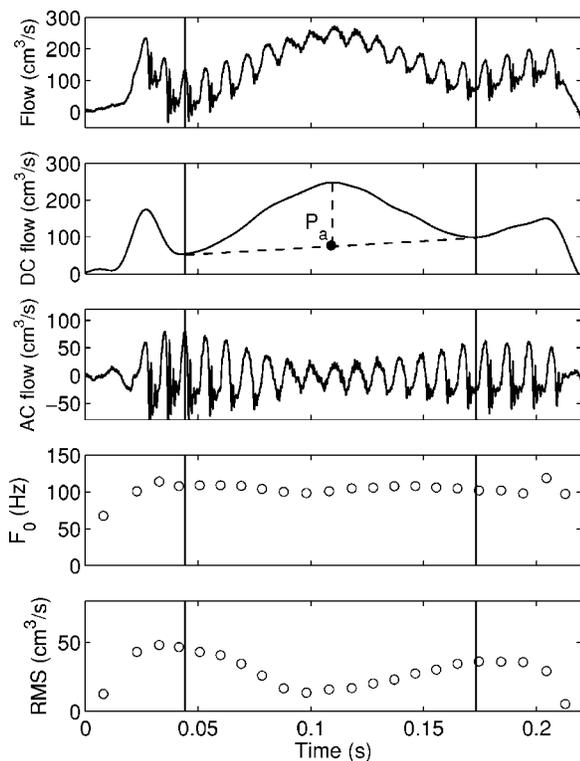
FIG. 1. Airflow data from subject MJB (male). From top to bottom: recorded airflow, dc component, ac component, fundamental frequency F_0, and rms value of ac amplitude. The vertical full lines mark the period selected as a target for the simulations. P_a defines a reference point (see the text).

FIG. 2. The airflow data from subject FJH (female). From top to bottom: recorded airflow, dc component, ac component, fundamental frequency F_0, and rms value of ac amplitude. The vertical full lines mark the period selected as a target for the simulations, and the dashed lines mark the devoiced period. P_a defines a reference point (see the text).

while the subjects repeated the utterance ''A papa hopper,'' with stress on the third syllable. The signals were filtered at 4.8 kHz, and sampled at 10 kHz. All productions were auditorily assessed to verify that they were perceptually acceptable versions of the intended utterance. We selected ten typical records from each subject, of various loudness levels, and showing both voiced and devoiced /h/.

The data was read into Matlab software, which was used for all further processing.

The dc flow component of each record was obtained by low-pass filtering it at a 50 Hz cutoff to eliminate glottal pulses, using a sixth-order Butterworth filter. The ac flow component was next computed as the difference between the unfiltered original signal and the filtered signals. The individual cycles of the ac component were then identified by using a zero-crossing algorithm with low pass filtering (Titze and Liang, 1993). From each cycle, we computed its rms value, and the fundamental frequency, as the inverse of its temporal length.

Figure 1 shows airflow data for a male subject, for the portion corresponding to the utterance /aha/.

Let us recall that airflow might be roughly considered, as directly proportional to the glottal area, when other parameters are constants. According to Titze (1988),

$$\Delta P = \frac{k_t \rho U_g^2}{2 A_g^2}, \tag{1}$$

where ΔP is the transglottal pressure, k_t is a transglottal pressure coefficient, ρ is the air density, U_g is the airflow

(volume velocity), and A_g the glottal area. Assuming also that the glottal air flow is equal to the oral flow, and that the transglottal pressure is constant, then the airflow results are proportional to the glottal area. Thus, the dc flow plot in Fig. 1 (second panel from the top) may be considered to represent the neutral glottal area, clearly showing the abduction–adduction gesture for the production of /h/.

In this case, the /h/ is fully voiced. The ac amplitude (the third panel from the top) decreases a bit during the glottal abduction–adduction, and there is almost no variation of the fundamental frequency F_0.

Figure 2 shows data corresponding to a female subject. In this case, the /h/ is devoiced. The dashed lines mark the period that we considered devoiced. In all cases, devoiced periods were identified as a region in which the computed values of F_0 were erratic (sudden jumps and discontinuities) and the rms values of ac amplitude reached a constant level close to zero (noise level), as may be seen in the F_0 and rms plots. Such regions were identified and marked by a visual inspection of the results.

We can also note a slight hysteresis effect for voice offset–onset in this example: voice stops and starts at different levels of the dc flow (169.5 and 152.2 cm^3/s, respectively). Assuming that the dc flow roughly represents the glottal neutral area, then we may say that voice offset and onset are achieved at different laryngeal configurations (different glottal areas). Further, the conditions for voice onset are more restricted than those for voice offset: to restart voice, the vocal folds must be driven closer together than the position at which voice stopped. This hysteresis effect has

TABLE I. Summary of recorded data. The table shows mean values of airflow parameters, grouped in fully voiced /h/ and devoiced /h/ for each subject. T is the length of the target region.

Subject		F_0 (Hz)	AC (cm³/s)	dc (cm³/s) base	peak	T (s)	Records
			Female subjects				
FEM	voiced	205	47.6	71.5	205.9	0.164	6
	devoiced	188	29.0	73.1	233.0	0.172	4
FGR	voiced	188	36.0	57.5	160.7	0.159	9
	devoiced	199	48.2	64.4	347.6	0.205	1
FJH	voiced	197	30.5	30.8	123.5	0.124	6
	devoiced	198	34.5	28.0	153.2	0.131	4
FRS	voiced	238	94.2	203.4	503.9	0.162	3
	devoiced	242	116.3	139.8	730.7	0.162	7
			Male subjects				
MJB	voiced	121	45.4	69.0	292.5	0.136	6
	devoiced	114	32.4	81.0	288.9	0.146	4
MJW	voiced	140	182.3	69.7	830.0	0.133	4
	devoiced	133	134.9	4.4	572.7	0.140	6
MSFA	voiced	123	61.6	71.2	342.4	0.144	8
	devoiced	102	44.0	71.9	362.4	0.162	2
MSO	voiced	149	25.5	68.2	230.8	0.158	4
	devoiced	155	25.9	71.4	277.5	0.170	6

FIG. 3. Two-mass model of the vocal folds (top) and vocal tract model for vowel /a/ (bottom).

been modeled using a subcritical Hopf bifurcation phenomenon (Lucero, 1999a), and it is common in cases of flow-induced vibrations (Thompson and Stewart, 1986).

In both figures, the vertical full lines mark the glottal abduction–adduction region, which was used later as target to fit the two-mass model. This region was defined between the two minima of the dc flow immediately before and after the /h/ peak. In the dc flow plots, P_a defines a weighted average between the two minima, and was used as a base reference value.

Table I summarizes the characteristics (mean values) of the recorded data. The dc base value is the weighted average indicated as P_a in Figs. 1 and 2, and the peak value corresponds to the peak glottal abduction. The ac rms and F_0 values were computed using a weighted average of values at the start and end of the target period, similarly to the dc base value. Main differences between male and female subjects appear in the fundamental frequency (higher in women than men, with no overlap between groups), and in the peak airflow value (higher in men than in women, with the exception of female subject FRS).

III. MODELS

A. Vocal folds

The larynx was modeled using a modified version of the two-mass model of the vocal folds (Ishizaka and Flanagan, 1972). Figure 3 (top) shows a sketch of the model. Each vocal fold is represented by two mass-damper-spring systems (m_1-b_1-s_1 and m_2-b_2-s_2), coupled through a spring (k_c). The two vocal folds are assumed identical, and they move symmetrically with respect to the glottal midline, in the horizontal direction.

When the glottis is open, the equations of motion may be written as

$$m_1\ddot{x}_1 + b_1(x_1,\dot{x}_1) + s_1(x_1) + k_c(x_1 - x_2) = f_1,$$
(2)

$$m_2\ddot{x}_2 + b_2(x_2,\dot{x}_2) + s_2(x_2) + k_c(x_2 - x_1) = f_2,$$

where b_i, s_i, f_i, with $i = 1, 2$, denote the forces related to the tissue damping, elasticity, and the airflow, respectively, m_i are the masses, x_i are their horizontal displacements measured from a rest (neutral) position $x_0 > 0$, and k_c is the coupling stiffness. As in the work of Ishizaka and Flanagan work, we adopted a cubic characteristic for the tissue elastic forces, of the form

$$s_i(x_i) = k_i x_i(1 + 100x_i^2), \quad i = 1,2,$$
(3)

where k_i are stiffness coefficients and x_i is measured in cm.

The stiffness coefficients and masses were computed through a Q scaling factor (Ishizaka and Flanagan, 1972): $k_i = Q\hat{k}_i$, $k_c = Q\hat{k}_c$, and $m_i = \hat{m}_i/Q$. This Q factor may be regarded as a scaling factor for the natural frequencies of the model, and it provides a convenient way to control the oscillation frequency. It will be used later as one of the control parameters of the model.

For the damping forces, instead of the usual linear term $r_i\dot{x}_i$, we adopted a nonlinear characteristic of the form

$$b_i(x_i,\dot{x}_i) = r_i(1 + 150|x_i|)\dot{x}_i, \quad i = 1,2,$$
(4)

where r_i are damping coefficients. The reason was the need to limit the amplitude of the vocal fold oscillation as the glottal width increased. As an illustration, Fig. 4 shows a simulation result using a linear damping characteristic, and a glottal half-width increase from 0.009 to 0.102 cm. An increase in the amplitude of the glottal pulses when the glottal abduction begins may be noted. This increase is a shortcoming of the standard two-mass model, since it is not present in the recorded airflow data. It was also noted by McGowan et al. (1995) in their modeling work. It appears because, prior to the glottal area increase, the collision between the opposite vocal folds is the major factor that limits the vocal fold oscillation. When the glottal width increases, at the beginning of the abduction gesture, the distance from the vocal fold rest position to the midline increases, and thus there is
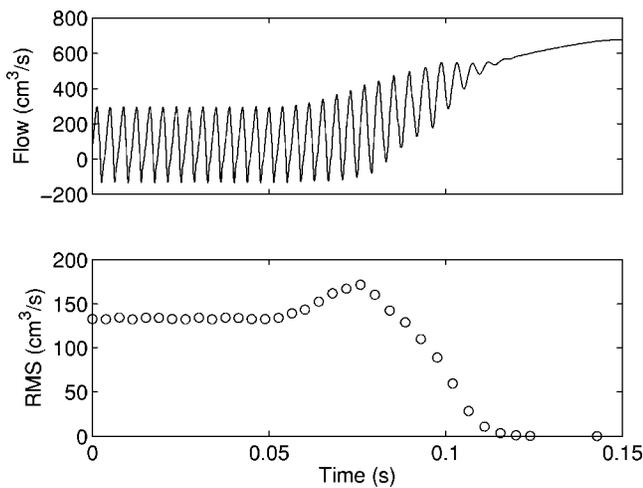
FIG. 4. Simulated airflow for vocal fold abduction, and linear vocal fold damping. Top: simulated airflow. Bottom: rms value of the ac component.
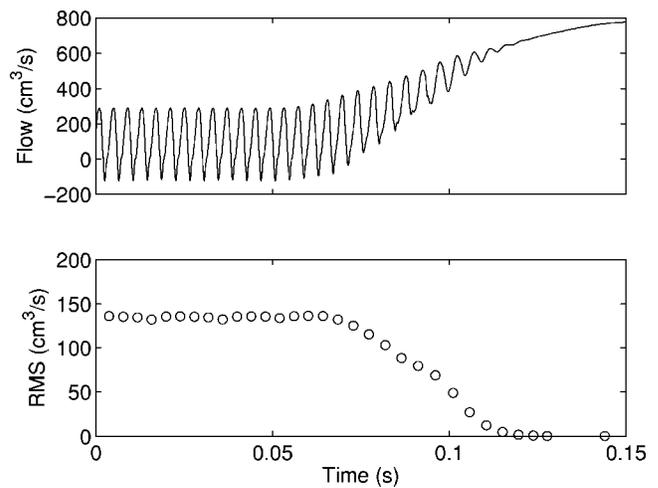
FIG. 5. Simulated airflow for vocal fold abduction, and nonlinear vocal fold damping. Top: simulated airflow. Bottom: rms value of the ac component.

more space for the oscillation. Eventually, for larger glottal widths, other factors come to limit the oscillation, such as the spring nonlinear characteristics and aerodynamic effects, and so the oscillation amplitude becomes smaller.

We found it necessary to eliminate this flow increase, to get reasonably good fits of the model to the data. McGowan *et al.* (1995) solved this issue by incorporating a representation of the anterior–posterior dimension of the vocal fold into the two-mass model. The posterior region was assigned a glottal area wider than the anterior region, and it was not allowed to vibrate. According to the authors, with that rudimentary upgrade of the two-mass model into a three-dimensional model, it was possible to abduct the vocal folds without increasing the amplitude of the airflow pulses.

Looking for alternative simpler solutions, we found that a nonlinear damping characteristic of the general form $r_i(1+\kappa|x_i|)\dot{x}_i$, where κ is a coefficient, produces the same result. The factor $r_i(1+\kappa|x_i|)$ acts as an equivalent damping coefficient, dependent on the displacement x_i. So, the larger the oscillation amplitude, the larger the damping on the system. This imposes a limit to the oscillation amplitude, by increasing the losses of the energy that fuels the oscillation, at large amplitudes. Note also that, at very low amplitudes ($x_i \rightarrow 0$) the damping factor approaches a linear characteristic. This has the advantage that the threshold pressure and other conditions to start the vocal fold oscillation, which depend on the damping term, are not affected. Another strategy to limit the oscillation amplitude could be to increase the nonlinear part of the stiffness characteristics; however, this approach would also make the oscillation frequency more dependent on the oscillation amplitude. Using the damping term, on the other hand, does not have a direct effect on the oscillation frequency, and provides a simple and convenient way to control the oscillation amplitude through parameter κ. Figure 5 shows a simulation result with the same configuration of the model as for Fig. 4, but with a nonlinear damping characteristic (with $\kappa=150$), and an increased subglottal pressure in order to get the same airflow amplitude. The value of $\kappa=150$ was selected by inspection of the simulation results.

The adoption of a nonlinear damping term for the fold tissues is also in agreement with experimental data. It is known that the vocal fold tissues, like soft tissues in general, have strong nonlinearities in their viscoelastic properties (Alipour-Haghighi and Titze, 1985; Chan and Titze, 2000; Titze, 1994). More specifically, data obtained by Alipour-Haghighi and Titze (1985, Fig. 4) show that the time constant of tissue relaxation curves increases with the level of strain imposed. This result may be modeled by a damping factor that increases with strain, as in Eq. (4). Note that a linear damping term combined with the nonlinear stiffness in Eq. (3), which is the standard version of the two-mass model, would produce the opposite effect: relaxation curves with time constants that decrease with the level of strain. Experimental measures in other soft tissues and neuromuscular systems have also been interpreted as demonstrating strain-dependent viscous damping (e.g., Bilston *et al.*, 2001; Gielen and Houk, 1984; Lin and Rymer, 2001; Nasseri *et al.*, 2002). The proper form of the damping term in the vocal fold model is beyond the objectives of this paper. Here we have adopted the form in Eq. (4), because it provided good results when fitting the model to the data. However, other alternatives are possible, which should be evaluated in future research.

The contact between the opposite vocal folds is modeled as in the work of Ishizaka and Flanagan work (1972). We assume that at their rest position, both masses are at a distance x_0 from the glottal midline. Then, each mass i collides with its opposite counterpart at a displacement $x_i = -x_0$. During contact, the stiffness is increased,

$$s_i(x_i) = k_i x_i(1+100x_i^2) + 3k_i(x_i+x_0)[1+500(x_i+x_0)^2],$$

$$x_i < -x_0 \quad \text{and} \quad i=1,2; \tag{5}$$

and the damping coefficient r_i is increased by adding 1 to the damping ratio: if $r_i = 2\xi_i\sqrt{k_i m_i}$ is the value of the damping coefficient when the glottis is open, for a given damping ratio ξ_i, then during contact of mass i it assumes the value $r_i = 2(\xi_i+1)\sqrt{k_i m_i}$.

B. Glottal and vocal tract aerodynamics

The glottal aerodynamics for the open glottis is modeled using a version of the Ishizaka and Flanagan equations (1972), updated with experimental results and a simplified version of the boundary layer model (Pelorson *et al.*, 1994, 1995; Lous *et al.*, 1998). For clarity, below we briefly describe the main equations, and leave the details for the cited references.

Letting the subglottal pressure be P_s, then the drop to pressure P_{11} at the glottal entry (at the lower edge of mass m_1) is, according to Bernoulli's equation,

$$P_s - P_{11} = \frac{\rho u_g^2}{2a_1^2}, \tag{6}$$

where ρ is the air density, u_g the volume velocity of glottal airflow, and a_1 is the cross-sectional lower glottal area, given by $a_1 = 2l_g(x_1 + x_0)$, where l_g is the vocal fold length. Note that flow contraction (*vena contracta*) is not considered, following Pelorson *et al.* (1995).

Along mass m_1, pressure drops to a value P_{12} due to air viscosity, given by

$$P_{11} - P_{12} = \frac{12\mu d_1 l_g^2 u_g}{a_1^3}, \tag{7}$$

where μ is the air viscosity, and d_1 is the width of mass m_1.

Two cases must be considered next, according to the glottal shape. Let us consider first the case in which the glottis is convergent or slightly divergent, i.e., $a_1 > k_s a_2$, where $k_s > 1$ is a suitable constant, and $a_2 = 2l_g(x_2 + x_0)$ is the cross-sectional upper glottal area. At the boundary between both masses there is a pressure variation, given by

$$P_{21} - P_{12} = \frac{\rho u_g^2}{2}\left(\frac{1}{a_1^2} - \frac{1}{a_2^2}\right), \tag{8}$$

where P_{21} is the air pressure at the lower edge of mass m_2. Next, there is a pressure drop along mass m_2 due to air viscosity, similar to Eq. (7),

$$P_{21} - P_{22} = \frac{12\mu d_2 l_g^2 u_g}{a_2^3}, \tag{9}$$

where P_{22} is the pressure at the glottal exit. At this point, due to the abrupt area expansion, the flow detaches from the glottal wall and forms a jet stream. We assume that all energy is lost in the stream due to turbulence (Pelorson *et al.*, 1994), and so

$$P_{22} = P_0, \tag{10}$$

where P_0 is the pressure input to the vocal tract.

When the glottis is more divergent, with $a_1 \leq k_s a_2$, we assume that the point of flow separation from the glottal wall moves inside the glottis and occurs at the boundary between both masses. This is a gross simplification of the boundary layer model, which shows that the point of flow separation moves upstream the glottis as a continuous function of the glottal angle of divergence. According to the Pelorson *et al.* (1994, 1995) results, in a divergent glottis the ratio between

the glottal area at the point of airflow separation and the minimum glottal area becomes asymptotically constant at high Reynolds numbers. In measurements on a physical model of the larynx with a cylindrical profile for the vocal folds, they obtained the approximate ratio of 1.1. In a previous work (Lucero, 1999a), the flow separation equations were solved, assuming a linear variation of the glottal area along the glottis, and an approximate ratio of 1.3 was obtained. Here, we consider an intermediate value and thus assume $k_s = 1.2$, following also Lous *et al.* (1998).

In this case, again we assume that all airflow energy is lost due to turbulence from the point of detachment, and so

$$P_{21} = P_{22} = P_0. \tag{11}$$

Finally, the forces f_i acting on the masses are computed as $f_i = (P_{i1} + P_{i2})/2$.

The forces for the case of glottal closure are computed as follows (Ishizaka and Flanagan, 1972):

$$f_1 = d_1 l_g P_s, \quad x_1 \leq -x_0 \quad \text{or} \quad x_2 \leq -x_0; \tag{12}$$

$$f_2 = \begin{cases} d_2 l_g P_s, & \text{if } x_1 > -x_0, \quad x_2 \leq -x_0; \\ 0, & \text{if } x_1 \leq -x_0; \end{cases} \tag{13}$$

The vocal tract was represented by a standard two-tube configuration for vowel /a/ (Flanagan, 1972; Titze, 1994), shown in Fig. 3. Its equations were derived using a transmission line analogy, terminated in a radiation load of a circular piston in an infinite baffle. The elements of the transmission line were computed from the cross-sectional areas S_1, S_2 and lengths L_1, L_2 of the vocal tract tubes, using the standard equations of the analogy (Ishizaka and Flanagan, 1972; Flanagan, 1972).

We asumed that the shape of the vocal tract did not change over the utterance /aha/. This is something of an oversimplification: Both formants F1 and F2 were significantly higher across speakers in the stressed vowel compared to the unstressed vowel. Nevertheless, the mean formant differences as a function of stress were small (about 60 Hz for the women, and 40 Hz for the men). The /a/-like quality of the unstressed vowel in this utterance probably reflects coarticulation with the preceding and following /a/ vowels. There was also a general tendency across speakers for F1 and F2 to be highest in the loud condition, and lowest in the soft condition; this pattern was significant for F1 in the stressed vowel. Again, however, the overall formant differences were small (less than 75 Hz). Finally, in those cases where the formants were visible throughout the production of /h/, F1 and F2 generally showed little movement between the vowels. Thus, we are confident that variability in supraglottal postures as functions of stress and loudness was fairly minor in the modeled data. Moreover, we note that our objective was to fit the major variations in speakers' aerodynamic data, rather than details of token-to-token variability.

C. Male and female configurations

The following values were adopted for the male configuration of the vocal fold model (Flanagan, 1972; Ishizaka and Flanagan, 1972): $\hat{m}_{1M} = 0.125$ g, $\hat{m}_{2M} = 0.025$ g, $\hat{k}_{cM} = 25$ N/m, $\hat{k}_{1M} = 80$ N/m, $\hat{k}_{2M} = 8$ N/m, $\xi_{1M} = 0.1$, ξ_{2M}

=0.6, $l_{gM} = 1.4$ cm, $d_{1M} = 0.25$ cm, $d_{2M} = 0.05$ cm, and for the vocal tract model (Flanagan, 1972; Goldstein, 1980): $S_{1M} = 1$ cm^2, $S_{2M} = 7$ cm^2, $L_{1M} = 8.9$ cm, $L_{2M} = 8.1$ cm. Here, subscript M denotes a male configuration.

To simulate female voices, McGowan *et al.* (1995) used the same male configuration of the model, as they were mainly interested in reproducing the gross dc airflow. Here, we intend to take a closer look at how onset and offset of the vocal fold vibration is controlled, and thus the size of the larynx and vocal tract might be an important factor. Female airways are smaller in cross-sectional area, and thus the pressure losses for viscosity are larger. The energy of the flow available for the oscillation is then smaller, and this fact may affect its control. Hence, we reduced the dimensions of the larynx and vocal tract to simulate female voices using scaling factors. According to Titze (1989), the size relation between male and female larynges is in the range 1.2–1.6, depending on the dimension. We adopted then a single scaling factor for the laryngeal dimensions, with a value $\beta_l = 1.4$. Masses were accordingly computed dividing by β_l^3, to compensate for the volume reduction. For the tissue stiffness, we assumed a constant elasticity modulus for men and women. This assumption is again a simplification, since Titze (1989) reported a slightly stiffer tissue for females than for males, probably as a result of differences in tissue composition. Then, for a constant elasticity modulus, the stiffness coefficient is directly proportional to the cross-sectional area of the tissues, and inversely proportional to their length. The reduction of all dimensions by a factor β_l implies that stiffness is also reduced by this same factor. We assumed the same tissue damping ratios for men and women.

Denoting by subscript F the female configuration, the parameters of the vocal fold model are: $\hat{m}_{1F} = \hat{m}_{1M}/\beta_l^3 = 0.0456$ g, $\hat{m}_{2F} = \hat{m}_{2M}/\beta_l^3 = 0.0091$ g, $\hat{k}_{cF} = \hat{k}_{cM}/\beta_l = 17.85$ N/m, $\hat{k}_{1F} = \hat{k}_{1M}/\beta_l = 57.14$ N/m, $\hat{k}_{2F} = \hat{k}_{2M}/\beta_l = 5.71$ N/m, $\xi_{1F} = \xi_{1M} = 0.1$, $\xi_{2F} = \xi_{2M} = 0.6$, $l_{gF} = l_{gM}/\beta_l = 1$ cm, $d_{1F} = d_{1M}/\beta_l = 0.179$ cm, $d_{2F} = d_{2M}/\beta_l = 0.036$ cm.

For the female vocal tract we used data from Goldstein (1980): $L_{1F} = 6.3$ cm, $L_{2F} = 7.8$ cm. The cross-sectional areas were computed by scaling the male areas according to the total length relation of male to the female vocal tracts $\beta_v = (8.9 + 8.1)/(6.3 + 7.8) = 1.21$. Thus, $S_{1F} = S_{1M}/\beta_v^2 = 0.688$ cm^2, $S_{2F} = S_{2M}/\beta_v^2 = 4.816$ cm^2.

IV. FIT TO COLLECTED SIGNALS

A. Control and target parameters

As control parameters for fitting the model to the data, we selected the subglottal pressure P_s, glottal half-width at the rest position of the vocal folds x_0, and Q factor. The glottal half-width characterizes the degree of abduction–adduction of the vocal folds, and should be the major control parameter for the flow increase during the production of /h/. The Q factor represents the degree of tension of the vocal folds, and is the main control for the fundamental frequency F_0 of the vocal fold oscillation. Subglottal pressure is a main control for the amplitude of the oscillation, and thus for the amplitude of the airflow ac component.

Vocal fold abduction and tension should be the main factors used by speakers to control phonation when producing /aha/ (Koenig, 2000; Löfqvist *et al.*, 1989; McGowan *et al.*, 1995). The value of subglottal pressure, on the other hand, should derive from the lung pressure, which should be roughly constant during the short duration of the utterance, as should the relative value between the resistances of the airways upstream and downstream the glottal entrance. As the vocal folds abduct for /h/ and glottal resistance decreases, subglottal pressure may decrease, and past experimental work has reported short term decreases in P_s when both the glottis and the supraglottal tract are open (Löfqvist, 1975; Ohala and Ohala, 1972). In the current work, we have chosen to use the subglottal pressure as a control parameter rather than explicitly controlling it during the abduction because the existant data do not permit precise quantification of how large its decrease should be: Löfqvist's (1975) measurements were made on a single male speaker of Swedish, producing voiceless obstruents, whereas Ohala and Ohala (1972) reported data on a single female speaker of Hindi. Our own past measurements (Löfqvist *et al.*, 1995; Koenig, 2000) further indicate that speakers may vary considerably in the degree of abduction they use during the consonant /h/ (cf. also Manuel and Stevens, 1989). Moreover, the well-known dimensional differences between adult male and female larynges (Titze, 1989; Kahane, 1982) imply that glottal areas during abduction may be significantly larger in men than women. Rather than speculate on the extent of group differences and/or intragroup variation, we approached this problem by allowing the subglottal pressure to vary as a fitting parameter.

As targets in the data, we chose the dc flow amplitude, and the rms and fundamental frequency of the ac component of the flow.

It is instructive to analyze briefly how the targets depend on the control parameters. Let us consider the flow dc component, ac component (rms value), and fundamental frequency F_0 as functions of the control parameters, in the form $(dc, ac, F_0) = \Gamma(P_s, x_0, Q)$, where Γ denotes a vector-valued function. We next compute their values when varying one of the control parameters, while keeping the other two fixed at some reference values. As reference, we used the standard values of the two-mass model $\hat{P}_s = 800$ Pa, $\hat{x}_0 = 0.02$ cm, $\hat{Q} = 1$ (Ishizaka and Flanagan, 1972).

Figure 6 shows results for the male configuration. The upper plot shows that both the ac and dc components increase almost linearly with the subglottal pressure. The fundamental frequency is almost constant, with a small negative slope. There is a threshold value for a subglottal pressure of 188 Pa (the threshold values were computed through additional simulations at a higher precision of the control parameters), which must be overcome for the start of the vocal fold oscillation. The middle plot shows that, when the glottal half-width is increased by abducting the vocal folds, the dc flow increases, since the glottal resistance decreases due to the larger glottal cross-sectional area. However, the vocal fold oscillation becomes more difficult (e.g., see Titze, 1988), so the ac amplitude decreases. There is a threshold value of 0.14 cm, above which the oscillation is no longer
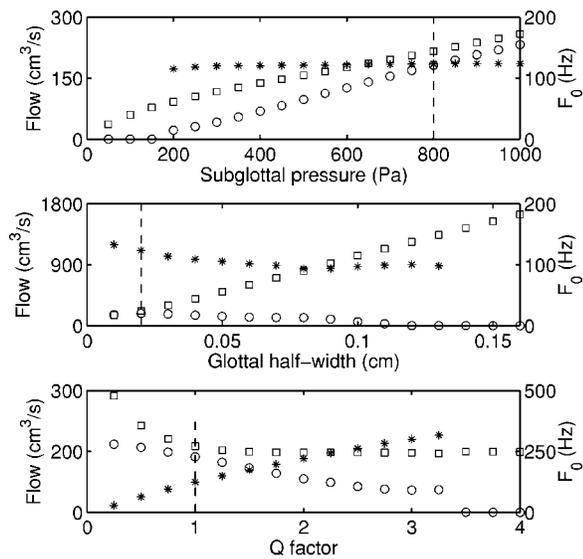
FIG. 6. Output targets as a function of the control parameters, for a male configuration of the model. Squares: dc component of oral airflow (scale at the left). Circles: rms value of the ac component of oral airflow (scale at the left). Asterisks: fundamental frequency of the ac component (the scale at the right). In the regions without asterisks, the AC component has a zero amplitude. The vertical dashed line marks the standard values of the model.

FIG. 7. Output targets as a function of the control parameters, for a female configuration of the model. Squares: dc component of oral airflow (scale at the left). Circles: rms value of the ac component of oral airflow (scale at the left). Asterisks: fundamental frequency of the ac component (scale at the right). In the regions without asterisks, the ac component has a zero amplitude. The vertical dashed line marks the standard values of the model.

possible. Considering, finally, variations of the Q factor, in the bottom plot, we see that its increase causes a linear increase of fundamental frequency. Let us recall that the Q factor is a linear scaling factor for the natural frequencies of the vocal folds. It also affects the dc and ac flows, causing them to decrease. As the Q factor increases, the vocal fold stiffness also increases, and so the oscillation is inhibited. Here too, there is a threshold value of 3.4, which sets an upper limit for the oscillation.

Figure 7 shows results for the female configuration. Results are qualitatively similar to the male case. The airflow values are lower than the male values, probably because of the increased resistance of the narrower airways. Frequency values are higher; although the stiffness coefficients are smaller than the male values by a factor of $1/\beta_l$, the masses are smaller by a factor of $1/\beta_l^3$, which results in higher natural frequencies. In fact, the fundamental frequency relations between the female and male cases for the same values of the control parameters are approximately equal to $\beta_l = 1.4$. The threshold value of the subglottal pressure is 240 Pa, larger than the male value. The difference is relatively small, 27% higher in women. The results seem in agreement with measures in men vs. women, which have found almost no gender differences in values of subglottal pressure (Koenig, 2000). Measured airflow values, on the other hand, are, in general, smaller in women, as found in the above simulations (for both dc and ac components).

The threshold values for the glottal half-width and Q factor are smaller than the male values, with values of 0.08 cm and 2.6, respectively. These results, together with the larger threshold pressure, imply that the adult female larynx has a more restricted region of vocal fold oscillation. Thus, voice would be harder to sustain in women than in men. In particular, the smaller threshold for the glottal width (43% reduction in women compared to men) would explain the
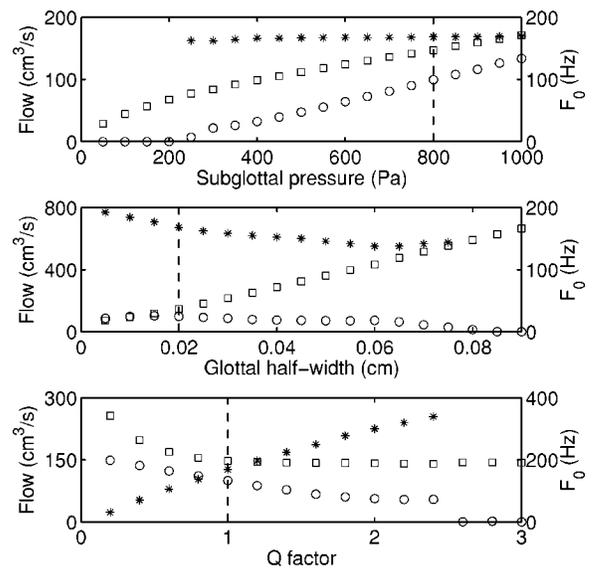
higher incidence of devoicing during glottal abduction–adduction for /h/ in running speech in women as compared to men (Koenig, 2000). As the vocal folds abduct, they easily reach the oscillation offset threshold in women, whereas men would require more extreme degrees of abduction.

B. Algorithms

Fitting the model to the selected target parameters poses the difficulty that the dc flow is computed at each data point (time step) of the simulation results. The rms value and fundamental frequency, on the other hand, require a certain quantity of data points, since they are computed after each whole cycle. Thus, a dynamic algorithm that tracks the target signal and tunes the control parameters to get the correct output point by point (e.g., as in Pitermann and Munhall, 2001) would not work. Instead, the algorithm must track the target signals at larger steps, each of which must contain at least one ac cycle. However, the length of the ac cycles, and even the presence or absence of an ac component, depend on the same values of the control parameters that the algorithm must determine.

After trying several strategies, we found that the following one produced good results for all records. The target parameters were determined from the measured records at five instants of time: the two minima of the dc flow immediately before and after the /h/ peak (vertical solid lines in Figs. 1 and 2), the maximum at the /h/ peak, and the two midpoints between the peak and the two minima.

The algorithm worked then in two main stages. In the first stage, each set of target parameters for the five time points was considered separately. For each set, simulations of oral airflow were produced using constant control parameters. The dc component, and rms and fundamental frequency of the ac component were computed from the simu-
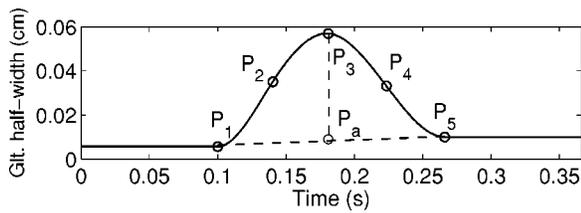
FIG. 8. Control time series for the glottal half-width. Points P_i are determined by the optimization algorithm described in the text. Before P_1 and after P_5, the parameter has a constant value, for a period of 0.1 ms. Between P_1 and P_5, intermediate points are determined using cubic spline interpolation. P_a is a weighted average of P_1 and P_5, used as a reference base value.

lated results and compared with the targets. In the devoiced regions (no ac component), fundamental frequency was not considered. As a cost function, a weighted sum of the square difference between these simulated parameters and the measured targets was computed,

$$C = \alpha_1 (\mathrm{dc}^* - \mathrm{dc})^2 + \alpha_2 (\mathrm{ac}^* - \mathrm{ac})^2 + \alpha_3 (F_0^* - F_0)^2, \tag{14}$$

where the asterisk denotes the target values. The weights were given the values $\alpha_1 = 1/\overline{\mathrm{dc}}$, $\alpha_2 = 1/\overline{\mathrm{ac}}$, $\alpha_3 = 1/\overline{F}_0$, where the overbar denotes mean values over the target region. In addition, upper and lower limits were imposed on the control parameters, to avoid regions of unrealistic values. Optimal values of the control parameters were then determined using a Nelder–Mead simplex method (Kincaid and Cheney, 2002) implemented in MATLAB, to minimize the cost function. These values were used in the next stage as initial values for a refined optimization.

In the second stage, a time series for each control parameter was constructed, as illustrated in Fig. 8 for the case of the glottal half-width. There, points P_1 to P_5 represent the five points computed in the previous stage. The control time series were used to generate simulations of oral airflow, and the optimal values of the three control parameters at points P_2 to P_5 were recomputed sequentially. Values at P_1 were left as computed in the previous stage, because before this point the control parameters have constant values, as in the previous stage. After this point, they must be recomputed, because now the parameters change with time (according to the control time series). These computations were performed using the same simplex method and cost function as in the initial stage, and using the values determined in that stage as initial values. The control time series were updated each time the values of the control parameters were changed, and new simulations were generated using the updated time series.

The capability of this algorithm for finding correct control parameter values was verified using oral airflow records generated by the same models, in both male and female configurations. Using the default termination tolerances provided in the MATLAB scripts, the algorithm computed the control parameters with errors less than 0.5%, relative to their standard values.

V. RESULTS

Figures 9 and 10 show simulation results for the data in Figs. 1 and 2, respectively. The vertical solid lines mark the
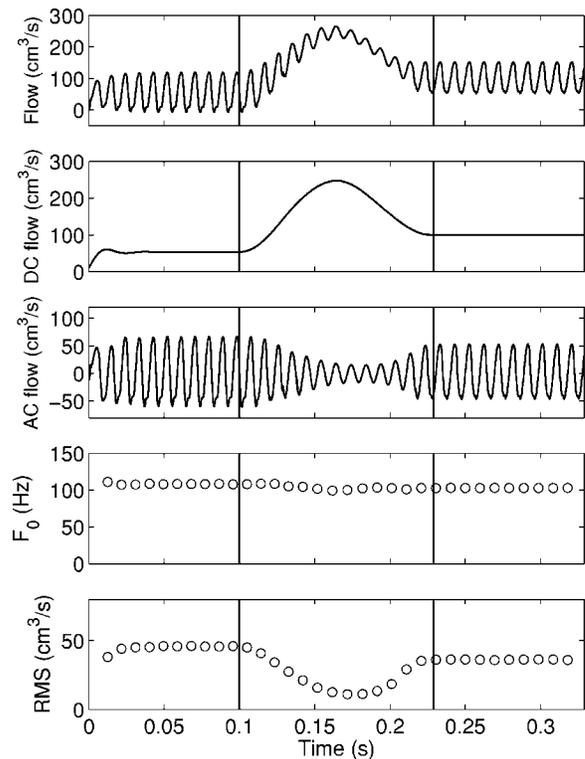
FIG. 9. Simulated airflow for data in Fig. 1. From top to bottom: airflow, dc component, ac component, fundamental frequency F_0, and the rms value of ac amplitude. The vertical lines mark the period that correspond to the simulation target.

target region, as in Figs. 1 and 2. In the case of fully voiced /h/, as in Fig. 9, a good fit of the model was possible and the simulated signals reproduced well the targeted data. In the devoiced /h/, as in Fig. 10, there is some discrepancy in the restart of the oscillation. Comparing with Fig. 2, we may note that the ac amplitude grows slowly after the /h/ peak, causing a delay in the rebuilding of the ac component. As a consequence, the offset–onset hysteresis gap is even larger in the simulated results. This delay was typical of the simulations in the devoiced /h/ cases, and was also noted by McGowan et al. (1995) in their simulations. In their work, the restart of the oscillation was facilitated by setting an overshoot for the glottal width at the end of the vocal fold adduction. Here, none of the results for that parameter showed such an overshoot. The delay seems to be a shortcoming of the vocal fold model used, which does not seem capable of a quicker restart of the vocal fold oscillation. Let us also note that the onset of the oscillation depends not only on the values of the control parameters, but also on the velocity of variation of those parameters (Lucero, 1999b): The smoother the variation, the smaller the gap between offset and onset thresholds. In the present case, the simulation results suggest that the variation at the abduction–adduction gesture is somewhat drastic for the model, which results in a large hysteresis effect.

Figure 11 shows the computed control parameters, for all the records of a male subject (the same subject as in Figs. 1 and 9). All other subjects presented qualitatively similar results, and so are not shown here. In general, all three parameters move in the direction of suppressing oscillation.
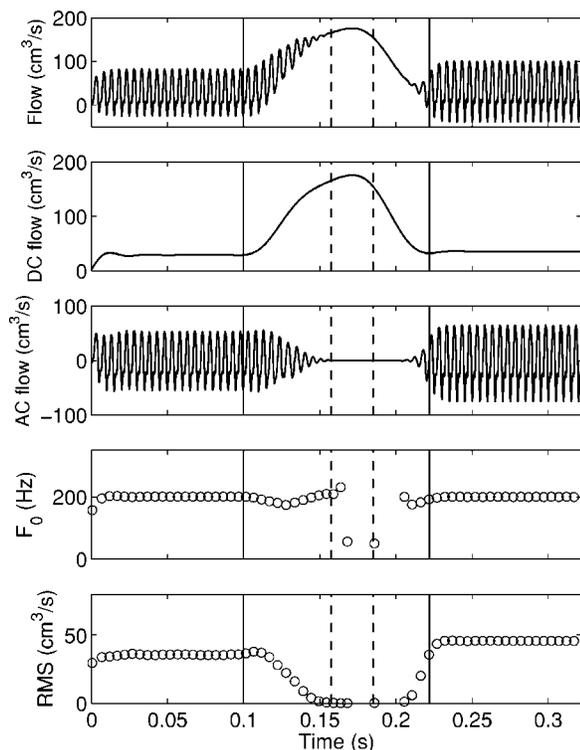
FIG. 10. Simulated airflow for data in Fig. 2. From top to bottom: airflow, dc component, ac component, fundamental frequency F_0, and rms value of ac amplitude. The vertical lines mark the period that correspond to the simulation target. The vertical dashed lines mark the devoiced period in Fig. 2.

The glottal half-width x_0 grows and then decreases, following a glottal abduction–adduction curve. At the same time, the subglottal pressure varies in the inverse direction. As discussed above (Sec. IV A), this reduction of glottal pressure at glottal abduction has been observed experimentally (Löfqvist, 1975; Ohala and Ohala, 1972), and is probably caused by a reduction of glottal resistance during abduction. The correlation of the increase in glottal half-width, computed as the difference between the peak abduction value and the base value, and the decrease of subglottal pressure, com-
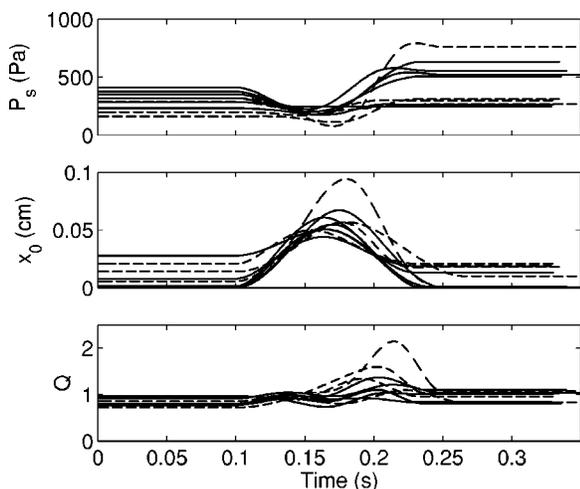
FIG. 11. Optimal control parameters for subject MJB (male). Top: subglottal pressure P_s. Middle: glottal half-width x_0. Bottom: Q factor. The full curves correspond to cases of fully voiced /h/, and the dashed curves to cases of devoiced /h/.

TABLE II. Summary of fitting results. The table shows mean values of airflow parameters, grouped in fully voiced /h/ and devoiced /h/ for each subject, computed from the simulated airflow as in Table I.

Subject		F_0 (Hz)	AC (cm^3/s)	DC (cm^3/s) base	DC (cm^3/s) peak
		Female subjects			
FEM	voiced	207	42.5	77.8	203.5
	devoiced	194	26.5	74.4	232.5
FGR	voiced	189	35.8	57.5	159.4
	devoiced	200	40.5	64.2	346.9
FJH	voiced	200	26.8	33.4	123.1
	devoiced	200	32.7	30.7	153.3
FRS	voiced	236	96.3	203.0	493.9
	devoiced	249	111.6	151.9	726.2
		Male subjects			
MJB	voiced	120	50.7	64.8	291.5
	devoiced	115	34.4	77.0	288.0
MJW	voiced	144	148.2	150.8	824.9
	devoiced	143	102.4	89.7	571.9
MSFA	voiced	123	61.1	72.4	341.5
	devoiced	103	43.8	71.7	362.1
MSO	voiced	149	19.8	68.4	230.8
	devoiced	172.6	14.9	71.3	277.3

puted analogously, is 0.901 for this subject ($p = 0.0038$), and 0.800 ($p < \text{Eps} = 2.2 \times 10^{-16}$) for all subjects. There is also an increase of the Q factor. An increase in the Q factor may correspond both to active (muscularly controlled) changes in vocal-fold characteristics and to passive changes in the vocal-fold tissues, e.g., as they are stretched during abduction. Although we cannot separate these two effects based on the current data, the higher Q factor is at least consistent with past studies, which have argued that the cricothyroid muscle may be activated around abduction movements to help suppress the vocal fold oscillation (Löfqvist *et al.*, 1989; McGowan *et al.*, 1995). In the cases of devoicing (broken lines), a stronger movement of parameters may be seen so that the threshold level is reached (smaller subglottal pressure, larger glottal half-width, and/or larger Q value).

Table II summarizes fitting results for F_0, ac rms, and dc flow, for a comparison with the respective values in the recorded data shown in Table I. In general, a good fit could be achieved for F_0 and peak dc flow for all subjects. The fit was a bit poorer for ac rms flow and base dc flow. The worst case is the simulation for male subject MJW, which, according to Table I, combines large ac flow with comparatively low dc values at the base level. The large error of the results in Table II reveals a difficulty of the model in producing this particular combination of ac and dc flow.

Table III shows the computed control parameters. In general, the estimated subglottal pressures were higher for female subjects than for male subjects. The only exception is again the simulation for subject MJW, which has high values of subglottal pressures. The combination of large ac flow values with low dc values at the base level demands an almost zero glottal width at the base level, as seen in Table III, to obtain a small dc flow, combined with a large pressure, to obtain a large ac flow. Female subject FRS also has large subglottal pressures compared to the others. They are also a result of large flow values, as seen in Table I. Note that in

TABLE III. Summary of computed control parameters. The table shows mean values, grouped in fully voiced /h/ and devoiced /h/ for each subject.

Subject		P_s (Pa)		x_0 (cm)		Q	
		base	peak	base	peak	base	peak
Female subjects							
FEM	voiced	593	450	0.010	0.037	1.15	1.39
	devoiced	429	294	0.013	0.053	1.11	1.93
FGR	voiced	432	374	0.010	0.031	1.03	1.12
	devoiced	499	416	0.010	0.065	1.13	1.09
FJH	voiced	419	291	0.004	0.028	1.05	1.16
	devoiced	491	282	0.002	0.035	1.02	0.99
FRS	voiced	1227	716	0.020	0.071	1.50	1.91
	devoiced	1239	608	0.013	0.120	1.50	1.86
Male subjects							
MJB	voiced	393	228	0.005	0.054	0.89	0.89
	devoiced	337	177	0.011	0.062	0.91	1.15
MJW	voiced	1123	623	0.000	0.091	1.10	0.95
	devoiced	774	393	0.000	0.079	1.03	1.13
MSFA	voiced	423	342	0.006	0.051	0.91	0.88
	devoiced	273	211	0.012	0.069	0.78	0.64
MSO	voiced	287	249	0.011	0.041	1.20	1.13
	devoiced	355	221	0.010	0.054	1.39	1.44

FIG. 12. Plots of mean values of control parameters for each subject, taken from Table III. The lines start at the average between starting and ending values of the parameters (point P_a in Fig. 8), and end at the values at the peak glottal abduction (point P_3 in Fig. 8). The circles indicate peak values for devoiced /h/, and the squares indicate peak values for voiced /h/. The top two plots are different views of the female results, and the bottom plots are different views of the male results.

this case the ac and dc (base) flow values are both large. Thus, the glottal width is also large, to reproduce that combination of flow values. Comparing peak glottal widths, men have, in general, larger values than women, with the exception of FRS, showing a larger vocal fold abduction. Q values are a bit higher in women, which, combined with a higher natural frequency, results in the greater fundamental frequency values for the female data in Table I.

Figure 12 show variations of the control parameters, from their values in Table III. The lines start at their base values and end at the peak abduction values. In general, the trajectories start at the regions of small x_0, high P_s, and small Q, and move toward high x_0, small P_s, and high Q. Trajectories that correspond to devoiced cases are longer than the voiced cases.

VI. CONCLUSION

In this paper we have shown a technique to fit a model of the larynx and vocal tract to speech data, which allows us to infer how the larynx is dynamically controlled in running speech. As noted in the Introduction, this is a more demanding test for the model, and might reveal aspects not seen in simulations using static configurations. In fact, it has shown the need for a nonlinear damping characteristic, to reproduce the ac flow pattern at glottal abduction.

The results show that the two-mass model plus two-tube approximation of the vocal tract, in spite of their simplicity, are generally capable of reproducing the speech data, with control parameters of physiologically realistic values and dynamical behavior. One aspect that could not be fitted with a good approximation was the restart of the oscillation after its offset, where the model had a slow behavior. The slower oscillation onset resulted in longer unvoiced periods and a larger hysteresis effect than observed in the data. Also, the model failed to reproduce airflow records combining large ac flow with low dc values.

The results also show that in both male and female speakers, devoicing at the abduction–adduction gesture for /h/ is achieved by the combined action of the three parameters: vocal fold abduction, decrease of subglottal pressure (which is an aerodynamic consequence of the glottal width increase), and an increase of vocal fold tension. Each of these actions has the effect of inhibiting the vocal fold oscillation, and even suppressing it when reaching an offset threshold. The simulation results also suggest variability across speakers in the factors that are manipulated during /h/; for example, some speakers show substantial variation in the Q factor during /h/, whereas others show a minimal change. This is consistent with our recent measurements, which indicate that individual speakers have unique strategies for achieving voicing and devoicing in running speech (Koenig and Lucero, 2004). Women have, in general, more restricted conditions for the vocal fold oscillation, which would explain the larger occurrence of devoicing in glottal abduction–adduction gestures, compared to men (Koenig, 2000). This restriction might be consequence of a smaller medial surface of the vocal folds, where energy is transferred from the flow to the fold tissues (Lucero, 2004). It also might be related to larger losses for air viscosity in the glottal airway, due to the smaller cross-sectional area, which would leave less energy to fuel the oscillation.

In general, fitting models to speech data is an important technique for making quantitative predictions of speech biomechanics, and has been used in previous studies (Döllinger _et al._, 2002; Mergell _et al._, 2000; Tao _et al._, 2004; Trevisan _et al._, 2001). Our technique may be extended to other instances of running speech to study voicing control under

different conditions. Such studies should also be accompanied by a deeper analysis of the stability and convergence of the numerical algorithm, in order to determine is the range of application to more general speech signals. Since the fitting algorithm uses ac rms and F_0 of individual ac cycles as target parameters, it may be inappropriate for cases of disordered voices that have extreme degrees of cycle-to-cycle variability. In such cases, algorithms with other target parameters should be used (e.g., Mergell *et al.*, 2000).

Even if we limit ourselves to speakers with normal voices, this work has an applicability to other research questions. Oral airflow signals are relatively easy to obtain, and they provide a noninvasive means of assessing laryngeal behavior. This is particularly important when studying special populations such as children. Modeling can refine our estimates of parameters that are difficult to measure in living speakers and tissues. Two examples are the subglottal pressure decrease during abducted conditions and the tissue damping. In future work, we plan to extend these methods to investigate the development of the larynx and phonatory control.

ACKNOWLEDGMENTS

Alipour-Haghighi, F., and Titze, I. R. (**1985**). "Viscoelastic modeling of canine vocalis muscle in relaxation," J. Acoust. Soc. Am. **78**, 1939–1943.

Bilston, L. E., Liu, Z. Z., and Phan-Thien, N. (**2001**). "Large strain behavior of brain tissue in shear: Some experimental data and differential constitutive model," Biorheology **38**, 335–345.

Chan, R. W., and Titze, I. R. (**2000**). "Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling," J. Acoust. Soc. Am. **107**, 565–580.

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

Döllinger, M., Hoppe, U., Hettlich, F., Lohscheller, J., Schuberth, S., and Eysholdt, U. (**2002**). "Vibration parameter extraction from endoscopic image series of the vocal folds," IEEE Trans. Biomed. Eng. **49**, 773–781.

Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception* (Springer-Verlag, New York).

Gielen, C. C. A. M., and Houk, J. C. (**1984**). "Nonlinear viscosity of human wrist," J. Neurophysiol. **52**, 553–569.

Goldstein, U. (**1980**). "An articulatory model for the vocal tracts of growing children," Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal folds," Bell Syst. Tech. J. **51**, 1233–1268.

Kahane, J. C. (**1982**). "Growth of the human prepubertal and pubertal larynx," J. Speech Hear. Res. **25**, 446–455.

Kincaid, D., and Cheney, W. (**2002**). *Numerical Analysis: Mathematics of Scientific Computing*, 3rd ed. (Brooks/Cole, Pacific Grove, CA), pp. 722–723.

Koenig, L. L. (**2000**). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," J. Speech Lang. Hear. Res. **43**, 1211–1228.

Koenig, L. L., and Lucero, J. C. (**2002**). "Oral–laryngeal control patterns for fricatives in 5-year olds and adults," *Proceedings of the 7th International Conference on Spoken Language Processing*, pp. 49–52.

Koenig, L. L., and Lucero, J. C. (**2004**). "Measurements of voicing offset and onset in men and women: further data," *Abstracts of the 4th International Conference on Voice Physiology and Biomechanics*, p. 93.

Koizumi, T., Taniguchi, S., and Hiromitsu, S. (**1987**). "Two-mass models of the vocal cords for natural sounding voice synthesis," J. Acoust. Soc. Am. **82**, 1179–1192.

Lin, D. C., and Rymer, W. Z. (**2001**). "Damping actions of the neuromuscular system with inertial loads: Human flexor pollicis longus muscle," J. Neurophysiol. **85**, 1059–1066.

Löfqvist, A. (**1975**). "A study of subglottal pressure during the production of Swedish stops," J. Phonetics **3**, 175–189.

Löfqvist, A., Baer, T., McGarr, N. S., and Seider Story, R. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.

Löfqvist, A., Koenig, L. L., and McGowan, R. S. (**1995**). "Vocal tract aerodynamics in /aCa/ utterances: Measurements," Speech Commun. **16**, 49–66.

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," Acta Acust. (Beijing) **84**, 1135–1150.

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

Lucero, J. C. (**1999b**). "Bifurcations at voice onset–offset," J. Acoust. Soc. Am. **105**, 1161.

Lucero, J. C. (**2004**). "Dynamics of a two-mass model of the vocal folds for men, women, and children," *Abstracts of the 4th International Conference on Voice Physiology and Biomechanics*, pp. 191–195.

Manuel, S. Y., and Stevens, K. N. (**1989**). "Acoustic properties of /h/," J. Acoust. Soc. Am. **86**, S49.

McGowan, R. S., Koenig, L. L., and Löfqvist, A. (**1995**). "Vocal tract aerodynamics in /aCa/ utterances: Simulations," Speech Commun. **16**, 67–88.

Mergell, P., Herzel, H., and Titze, I. R. (**2000**). "Irregular vocal fold vibration—High-speed observation and modeling," J. Acoust. Soc. Am. **108**, 2996–3002.

Nasseri, S., Bilston, L. E., and Phan-Thien, N. (**2002**). "Viscoelastic properties of pig kidney in shear, experimental results and modeling," Rheol. Acta **41**, 180–192.

Ohala, M., and Ohala, J. (**1972**). "The problem of aspiration in Hindi phonetics," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo, Vol. 6, pp. 39–46.

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

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Bailliet, H. (**1995**). "Description of the flow through in-vitro models of the glottis during phonation," Acta Acust. (Beijing) **3**, 191–202.

Pitermann, M., and Munhall, K. G. (**2001**). "An inverse dynamics approach to face animation," J. Acoust. Soc. Am. **110**, 1570–1580.

Tao, C., Zhang, Y., Du, G., and Jiang, J. J. (**2004**). "Estimating model parameters by chaos synchronization," Phys. Rev. E **69**, 036204.

Thompson, J. M. T., and Stewart, H. B. (**1986**). *Nonlinear Dynamics and Chaos* (Wiley, New York), pp. 108–131.

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

Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.

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

Titze, I. R., and Liang, H. (**1993**). "Comparison of F0 extraction methods for high-precision voice perturbation measurements," J. Speech Hear. Res. **36**, 1120–1133.

Trevisan, M. A., Eguia, M. C., and Mindlin, G. B. (**2001**). "Nonlinear aspects of analysis and synthesis of speech time series data," Phys. Rev. E **63**, 026216.