ESTIMATION OF LUMPED VOCAL FOLD MECHANICAL PROPERTIES FROM NON-INVASIVE MICROPHONE RECORDINGS

Steven M. Lulich

Department of Speech and Hearing Sciences, Indiana University, Bloomington, USA
slulich@indiana.edu

ABSTRACT

In this paper it is argued that vertical motions of the vocal folds allow sound to be transmitted between the subglottal airways and the vocal tract during the closed phase of vocal fold vibration. From this insight, a pair of equations relating the lumped inertance and compliance of the vocal folds to the frequencies of speech resonances ($F0$ and formants) is derived. These equations include terms for the input impedances of the vocal tract and the subglottal airways measured from the glottis. If these input impedances can be estimated with sufficient accuracy, then the lumped inertance and compliance of the vocal folds can be estimated to arbitrary accuracy depending on the accuracy of formant and $F0$ measurements, and depending on the validity of the assumptions that losses are negligible and that the vocal folds are mechanically isotropic in the coronal plane.

***Index Terms*—** vocal folds, subglottal, vocal tract, models of speech and voice production

1. INTRODUCTION

Vocal fold vibration depends critically on the mechanical properties of the vocal fold tissues, coupled with the time-varying aerodynamic pressure produced by flow through the larynx during phonation. This time-varying pressure is itself subject to influence from the acoustic load of the vocal tract and of the subglottal airways. Titze [1] showed from a time-domain analysis that a vocal fold modeled as a mass-spring-dashpot mechanical system (with mass M, spring constant K, and viscosity B) has an effective mass, M^*, spring constant, K^*, and viscosity, B^*, which are affected by the properties of the acoustic load:

$$M^* = M + 2l_{vf}I_2b(\xi_0 + \bar{\xi}) \quad (1)$$
$$B^* = B - 2l_{vf}I_2\bar{v} + 2l_{vf}R_2b(\xi_0 + \bar{\xi}) \quad (2)$$
$$K^* = K - 2l_{vf}R_2\bar{v} \quad (3)$$

In these equations, l_{vf} is the length of the vocal folds ($l_{vf} \approx 1.6cm$ for adult males), and ξ_0 and $\bar{\xi}$ are the initial (pre-phonatory) and mean phonatory glottal half-width, respectively (ξ is the instantaneous glottal half-width). Other parameters are described in [1]. I_2 and R_2 are the lumped inertance and resistance, respectively, of the vocal tract load impedance as defined in [1]. The acoustic load of the subglottal airways was not included in this derivation, although subsequent work on non-linear source-filter interaction has done so [2].

Motion of the vocal folds in the medial-lateral ('horizontal') direction has been of primary concern to speech and voice researchers, since it is this motion which modulates the airflow through the glottis and enables the voice source to be produced efficiently. In contrast, the focus of this paper is vocal fold motion in the inferior-superior ('vertical') direction. It is argued that such vertical motions allow sound to be transmitted between the subglottal airways and the vocal tract even during the closed phase of vocal fold vibration, as proposed by [3]. Under such a condition, it is possible to derive Equation 1 in a very straight-forward way in the frequency domain by assuming that the vocal folds are mechanically isotropic in the coronal plane and that viscous and acoustic losses are negligible [3]. Moreover, this leads immediately to a pair of analytic equations relating the lumped inertance and compliance of the vocal fold tissue to the frequencies of speech resonances ($F0$ and formants). These equations include terms for the input impedances of the vocal tract and the subglottal airways (measured from the glottis). If these input impedances can be estimated with sufficient accuracy, then the lumped inertance and compliance of the vocal fold tissues can be estimated to arbitrary accuracy depending on the accuracy of formant and $F0$ measurements, and depending on the validity of the two assumptions given above.

Section 2 of this paper presents the argument that sound transmission between the subglottal airways and the vocal tract can be mediated by the vertical motions of the vocal fold tissues during the closed phase, in addition to the motion of the glottal air column during the open phase [3]. Section 3 presents the theoretical ramifications of this insight, including a derivation of Equation 1 and the derivation of equations determining the inertance and compliance of the vocal folds. The use of these new equations is illustrated for a special case in Section 4. Section 5 provides a summary and conclusion, and the relation of this paper to prior work is outlined in Section 6.

2. SOUND TRANSMISSION THROUGH THE VOCAL FOLD TISSUES

The coupling between subglottal and supraglottal airways is generally assumed to occur only during the open phase of the vocal fold vibration cycle [4, 5, 6]. During the open phase (for small amplitude vibrations), the glottal air column between the membranous vocal folds is modeled as a lumped acoustic mass and resistance in series [7].

Acoustic coupling between the subglottal and supraglottal airways need not be mediated solely by the open phase glottal air column, but may also be mediated by either the posterior glottal opening or the vocal fold tissue itself, so that coupling may be possible throughout the vocal fold vibration cycle. During the closed phase, coupling may be dominated by the vocal fold tissue or the posterior glottal opening (henceforth, the cartilaginous glottis), while during the open phase coupling is dominated by the glottal air column [3]. Vocal fold vibration therefore modulates the coupling mechanism and the strength of coupling between subglottal and supraglottal airways within a cycle. The third mechanism - that is, the coupling via the vocal fold tissue itself - is the focus of the present study.

A new model of the *laryngeal impedance* as proposed by [3] is shown in Figure 1. It is a modification of the model introduced by [8] (see also [6], p. 197) and used by [9], [10], and [11] in studies of subglottal-supraglottal coupling. In the original model, the subglottal impedance, Z_{sg}, and the vocal tract impedance, Z_{vt}, were connected in series by the glottal impedance, Z_g , with a dipole source represented by two ideal volume velocity sources, U_s, straddling the glottal impedance and with opposite sign. In the modified model, the glottal impedance, Z_g, is replaced by a 'laryngeal impedance', Z_{lar}, consisting of three parallel impedances representing the membranous glottis, Z_{mg}, the cartilaginous glottis, Z_{cg}, and the vocal fold tissue, Z_{vf}.

The impedances of the membranous glottis and the cartilaginous glottis are modeled by Eqs. 4 and 5, where l_{mg} and $l_{cg} \approx 2L_{mg}/3$ [12] are the anteroposterior lengths of the membranous and cartilaginous glottis, respectively, U_{mg} and U_{cg} are the volume velocities through them, and h is the height of the glottis, which is assumed to be the same for both the membranous and cartilaginous portions of the glottis. The cartilaginous glottis is assumed to be (isosceles) triangular with a base of $2b_{cg}$, so that the area of the cartilaginous glottis is $A_{cg} = l_{cg}b_{cg}$. ξ is the glottal half-width, so that $A_{mg} = 2l_{mg}\xi$ is the area of the membranous glottis (assuming that the membranous glottis is rectangular). ρ is the density of air and $\mu = \eta/\rho$ is the kinematic viscosity, and η is the dynamic viscosity. K_g is a constant which depends on the geometry of the glottal entry and exit.

$$Z_{mg} = \left[\frac{12\mu h}{l_{mg}(2\xi)^3} + K_g\frac{\rho U_{mg}}{(2l_{mg}\xi)^2}\right] + j\omega\frac{\rho h}{2l_{mg}\xi} \quad (4)$$

$$Z_{cg} = \left[\frac{12\mu h}{l_{cg}b_{cg}^3} + K_g\frac{\rho U_{cg}}{(l_{cg}b_{cg})^2}\right] + j\omega\frac{\rho h}{l_{cg}b_{cg}} \quad (5)$$

The vocal folds are assumed to form a circular plate when the glottis is completely closed,with radius equal to the radius of the trachea, $r = r_{trachea}$. A typical value of the tracheal radius is $r_{trachea} \approx 0.8cm$ [13], so that the diameter, $2r \approx 1.6cm$, of the vocal fold plate is roughly equal to the length of the vocal folds. The (axial) cross-sectional area of the vocal folds when the glottis is completely closed is $A_{vf} = \pi r^2$. The cross-sectional area per vocal fold is $A_{vf}/2$, and the impedance of the vocal folds is given by Eq. 6.

$$Z_{vf} = \frac{1}{2}\left[j\omega\frac{m}{(A_{vf}/2)^2} + \frac{1}{j\omega}\frac{k}{(A_{vf}/2)^2} + R\right] \quad (6)$$

where m is the mass of one vocal fold, k is the spring constant, R is the viscous resistance, and the factor of $1/2$ indicates that each vocal fold impedance is assumed to be identical and summed in parallel. Note that Equation 6 gives the impedance of the vocal fold tissue in the inferior-superior direction, or the direction of sound transmission between the subglottal airways and the vocal tract.

Sound transmission through the closed vocal fold tissue is generally not recognized in the speech and voice production literature, so it is worth emphasizing that such a phenomenon is essentially identical to sound transmission through the closed velum [14] or through the vocal tract wall tissues such as the cheeks and neck [15]. Moreover, multi-mass models of vocal fold vibration as well as excised hemilarynx experiments show vertical motions of the vocal folds which are similar in magnitude to the horizontal motions [12, 16, 17].

3. THEORETICAL DERIVATIONS

In this section, it is assumed that viscous and acoustic losses are negligible. This means that the subglottal and vocal tract complex input impedances are replaced by purely imaginary reactances, and that $R = 0$ in Equation 6. Under this assumption, the natural frequencies of the subglottal-laryngeal-supraglottal system during complete glottal closure can be found by summing the subglottal reactance, X_{sg}, with the series laryngeal and vocal tract reactances, $X_{lar} + X_{vt} = X_{vf} + X_{vt}$, and setting the sum to zero:

$$[X_{sg} + X_{vf} + X_{vt}]_{\omega_n} = 0 \quad (7)$$

where $[\cdot]_{\omega_n}$ indicates that the reactances are evaluated at the natural frequency, ω_n.

3.1. Derivation of the effective vocal fold mass

In the case that $F0$ is low compared to the first formant, $F1$, and the first subglottal resonance, $Sg1$, and if $F0$ is considered to be a natural frequency of the system, $\omega_0/(2\pi)$, then lumped inertances can be substituted for X_{sg} and X_{vt} and Equation 6 can be substituted in Equation 7 to obtain the following:

$$j\omega_0 I_1 + j\omega_0 I_2 + \frac{1}{2}\left[j\omega_0\frac{m}{(A_{vf}/2)^2} + \frac{1}{j\omega_0}\frac{k}{(A_{vf}/2)^2}\right] = 0 \quad (8)$$

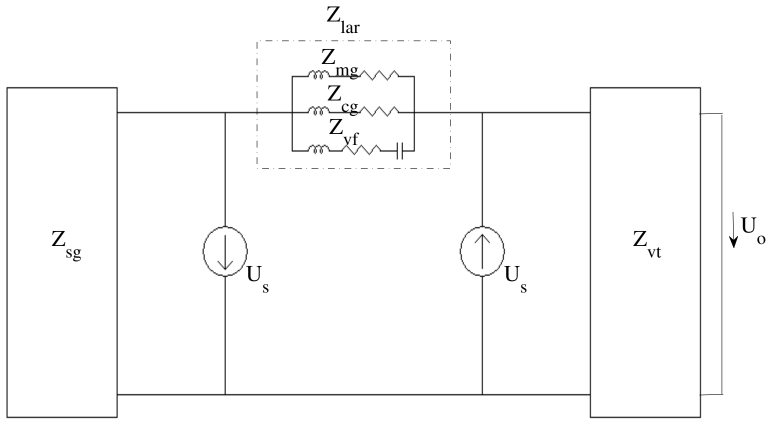


Fig. 1. Model of the subglottal and supraglottal airways coupled in series via the laryngeal impedance.

where I_1 is the inertive load of the subglottal airways and I_2 is the inertive load of the vocal tract. Rearranging terms results in Equation 9.

$$\frac{1}{2}\left[j\omega_n\frac{m+2(A_{vf}/2)^2(I_1+I_2)}{(A_{vf}/2)^2}+\frac{1}{j\omega_n}\frac{k}{(A_{vf}/2)^2}\right]=0 \quad (9)$$

It is now clear that the system operates with an effective mass, m^*, for each vocal fold which depends on the vocal tract and subglottal acoustic loading as well as the vocal fold mass,

$$m^* = m + 2(A_{vf}/2)^2(I_1+I_2) \quad (10)$$

which is equivalent to Equation 1. As in [1], the effective stiffness, k, does not depend on the inertance of the acoustic load (see Equation 3).

3.2. Equations for vocal fold inertance and compliance

Equation 7 can be rewritten in the following useful form:

$$\left[K_n + \frac{1}{2}\left(j\omega L + \frac{1}{j\omega C}\right)\right]_{\omega_n} = 0 \quad (11)$$

where $K_n = [X_{sg} + X_{vt}]_{\omega_n}$ is an imaginary number, $L = m/(A_{vf}/2)^2$ is the lumped vocal fold inertance, and $C = (A_{vf}/2)^2/k$ is the lumped vocal fold compliance. Equation 11 is valid for any resonance frequency, ω_n, of the complete subglottal-larynx-vocal tract system in the absence of losses. If two resonance frequencies are known, $F_m = \omega_m/(2\pi)$ and $F_n = \omega_n/(2\pi)$, then two equations of this form can be written down. If, in addition, the corresponding values of K_m and K_n at each resonance frequency are known or can be estimated, then the result is a system of two equations and only two unknowns: L and C. This system of equations can be solved to yield two closed-form equations for determining L and C:

$$L = \frac{j(K_nF_n - K_mF_m)}{\pi(F_n^2 - F_m^2)} \quad (12)$$

$$C = \frac{F_n^2 - F_m^2}{j4\pi F_mF_n(K_nF_m - K_mF_n)} \quad (13)$$

Resonances which may be used for F_m and F_n could include $F0$ and any formants (or any subglottal resonances).

4. ESTIMATION OF L AND C IN A SPECIAL CASE

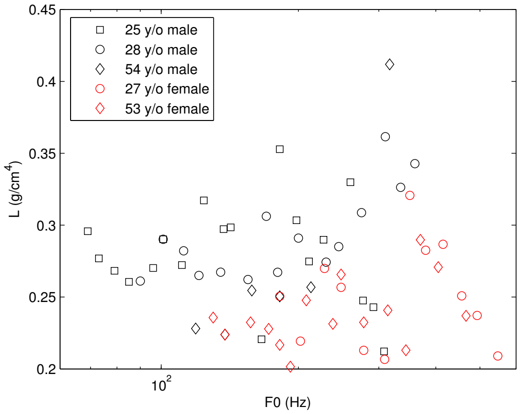
The special case is considered in which $F_m = F0$ and $F_n = F1$, and $K_m = 0$, which indicates no non-linear source-filter interaction. To estimate K_n, a model of the subglottal-larynx-vocal tract system was implemented. The subglottal airways were modeled as described in [19, 10] using the Weibel [13] lung geometry, and the vocal tract was modeled for the vowel [a] using the area function reported by [18]. The vocal folds were modeled as having area $A_{vf} = 1.44cm^2$, with mass $m = 0.06g$ and spring constant $k = 33,000cyne/cm$ [6]. The viscosity of the vocal folds was assumed to be $R \cdot A_{vf}^2 = 80dyne \cdot s/cm$, equal to that of the velum [14]. (Further details of the implementation can be found in [19].) The vowel transfer function spectrum was computed, and the resulting F1 was found to be $785Hz$. From these subglottal and vocal tract models, it was found that $X_{sg} = -26j \;\; g/s \cdot cm^4$ and $X_{vt} = -512j \;\; g/s \cdot cm^4$ at this frequency, so that K_n was equal to $-538j \;\; g/s \cdot cm^4$.

Five adult subjects (3 males, 2 females, aged 25-54) with normal voice quality participated in a small experiment. Each subject produced a sustained vowel [a] at various pitches in a musical major scale, from the lowest note they could produce to the highest. For each note, the fundamental frequency, $F0$, and the first formant, $F1$, were measured. It was assumed, for the sake of illustration, that $K_n = -538j \;\; g/s \cdot cm^4$ for each subject. Equations 12 and 13 were used to calculate the estimated values of L and C for each subject at each pitch in the major scale. The results are shown in Figure 2.[1]

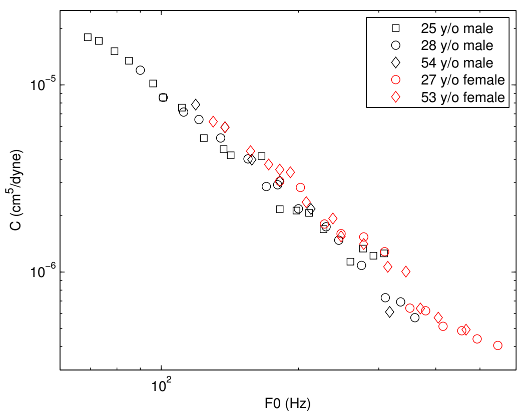
The data show that the vocal fold compliance estimates all follow a single master curve, regardless of gender or age, with compliance decreasing as $F0$ increases. The vocal fold inertance estimates are more variable, although the values for all subjects cover a similar range.

A further step was taken to estimate the mass, m, and

[1] Note that speaker F1 produced vowels with $F0$ up to $951Hz$, but instances with $F0 > 600Hz$ produced estimated values of L and C which were obviously incorrect, including negative values (data not shown). This is presumably due to the difficulty in measuring $F1$ accurately at such high $F0$. Similarly, the highest $F0$ values from speaker M1 are not shown ($F0 > 300Hz$).



(a) Estimated vocal fold inertance, L



(b) Estimated vocal fold compliance, C

Fig. 2. Estimated values of vocal fold inertance, L, and compliance, C, as a function of $F0$ for five speakers.

spring constant, k, of the vocal folds. For the adult males, it was assumed that the vocal fold length was $l_{vf} = 1.8cm$, and the width of the vibratory part of the vocal folds was $0.5cm$ [6, p. 9], leading to $A_{vf} = 0.9cm^2$. For females, $l_{vf} = 1.3cm$ and the width was $0.4cm$, leading to $A_{vf} = 0.52cm^2$. The mass and spring constant could then be calculated from $L = m/(A_{vf}/2)^2$ and $C = (A_{vf}/2)^2/k$. Table 1 reports the median mass, m, and spring constant, k, for each subject.

Stevens [6] suggests that the effective mass of a single vocal fold is on the order of $m = 0.025g$ for females and $m = 0.06g$ for males. For males, these values agree well with the median values obtained in the present study. For females, they are somewhat smaller than the median estimated

Table 1. Median mass, m, spring constant, k values, and age (in years) for the five speakers.

Speaker	age	m(g)	k(dyne/cm)
M1	25	0.0587	48,616
M2	28	0.0577	70,730
M3	54	0.0511	58,572
F1	27	0.0507	52,674
F2	53	0.0471	34,980

values. Similarly, [6] reports the compliance per unit length to be $l_{vf}/k = 2.5 \cdot 10^{-5}cm^2/dyne$ for females and $l_{vf}/k = 3 \cdot 10^{-5}cm^2/dyne$ for males. These correspond to spring constants $k = 52,000dyne/cm$ for females ($l_{vf} = 1.3cm$) and $k = 60,000dyne/cm$ for males ($l_{vf} = 1.8cm$). These values are in good agreement with the estimated values for both males and females.

5. SUMMARY AND CONCLUSION

In this paper, it has been argued that vertical motions of the vocal folds allow sound to be transmitted between the subglottal airways and the vocal tract even during the closed phase of vocal fold vibration. The derivation of Titze's effective mass (Equation 1) in the frequency domain further suggests the accuracy of this insight. A pair of equations relating the lumped vocal fold inertance and compliance to the frequencies of speech resonances was derived, and their potential for use in estimating vocal fold mechanical properties using non-invasive microphone recordings was illustrated.

Estimates of the lumped acoustic mass, L, do not show a systematic trend with increasing frequency, either within or across speakers, but vary between 0.2 and $0.4g/cm^4$. This may indicate that the estimation of L is currently accurate only to within $0.1g/cm^4$, or roughly 30%. Since L and C are related approximately according to the formula $F0 = 1/(2\pi)\sqrt{1/LC}$, errors of 30% in L must be compensated by errors of the same magnitude in C for a given $F0$. Estimation errors should decrease once the factors K_n and K_m can be more precisely determined, and when effects of acoustic losses are included.

6. RELATION TO PRIOR WORK

The work presented here has focused on the theoretical ramifications of the insight that the vocal fold tissues may transmit sound between the subglottal airways and the vocal tract [3]. It was shown how the focus on vertical motion makes it possible to derive analytic equations for estimating vocal fold lumped mechanical properties using non-invasive microphone recordings of vowels. This is in contrast to current methods which typically rely on high speed endoscopic imaging of horizontal vocal fold vibration and high-dimensional multi-mass models of the vocal folds in order to estimate vocal fold mechanical properties [20]. Such methods remain superior when accurate 3-dimensional models of vocal fold mechanics are required, but the approach developed in this paper shows promise for situations in which such methods are unavailable, such as during ambulatory monitoring of vocal health or fast and non-invasive vocal health screening applications.

7. ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant No. 0905250. I am grateful for the many stimulating discussions of the Voice Quality Study Group at MIT, which were very helpful in developing the ideas presented in this paper.

8. REFERENCES

[1] Ingo R. Titze, “The physics of small-amplitude oscillation of the vocal folds,” *The Journal of the Acoustical Society of America*, vol. 83, no. 4, pp. 1536–1552, 1988.

[2] Ingo R. Titze, “Nonlinear source-filter coupling in phonation: Theory,” *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 2733–2749, 2008.

[3] Steven M. Lulich, Matías Zañartu, Daryush D. Mehta, and Robert E. Hillman, “Source-filter interaction in the opposite direction: subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase,” *Proceedings of Meetings on Acoustics (POMA)*, vol. 6, pp. 060007, 2009.

[4] Gunnar Fant, Kenzo Ishizaka, J. Lindqvist, and J. Sundberg, “Subglottal formants,” *STL-QPSR*, vol. 1, pp. 1–12, 1972.

[5] Dennis H. Klatt and Laura C. Klatt, “Analysis, synthesis, and perception of voice quality variations among female and male talkers,” *The Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 820–857, 1990.

[6] Kenneth N. Stevens, *Acoustic Phonetics*, MIT Press, Cambridge, MA, 1998.

[7] James L. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer-Verlag, Berlin, second edition, 1972.

[8] Helen M. Hanson and Kenneth N. Stevens, “Subglottal resonances in female speakers and their effect on vowel spectra,” *Proceedings of the XIIIth international congress of phonetic sciences*, vol. Stockholm, 3, pp. 182–185, 1995.

[9] Xuemin Chi and Morgan Sonderegger, “Subglottal coupling and its influence on vowel formants,” *The Journal of the Acoustical Society of America*, vol. 122, pp. 1735–1745, 2007.

[10] Steven M. Lulich, “Subglottal resonances and distinctive features,” *Journal of Phonetics*, vol. 38, no. 1, pp. 20–32, 2010.

[11] Matías Zañartu, Julio C. Ho, Daryush D. Mehta, Robert E. Hillman, and George R. Wodicka, “An impedance-based inverse filtering scheme with glottal coupling,” *Journal of the Acoustical Society of America*, vol. 125, pp. 2638, 2009.

[12] Ingo R. Titze, *The Myoelastic-Aerodynamic Theory of Phonation*, National Center for Voice and Speech, Denver, CO, 2006.

[13] E. R. Weibel, *Morphometry of the Human Lung*, Springer, Berlin, 1963.

[14] Jianwu Dang and Kiyoshi Honda, “An improved vocal tract model of vowel production implementing piriform resonance and transvelar nasal coupling,” *Proceedings of ICSLP*, pp. 965–968, 1996.

[15] Jacob B. Munger and Scott L. Thomson, “Frequency response of the skin on the head and neck during production of selected speech sounds,” *The Journal of the Acoustical Society of America*, vol. 124, no. 6, pp. 4001–4012, 2008.

[16] T. Baer, *Investigation of phonation using excised larynxes*, Ph.D. thesis, MIT, 1975.

[17] Michael Döllinger and David A. Berry, “Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation,” *Journal of Voice*, vol. 20, no. 3, pp. 401–413, 2005.

[18] Brad H. Story, “Synergistic modes of vocal tract articulation for english vowels,” *Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3834–3859, 2005.

[19] Steven M. Lulich, *The role of lower airway resonances in defining vowel feature contrasts*, Ph.D. thesis, MIT, 2006.

[20] Anxiong Yang, Michael Stingl, David A. Berry, Jörg Lohscheller, Daniel Voigt, Ulrich Eysholdt, and Michael Döllinger, “Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model,” *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 948–964, 2011.