

Phonation thresholds as a function of laryngeal size in a two-mass model of the vocal folds (L)

Jorge C. Lucero^{a)}

Department of Mathematics, University of Brasília, Brasília 70910-900, Brazil

Laura L. Koenig^{b)}

Haskins Laboratories, 300 George Street, New Haven, Connecticut 06511 and Long Island University, Brooklyn, New York 11201-8423

(Received 4 April 2005; revised 11 July 2005; accepted 29 August 2005)

This letter analyzes the oscillation onset-offset conditions of the vocal folds as a function of laryngeal size. A version of the two-mass model of the vocal folds is used, coupled to a two-tube approximation of the vocal tract in configuration for the vowel /a/. The standard male configurations of the laryngeal and vocal tract models are used as reference, and their dimensions are scaled using a single factor. Simulations of the vocal fold oscillation and oral output are produced for varying values of the scaling factor. The results show that the oscillation threshold conditions become more restricted for smaller laryngeal sizes, such as those appropriate for females and children. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2074987]

PACS number(s): 43.70.Aj, 43.70.Bk [DOS]

Pages: 2798–2801

I. INTRODUCTION

The purpose of this letter is to analyze how phonation onset-offset conditions change according to laryngeal size, as in cases of phonation by men, women, and children. Since the biomechanical parameters of the vocal folds, the interaction between the vocal folds and airflow, and other terms of glottal aerodynamics depend on the anatomical dimensions of the glottis, variations of the oscillatory behavior of the vocal folds as functions of those dimensions might be expected. Such variations might influence the strategies for controlling voicing onset and offset during speech by women versus men, and might be important for understanding the development of laryngeal motor control in children.

In recent works, a version of the two-mass model of the vocal folds (Ishizaka and Flanagan, 1972) was used to simulate speech production of adults (Lucero and Koenig, 2005) and children (Lucero and Koenig, 2003) in the vicinity of an abduction gesture. The objective was to determine control strategies of voicing onset and offset used by speakers, and detect possible differences between female, male, and child speakers. There, an inverse dynamic approach was used, in which the model was fitted to collected speech records. The results showed that devoicing during the abduction-adduction gesture for /h/ is achieved by the combined action of vocal fold abduction, decrease of subglottal pressure, and increase of vocal fold tension. Each of these actions has the effect of inhibiting the vocal fold oscillation, suppressing it when reaching an offset threshold. Also, more restricted oscillation regions for women than for men were detected, probably as a consequence of their smaller laryngeal size.

To take a closer look at this result, the following sections will analyze the oscillation threshold conditions versus laryn-

geal dimensions. We remark that our goal is not to develop a new model or present detailed simulations of vocal output; rather, this letter is a brief analytical exercise to explore qualitative relations among laryngeal parameters in the well-established two-mass model of the vocal folds.

II. MODELS

The larynx is modeled using our version of the two-mass model (Lucero and Koenig, 2005), schematically shown in Fig. 1, coupled to a two-tube approximation of the vocal tract in configuration for the vowel /a/ (Titze, 1994). As in our previous work, the standard values of the models' parameters (Ishizaka and Flanagan, 1972; Titze, 1994) are used as reference for a male configuration. To vary their size, a single scaling factor β is used for all dimensions.¹ According to reported experimental data (Goldstein, 1980; Titze, 1989), an adult female configuration would then correspond to an approximate factor of $\beta=0.72$, and a 5-year-old configuration to $\beta=0.64$.

Masses are accordingly scaled by multiplying by β^3 , to compensate for the volume change. For the tissue stiffness, a constant elasticity modulus is assumed for all sizes. In this case, the stiffness coefficient is directly proportional to the cross-sectional area of the tissues, and inversely proportional to their length. Hence, scaling of all dimensions by a factor β implies that stiffness is also scaled by this same factor (see also Titze and Story, 2002). We also incorporate a Q factor for the natural frequencies of the model (Ishizaka and Flanagan, 1972), which may be used to control its oscillation frequency. All masses are divided and stiffness coefficients are multiplied by Q . Finally, a constant damping ratio ξ of tissues is assumed for all sizes, which implies that the damping coefficient $r=2\xi\sqrt{mk}$ (where m is the mass and k is the stiffness coefficient) must be multiplied by β^2 . The use of a single scaling factor for all dimensions and the assumptions of constant elasticity modulus and damping ratio are conve-

^{a)}Electronic mail: lucero@mat.unb.br

^{b)}Electronic mail: koenig@haskins.yale.edu

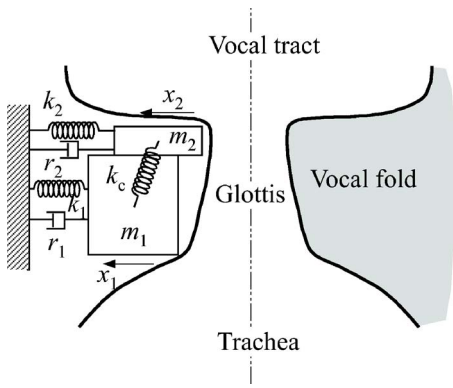


FIG. 1. Two-mass model of the vocal folds.

nient simplifications of the actual variations of size and tissue composition among men, women, and children (e.g., Hirano *et al.*, 1983; Titze, 1989). Such simplifications are justified by our intention to analyze the oscillatory behavior of the vocal folds versus laryngeal size qualitatively, rather than producing detailed simulations of vocal output.

Figure 2 shows plots of simulated oral airflow, as an example of the model's output. The simulations were obtained by varying the glottal half-width from 0.02 to 0.1 cm, and then back to the original value, following a sinusoidal pattern. This variation pattern imitates the glottal abduction-adduction gesture during the production of the utterance /aha/ in running speech (Lucero and Koenig, 2003, 2005). All other parameters were kept fixed at their standard values.

Comparing the plots, we see that the male flow has larger amplitude and lower fundamental frequency, as expected due to the larger size. As the size is reduced, the flow amplitude decreases because of the increased glottal resistance. Also, the oscillation frequency [computed from the ac component of the flow, as described in Lucero and Koenig (2005)] increases: From top to bottom, approximately 123, 165, and 187 Hz. Note that the oscillation frequency is

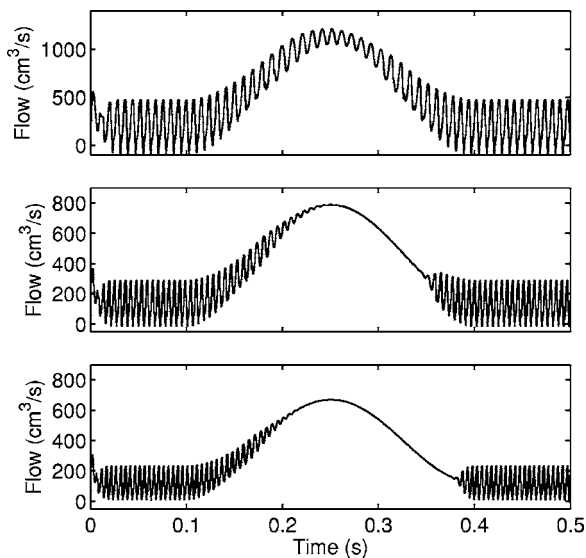


FIG. 2. Oral airflow patterns during a vocal fold abduction-adduction gesture. Top panel: $\beta=1$ (male adult), middle: panel $\beta=0.72$ (female adult), bottom panel: $\beta=0.64$ (5-year-old child).

$2\pi f = \sqrt{(\beta k)/(\beta^3 m)} = (\sqrt{k/m})/\beta$, and so it is inversely proportional to β . The frequency differences are actually smaller than expected, suggesting that the simplifications adopted leave out some aspects of the population differences. In the female case, the glottal pulses stop at the peak abduction, and restart at the end of the following adduction, in a clear oscillation hysteresis phenomenon. In the child case, the glottal pulses stop even earlier than the female case, at a lower value of the glottal width. The plots clearly show that the oscillation region becomes more restricted as the laryngeal size decreases.

III. STABILITY ANALYSIS

The dynamics of the two-mass model in the vicinity of its rest position has been analyzed in previous studies (e.g., Lucero, 1993; Steinecke and Herzel, 1995). According to the theory of dynamical systems (e.g., Perko, 1991), the stability of that position may be determined by taking the linear part of the equations of motion in its vicinity. Simplifying those equations by neglecting losses due to air viscosity, and assuming that the load presented by the vocal tract to the vocal folds is negligible, we find an equilibrium position at the rest position of the vocal folds. The linearized differential equations of the two-mass model around that position are

$$m_1 \ddot{x}_1 + r_1 \dot{x}_1 + (k_c + k_1)x_1 - k_c x_2 = \frac{2d_1 l_g P_s}{x_0} (x_1 - x_2),$$

$$m_2 \ddot{x}_2 + r_2 \dot{x}_2 + (k_c + k_2)x_2 - k_c x_1 = 0,$$

where m_i , with $i=1, 2$, are the masses, x_i are their horizontal displacements measured from a rest (neutral) position $x_0 > 0$, r_i are the damping coefficients, k_i and k_c are the stiffness coefficients, d_1 is the lower mass width, l_g is its length, and P_s is the subglottal pressure. Introducing the size scaling factor β and the Q factor, as explained earlier (i.e., doing the substitutions $m_i \rightarrow \beta^3 m_i/Q$, $r_i \rightarrow \beta^2 r_i$, $k_i \rightarrow \beta Q k_i$, $d_1 \rightarrow \beta d_1$, $l_g \rightarrow \beta l_g$) and computing the characteristic equation of this system, we obtain

$$\beta^4 s^4 + \beta^3 Q (\mu_1 + \mu_2) s^3 + \beta^2 Q^2 (\omega_1^2 + \omega_2^2 + \mu_1 \mu_2) s^2 + \beta Q^3 (\mu_1 \omega_2^2 + \mu_2 \omega_1^2) s + Q^4 (\omega_1^2 \omega_2^2 - \kappa) = 0,$$

where $\mu_1 = r_1/m_1$, $\mu_2 = r_2/m_2$, $\omega_1^2 = (k_1 + k_c - \Gamma)/m_1$, $\omega_2^2 = (k_2 + k_c)/m_2$, $\kappa = k_c(k_c - \Gamma)/(m_1 m_2)$, $\Gamma = 2\beta d_1 l_g P_s / (Q x_0)$, and s is a complex variable.

Changing the variable to $p = \beta s / Q$, the above noted characteristic equation simplifies to

$$p^4 + (\mu_1 + \mu_2) p^3 + (\omega_1^2 + \omega_2^2 + \mu_1 \mu_2) p^2 + (\mu_1 \omega_2^2 + \mu_2 \omega_1^2) p + (\omega_1^2 \omega_2^2 - \kappa) = 0.$$

This last polynomial equation has the general form $p^4 + a_1 p^3 + a_2 p^2 + a_3 p + a_4 = 0$. According to the Routh-Hurwitz criterion (Ogata, 1970), a pair of complex roots cross the imaginary axis from left to right when

$$a_1 a_2 a_3 - a_3^2 - a_1^2 a_4 = 0.$$

This fact signals the occurrence of a Hopf bifurcation, at which the rest position becomes unstable and a limit cycle is

produced, determining the onset threshold of the vocal fold oscillation. Equations (3) and (4) show that the onset threshold condition depends on the value of β only through $\Gamma(\beta)$: note that when substituting the values of the coefficients a_i in Eq. (4), Γ will appear within the expressions for ω_1 and κ ; and further, that Γ will be the only factor related to β . For the standard values of the parameters (Ishizaka and Flanagan, 1972), Eq. (4) has a solution at $\Gamma_{th}=47.80$ N/m. Using also the standard values of d_1 and l_g , we have the threshold relation

$$\frac{\beta P_s}{x_0 Q} \Big|_{\text{threshold}} = 6.83 \times 10^5 \text{ N/m}^3. \quad (5)$$

This relation indicates that, for smaller larynges (smaller values of β), the threshold value of the subglottal pressure to start the vocal fold oscillation must be higher (larger P_s), or the vocal folds must be driven closer together (smaller x_0 , or larger adduction), or the vocal fold tissues must be more relaxed (smaller Q). Let us also note that factor β appears in the expression for Γ due to the reduction in the medial surface of mass m_1 , on which the air pressure acts (if this surface were constant, then the previous conclusions would be just the opposite). Hence, smaller larynges have more restricted phonation regions because their glottal surface is smaller, and so they absorb less energy from the airflow to fuel the vocal fold oscillation.

IV. PHONATION THRESHOLD PRESSURE

The analysis of Sec. III was done under the simplifying assumption of neglecting the effects of air viscosity and vocal tract loading. That simplification was necessary to permit the analytical treatment, and obtain the qualitative relation between size and main control parameters expressed by Eq. (5). However, the question naturally arises: what happens when the model's full equations are considered? We consider here the oscillation threshold for the subglottal pressure; similar results may be obtained for the thresholds on x_0 and Q .

To determine the oscillation threshold pressure, simulations of vocal fold oscillation were performed, while varying the subglottal pressure from 0 to 1000 Pa and back to 0 over a period of 1 s, following a sinusoidal curve as shown in Fig. 3. The simulations were done using both the complete equations of the model, as given in our previous work (Lucero and Koenig, 2005), and also a simplified version without the effects of air viscosity and the vocal tract, which matches the conditions adopted in the previous stability analysis.

From the simulated glottal airflow, the rms amplitude of its ac component was computed cycle-by-cycle, using a zero-crossing algorithm with low pass filtering (Titze and Liang, 1993). The oscillation onset was determined as the instant of time at which the rms flow amplitude increased above a threshold value of $1 \text{ cm}^3/\text{s}$. Similarly, the offset was determined as the instant of time at which the rms flow decreased below $1 \text{ cm}^3/\text{s}$.

Figure 4 shows the computed values of the oscillation thresholds for the subglottal pressure. We can see that there are two different levels of the thresholds, one for onset, and

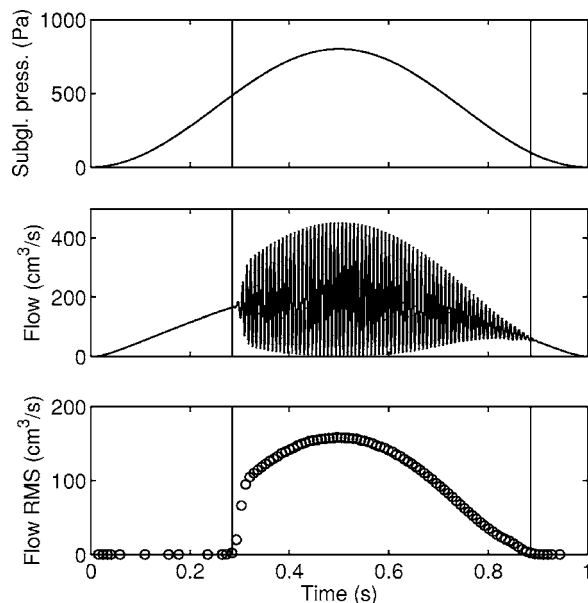


FIG. 3. Simulation results to compute the oscillation threshold value of subglottal pressure. Top panel: subglottal pressure, middle panel: glottal airflow, bottom panel: rms value of the ac glottal airflow. The left and right vertical lines mark the position of the oscillation onset and offset, respectively.

a lower value for offset. Both thresholds increase when the larynx size is reduced, as predicted by Eq. (5). The observed differences between the theoretical values from the stability analysis and simulated results for the same simplifying assumptions (the squares in the plot) come from two sources. One is the computing error inherent to the technique used to detect the thresholds, which overestimates the time of oscillation onset and underestimates the offset. Another source is the time-varying subglottal pressure, which is assumed constant in the stability analysis. It may be verified that the slower its rate of variation, the closer the simulated results are to the theoretical ones. We also note that the onset threshold is higher when using the full equations (circles in the

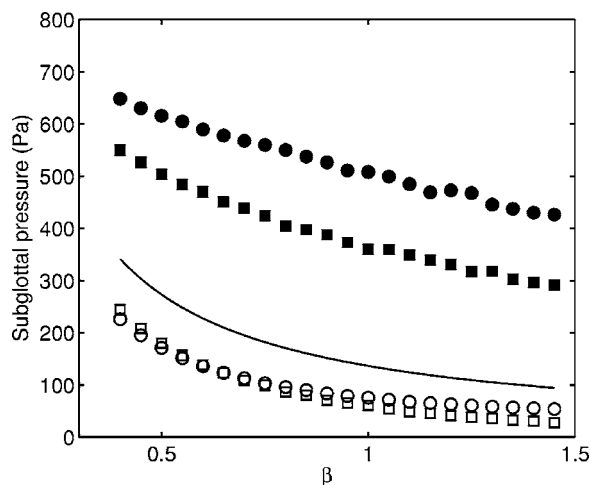


FIG. 4. Oscillation thresholds of subglottal pressure. Circles: thresholds when using the full equations of the model; squares: thresholds when neglecting the effects of air viscosity and the vocal tract. In both cases, the closed marks indicate the oscillation onset, and the open marks indicate the offset. Full line: value predicted by Eq. (5).

plot). One possible explanation is that pressure losses for air viscosity have the consequence of leaving less energy to fuel the oscillation, and therefore of increasing the onset threshold.

V. CONCLUSIONS

In general, the results show that the oscillation conditions of the vocal folds become more restricted as laryngeal size is reduced. This restriction seems to result from a reduction of the glottal area in contact with the airflow, where the energy is transferred from the flow to fuel the vocal fold oscillation. According to this result, women will have in general more restricted conditions for vocal fold oscillation due to a smaller laryngeal size. This would explain their larger occurrence of devoicing in glottal abduction-adduction gestures (Koenig, 2000). As the vocal folds are abducted, the offset threshold position would be easier to reach for women than men. Similarly, child vocal folds will have more restricted oscillation conditions than adults. This restriction could result in their higher values of subglottal pressures observed during speech (e.g., Stathopoulos and Sapienza, 1993; Stathopoulos and Weismer, 1985). In the literature, higher subglottal pressures in children have usually been attributed to higher resistances as a result of smaller airways, although Stathopoulos and Weismer (1985) also suggested that children may simply choose higher speaking volumes. Our results seem to offer a different explanation for the higher pressure values, or at least some motivation behind Stathopoulos and Weismer's suggestion: children may need those higher pressures to get phonatory patterns within acceptable perceptual limits.

The results also suggest the possibility of different strategies for speakers to achieve similar patterns of voicing offset-onset during speech. Equation (5) confirms the results of our previous paper (Lucero and Koenig, 2005), indicating that voice onset may be achieved by the combined action of glottal adduction, increase of subglottal pressure, and decrease of vocal fold tension (for voice offset, the actions are the opposite). Thus, speakers (within a size group) could adopt different vocal tract postures and adjust accordingly the vocal fold tension to arrive at the same phonatory result. Consistent with this hypothesis, our recent experimental data from women (Koenig *et al.*, 2005) provide evidence for interspeaker differences in voicing control.

We remark that the above-noted conclusions must be considered within the simplifying assumptions of the two-mass model. For example, the stability analysis assumes a constant subglottal pressure, independent of the glottal area. This simplification derives from a constant lung pressure and neglect of pressure variations in the subglottal airways. Recent works (Neubauer *et al.*, 2005) have detected a potentially significant influence of the subglottal system on the phonation thresholds, even suppressing onset-offset hysteresis effects. However, we believe that our main conclusions should hold in general qualitative terms, when using more

sophisticated models. Certainly, the influence of the subglottal system on phonation is an interesting subject, which deserves further exploration.

ACKNOWLEDGMENTS

This work was supported by Grant No. DC-00865 from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health, CT-Info/MCT/CNPq and Finatec (Brazil). The results reported here were presented at the Fourth International Conference on Voice Physiology and Biomechanics (ICVPB) in Marseille, 18–20 August 2004.

¹In our previous article (Lucero and Koenig, 2005), we used β as scaling factor from a female reference. We have here inverted the factor and adopt a male reference. This simplifies interpretation of the results, since the β values are directly proportional to laryngeal size.

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

Hirano, M., Kurita, S., and Nakashima, T. (1983). "Growth, development, and aging of human vocal folds," *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp 22–43.

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.

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., Mencl, W. E., and Lucero, J. C. (2005). "Multidimensional analysis of voicing offsets and onsets in female speakers," *J. Acoust. Soc. Am.* (in press).

Lucero, J. C. (1993). "Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region," *J. Acoust. Soc. Am.* **94**, 3104–3111.

Lucero, J. C., and Koenig, L. L. (2003). "Simulations of VhV sequences in children," *Proceedings of the 15th International Congress on Phonetic Sciences*, pp. 2605–2608.

Lucero, J. C., and Koenig, L. L. (2005). "Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control," *J. Acoust. Soc. Am.* **117**, 1362–1372.

Neubauer, J., Zhang, Z., and Berry, D. (2005). "Effects of subglottal acoustics on phonation onset," *J. Acoust. Soc. Am.* **117**, 2542.

Ogata, K. (1970). *Modern Control Engineering* (Prentice-Hall, Englewood Cliffs), pp. 252–258.

Perko, L. (1991). *Differential Equations and Dynamical Systems* (Springer, New York).

Stathopoulos, E. T., and Sapienza, C. (1993). "Respiratory and laryngeal measures of children during vocal intensity variation," *J. Acoust. Soc. Am.* **94**, 2531–2543.

Stathopoulos, E. T., and Weismer, G. (1985). "Oral airflow and air pressure during speech production: A comparative study of children, youths and adults," *Folia Phoniatr.* **37**, 152–159.

Steinecke, I., and Herzel, H. (1995). "Bifurcations in an asymmetric vocal fold model," *J. Acoust. Soc. Am.* **97**, 1878–1884.

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).

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.

Titze, I. R., and Story, B. H. (2002). "Rules for controlling low-dimensional vocal fold models with muscle activation," *J. Acoust. Soc. Am.* **112**, 1064–1076.