Proceedings of Meetings on Acoustics

Volume 6, 2009 http://asa.aip.org

157th Meeting
Acoustical Society of America
Portland, Oregon
18 - 22 May 2009
Session 3aSC: Speech Communication

3aSC8. Source-filter interaction in the opposite direction: subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase

Steven M. Lulich*, Matias Zanartu, Daryush D. Mehta and Robert E. Hillman

***Corresponding author's address: Speech Communication Group, Massachusetts Institute of Technology, Cambridge, MA 02139, lulich@speech.mit.edu**

Studies of speech source-filter interaction usually investigate the effect of the speech transfer function (loading) on vocal fold vibration and the voice source. In this study we explore how vocal fold mechanics affect the transfer function throughout the glottal cycle, with emphasis on the closed phase. Coupling between the subglottal and supraglottal airways is modulated by the laryngeal impedance. Although coupling is generally thought to occur only during the open phase of vocal fold vibration, a posterior glottal opening and the vocal fold tissue itself can allow sound transmission, thereby introducing coupling during the closed phase as well. The impedance of the vocal fold tissue at closure is shown to be small enough to permit coupling throughout the phonatory cycle, even in the absence of a posterior glottal opening. Open- and closed-phase coupling is characterized using mathematical models of the subglottal and supraglottal airways, and the parallel laryngeal impedances of the membranous glottis, posterior glottal opening, and vocal fold tissue. Examples from sustained vowels are presented, using synchronous recordings of neck skin acceleration, laryngeal high-speed videoendoscopy, electroglottography, and radiated acoustic pressure.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.3269926]
Received 28 Oct 2009; published 10 Nov 2009

I. INTRODUCTION

In studies of source-filter interaction during phonation, the typical procedure is to investigate how various aspects of the acoustic loading of the vocal tract and subglottal airways interact with the mechanical and geometric properties of the vocal fold tissue and the airflow through the glottis to produce specific patterns of vocal fold vibration (Titze, 1988, 2008; Zañartu *et al.*, 2007; Becker *et al.*, 2009; Zhang *et al.*, 2006). Titze (1988) derived formulas to describe the dependence of the effective vocal fold mechanical properties on the acoustic load of the vocal tract (Eqs. 43-45):

$$M^* = M + 2LI_2b(\xi_0 + \bar{\xi}) \tag{1}$$

$$B^* = B - 2LI_2\bar{v} + 2LR_2b(\xi_0 + \bar{\xi}) \tag{2}$$

$$K^* = K - 2LR_2\bar{v} \tag{3}$$

where M, B, and K are the intrinsic mass, damping, and stiffness of the vocal fold tissue per unit area, and M^*, B^*, and K^* are the effective mass, damping, and stiffness per unit area of the tissue-load system, respectively. L is the length of the vocal folds ($L \approx 1{,}6cm$ for adult males), ξ_0 and $\bar{\xi}$ are the initial (pre-phonatory) and mean phonatory glottal half-width, respectively (the instantaneous glottal half-width is ξ), and $\bar{v}$ is the mean glottal flow particle velocity. The constant b is a constant of proportionality relating the oscillatory component of glottal flow to the lateral motion of the vocal folds, $\tilde{v} = -b\dot{\xi}$, where $v = \bar{v} + \tilde{v}$ is the total glottal flow (this approximation essentially assumes a monopole phonation sound source). I_2 and R_2 are the lumped inertance and resistance, respectively, of the vocal tract load impedance as defined in Titze (1988).

This model assumes that the fundamental frequency of phonation is lower than the first formant, that the vibration of the vocal folds is small in amplitude (meaning that the vocal folds are slightly abducted and oscillate without completely closing), and it does not take into account the loading effects of the subglottal airways. More recent work (Titze, 2008; Titze *et al.*, 2008; Titze, 2006) has considerably expanded upon the theory outlined by Titze (1988), but the basic principles are the same, namely, that the acoustic load has specific effects on the effective mechanical properties of the vocal folds and hence on vocal fold vibration and phonation.

In this paper, we outline a complementary theory of tissue-load interaction in which our primary goal is to consider the effect of vocal fold mechanical properties on the (loading) acoustic transfer function from the glottis to the lips. The time domain representation of Titze (1988) and our frequency domain representation of source-tract interaction will be shown to be equivalent.

In the next section (Section II) we outline the mechanisms of supraglottal-subglottal coupling during vowel production and the role that vocal fold vibration plays in modulating this coupling. A frequency domain model of the subglottal airways, larynx, and vocal tract will be described. In Section III we present the results of a simulation showing how the laryngeal impedance changes during the course of a vocal fold vibration cycle, as well as the effects on vowel transfer functions. Results from a pilot study of human speech are presented in Section IV. In Section V we discuss the findings and draw some conclusions.

II. LARYNGEAL MODULATION OF SUBGLOTTAL-SUPRAGLOTTAL COUPLING

The coupling between subglottal and supraglottal airways is generally assumed to occur only during the open phase of the vocal fold vibration cycle (Fant *et al.*, 1972; Klatt and Klatt, 1990; Stevens, 1998). During the open phase (for small amplitude vibrations), the glottal air column between the vocal folds is modeled as a lumped acoustic mass and resistance in series (Flanagan, 1972):

$$Z_g = \left[\frac{12\mu h}{L(2\xi)^3} + K_g\frac{\rho U_g}{(2L\xi)^2}\right] + j\omega\frac{\rho h}{2L\xi} \tag{4}$$

where h is the vertical height of the glottis, U_g is the glottal volume velocity, ρ is the density of air and $\mu = \eta/\rho$ is the kinematic viscosity, and η is the dynamic viscosity. L and ξ are as defined above, so that $A_g = 2L\xi$ is the glottal area (we assume that the glottis has a rectangular shape). K_g is a constant which depends on the geometry of the glottal entry and exit, and we will assume a value of $K_g = 1{,}325$ (Ananthapadmanabha and Fant, 1982) throughout this paper.

We propose that coupling between the subglottal and supraglottal airways is not mediated solely by the open phase glottal air column, but may also be mediated by a posterior glottal opening and by the laryngeal (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, while during the open phase coupling may be dominated by the glottal air column. Vocal fold vibration therefore modulates the coupling mechanism and the strength of coupling between subglottal and supraglottal airways within a single cycle.

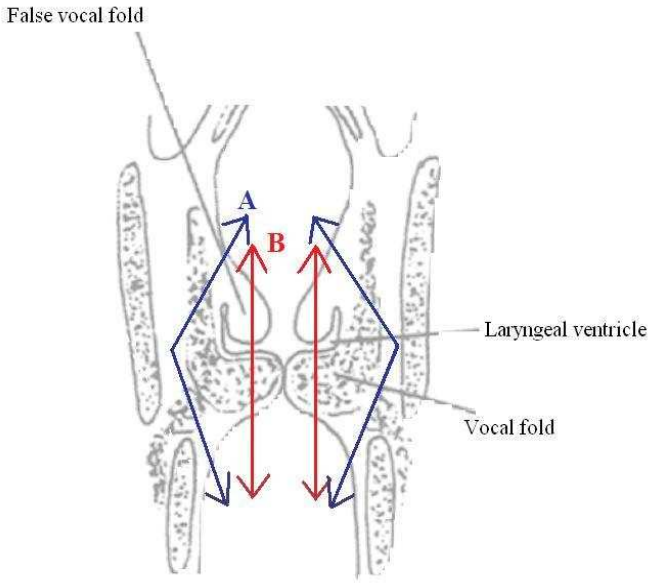

Figura 1. Schematic diagram of two paths of tissue-mediated sound transmission between the subglottal and supraglottal airways. The coronal plane is shown, with thyroid and cricoid cartilages on the periphery. The 'indirect pathís labeled 'Aánd the 'direct pathís labeled 'B'. (Adapted and modified from Dickson and Maue-Dickson, 1982).

In order to avoid possible confusion later on, we will refer to two portions of the glottis: the membranous glottis and the cartilaginous glottis. We will assume that the cartilaginous glottis is synonymous with the posterior glottal opening, which may be open throughout the glottal cycle even when the membranous glottis achieves complete closure, whereas the membranous glottis is associated with the time varying opening occurring in normal phonation. The air-borne impedance of the cartilaginous glottis will be referred to as Z_{cg} for the remainder of this paper, and likewise the air-borne impedance of the membranous glottis will be referred to as Z_{mg}. If the cartilaginous glottis remains abducted during the closed phase of the membranous glottis, then coupling between the subglottal and supraglottal airways is still possible.

Whatever the state of the membranous or cartilaginous glottis, a third means of subglottal-supraglottal coupling is possible. The tissues of the larynx, including the vocal folds themselves, may facilitate coupling throughout the vocal fold vibration cycle. We consider two paths of tissue-mediated sound transmission (hence coupling) between the subglottal and supraglottal airways, as illustrated in Figure 1. The path labeled 'Aís the 'indirect path', and indicates the sound transmitted from the trachea to its walls, bypassing the vocal folds and causing the walls of the supraglottal vocal tract to vibrate and thus radiate sound back into the vocal tract itself. The path labeled 'Bís the 'direct path', and indicates the sound transmitted directly from the trachea to the vocal fold tissue and then into the vocal tract. (Of course, sound transmission actually occurs in both directions, not just the direction indicated by the arrows in the figure.) In the remainder of this paper, we will assume that subglottal-supraglottal coupling via the 'indirect pathís negligible. We will also assume that the false vocal folds play a negligible role in modulating subglottal-supraglottal coupling. Further investigation of the validity of both assumptions is worth pursuing in the future.

We therefore consider three mechanisms of coupling between the subglottal and supraglottal airways: 1) membranous glottal coupling, 2) cartilaginous glottal coupling, and 3) vocal fold tissue coupling. The model we will explore in this paper is shown in Figure 2. It is a modification of the model introduced by Hanson and Stevens (1995) (see also Stevens, 1998, p. 197) and recently used by Chi and Sonderegger (2007), Lulich (2009), and Zañartu *et al.* (2009) in studies of subglottal-supraglottal coupling. In the original model, the subglottal impedance, Z_{sg}, and the vocal tract impedance, Z_{vt}, are connected in series by the glottal impedance, Z_g (as defined above, Equation 4), with a dipole source represented by two 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}.

We assume that the membranous glottis is rectangular, so that its impedance is given by

$$Z_{mg} = \left[\frac{12\mu h}{L_{mg}(2\xi)^3} + K_g \frac{\rho U_{mg}}{(2L_{mg}\xi)^2} \right] + j\omega \frac{\rho h}{2L_{mg}\xi} \tag{5}$$

where L_{mg} is the length of the membranous glottis (typically about two-thirds the total glottal length, Titze, 2006)

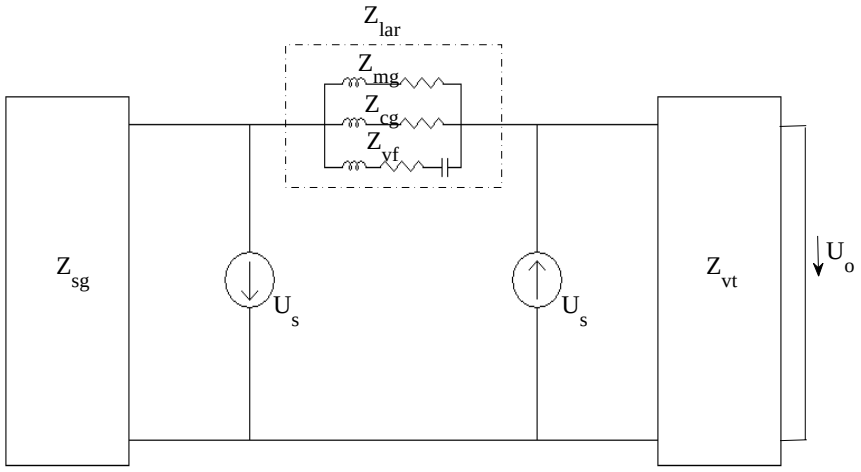

Figura 2. Frequency domain model of the subglottal and supraglottal airways coupled in series via the laryngeal impedance. The twin volume velocity sources model a dipole glottal sound source. The laryngeal impedance consists of the membranous glottal impedance, the cartilaginous glottal impedance, and the vocal fold tissue impedance summed in parallel.

and U_{mg} is the volume velocity through the membranous glottis. The height of the glottis, h, is assumed to be the same for both the membranous and cartilaginous glottis. The glottal half-width, ξ, refers only to the membranous glottis.

We assume that the cartilaginous glottis is (isosceles) triangular, with an anteroposterior length, L_{cg}, equal to about one-third the total length of the glottis, and a (posterior) base, $2b_{cg}$, so that the area of the cartilaginous glottis is $A_{cg} = L_{cg}b_{cg}$. We model the cartilaginous glottis with an equivalent rectangular area, so its impedance is given by

$$Z_{cg} = \left[\frac{12\mu h}{L_{cg}b_{cg}^3} + K_g\frac{\rho U_{cg}}{(L_{cg}b_{cg})^2}\right] + j\omega\frac{\rho h}{L_{cg}b_{cg}} \tag{6}$$

where U_{cg} is the volume velocity through the cartilaginous glottis. (Note that the first two terms on the right hand side of Equation 6 depend on the geometry of the opening, so that our use of an equivalent rectangular area introduces an error here.)

We assume that the vocal folds form a circular plate when the glottis is completely closed, as in Figure 3, with radius equal to the radius of the trachea, $r = r_{trachea}$. A typical value of the tracheal radius is $r_{trachea} \approx 0{,}8cm$ (Weibel, 1963), so that the diameter, $2r \approx 1{,}6cm$, of the vocal fold plate is roughly equal to the length of the vocal folds. The portion of the vocal folds through which sound is transmitted is assumed to be somewhat less deep (in the superior-inferior direction) than the height of the glottis during its open phase, $h_{vf} \leq h$. This is because the vocal folds are incompressible in the frequency range of interest, so that they must become thinner when adducted and thicker when abducted (Titze, 2006, p. 176). Rather than calculate the mass of the vocal folds as $m = \rho_{vf}\pi r^2 h_{vf}$, where ρ_{vf} is the density of the vocal fold tissue, we follow Stevens (1998) in assuming that for male speakers the effective mass per unit length of each vocal fold is roughly $0{,}06g/cm$, so that the total mass per vocal fold is given by $m = 0{,}06L$, where $L = L_{mg} + L_{cg}$ is the total length of the vocal folds. The stiffness of each vocal fold, k, is an intrinsic property of the tissue and does not depend on the vocal fold dimensions. The (axial) cross-sectional area of the vocal folds when the glottis is completely closed is $A_{vf} = \pi r^2$. When the glottis is not completely closed, the mass

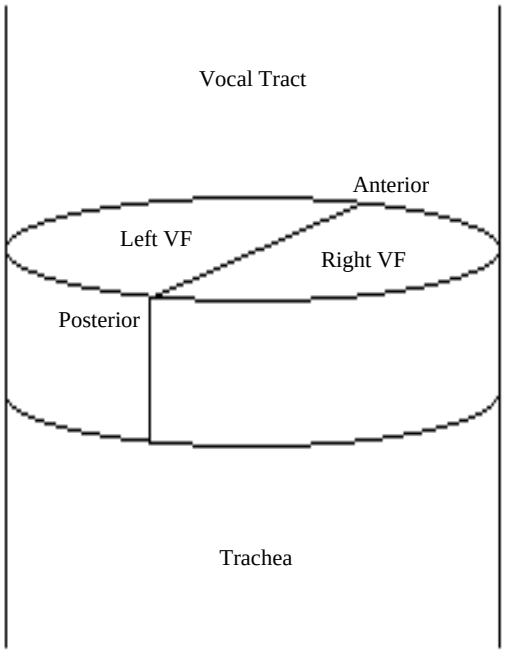

Figura 3. Model of the vocal folds as a circular plate with area $A_{vf} = \pi r^2$ during the closed phase, where r is the radius of the plate.

and stiffness of the vocal folds are assumed to remain unchanged, but the area is reduced to $A_{vf} = \pi r^2 - (A_{mg} + A_{cg})$, where $A_{mg} = 2L_{mg}\xi$ is the area of the membranous glottis and $A_{cg} = L_{cg}b_{cg}$ is the area of the cartilaginous glottis, as described above. The cross-sectional area per vocal fold is therefore $A_{vf}/2$, and the impedance of the vocal folds is given by

$$Z_{vf} = \frac{1}{2}\left[j\omega\frac{m}{(A_{vf}/2)^2} + \frac{1}{j\omega}\frac{k}{(A_{vf}/2)^2} + R\right] \tag{7}$$

where R is the viscous resistance of the vocal fold tissue and the factor of $1/2$ indicates that each vocal fold impedance is assumed to be identical and summed in parallel.

If we ignore losses throughout the vocal tract, the subglottal airways, and the vocal folds, we can determine the natural frequencies of the whole system during complete glottal closure 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:

$$[X_{sg} + X_{vf} + X_{vt}]_{\omega_n} = 0 \tag{8}$$

where $[\cdot]_{\omega_n}$ indicates that the reactances are evaluated at the natural frequency, ω_n. If the fundamental frequency, $F0$, 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, we can substitute lumped inertances for X_{sg} and X_{vt} in Equation 8 and obtain the following:

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

where I_1 is the inertive load of the subglottal airways and I_2 is the inertive load of the vocal tract. After a final

rearrangement of terms we obtain

$$\frac{1}{2}\left[j\omega_n \frac{m + 2(A_{vf}/2)^2(I_1 + I_2)}{(A_{vf}/2)^2} + \frac{1}{j\omega_n}\frac{k}{(A_{vf}/2)^2}\right] = 0 \tag{10}$$

in which it becomes clear that the system operates with an effective mass, m^*, which depends on the vocal tract and subglottal acoustic loading as well as the vocal fold mass:

$$m^* = m + 2(A_{vf}/2)^2(I_1 + I_2) \tag{11}$$

Compare this result with Equation 1. After taking account of the fact that M is the mass per unit area, these two equations correspond exactly. In Equation 1, the acoustic load is assumed to come only from the vocal tract, and the cross-sectional area of this load is assumed to be related to the glottal area, $2L\xi$. In Equation 11, the acoustic load of both the subglottal and supraglottal airways are combined, with a cross-sectional area determined by the vocal fold area, A_{vf}. The second (aerodynamic) term of Equation 1 is a factor of 2 smaller than the equivalent term in Equation 11 because it is concerned with a single vocal fold, whereas Equation 11 is concerned with both vocal folds in parallel. As in Titze (1988) (see Equation 3), the effective stiffness, k, does not depend on the inertance of the acoustic load.

An important difference in the derivation of Equations 1-3 and Equation 10 is that the former are concerned with the lateral motion of the vocal folds whereas the latter is concerned with their vertical motion. Therefore, although the results have an identical mathematical form, it is possible that the corresponding mechanical properties of mass, stiffness, and damping are not identical. We will, however, assume identity for the reason that the vocal folds are known to be (approximately) mechanically isotropic in the coronal plane (Titze, 2006, p. 170). Under this assumption, the fundamental frequency of vocal fold vibration, $F0$, calculated using Titze's method is identical to the fundamental frequency of vocal fold vibration calculated using the method proposed here.

A final point worth emphasizing is that, within the current framework, the fundamental frequency of vocal fold vibration can be considered a formant just like the first, second, and higher formants, in that they are all natural frequencies of the combined tissue-load system.

III. NUMERICAL SIMULATION

A. Methods

The purpose of the following simulations was to gain some insight into how the time-varying laryngeal impedance might affect the coupling between subglottal and supraglottal airways, and the transfer function of vowels. We begin by assuming the glottal area function for a single cycle described by Ananthapadmanabha and Fant (1982):

$$A_{mg}(t) = \begin{cases} A_{max}\left[0{,}5 - 0{,}5cos\left(\pi\dfrac{t}{T_o}\right)\right], & 0 < t < T_o \\ A_{max}\left[cos\left(\pi\dfrac{t - T_o}{2T_c}\right)\right], & T_o < t < T_o + T_c \\ 0, & T_o + T_c < t < T_0 \end{cases} \tag{12}$$

where A_{max} is the maximum area of the membranous glottis, T_o is the duration of the opening phase, T_c is the duration of the closing phase ($T_o + T_c$ is the total open phase duration), and T_0 is the fundamental period ($T_0 - T_o - T_c$ is the closed phase duration). A subglottal pressure, P_s, was assumed to drive the glottal flow. The parameter values used in this simulation are given in Table I. Note that $L = L_{mg} + L_{cg} = 1{,}6cm$, and the base of the cartilaginous glottis, $b_{cg} = 0{,}044cm$, corresponds to an angle between the two vocal processes slightly greater than 10 degrees.

We calculated the quasi-steady glottal flow volume velocity at several times during the glottal cycle separately for the membranous and cartilaginous glottis. The cartilaginous glottis was assumed to be constant throughout the cycle, so that the flow, U_{cg}, through it was also constant. The membranous glottal flow, U_{mg}, however, was not constant, and was calculated by solving the following quadratic equation relating pressure and flow:

$$P_s = \frac{12\mu h}{L_{mg}(2\xi)^3}U_{mg} + K_g\frac{\rho}{(2L_{mg}\xi)^2}U_{mg}^2 \tag{13}$$

where we have assumed that the transglottal pressure is fixed and equal to the subglottal pressure, P_s. (Source-filter interaction is typically accounted for by allowing the transglottal pressure to fluctuate, whereas in this case we account

Cuadro I. Parameters used in the simulation study.

A_{max}	$0{,}25cm^2$	L_{mg}	$1{,}1cm$
T_0	$10ms$	L_{cg}	$0{,}5cm$
T_o	$5{,}25ms$	b_{cg}	$0{,}044cm$
T_c	$1{,}75ms$	h	$0{,}934cm$
P_s	$8cmH_20$	r	$0{,}8cm$
μ	$1{,}86 \cdot 10^{-4} cm^2 \cdot s^{-1}$	m	$0{,}096g$
ρ	$1{,}14 \cdot 10^{-3} g \cdot cm^{-3}$	k	$77000g \cdot s^{-2}$
K_g	1,325	R	$30g \cdot cm^{-4} \cdot s^{-1}$

for it by its effect on the transfer function, cf. Walker and Murphy, 2007.) The cartilaginous glottal flow was calculated using the same equation, substituting L_{cg}, U_{cg}, and b_{cg} for L_{mg}, U_{mg}, and 2ξ, respectively. We then calculated the membranous and cartilaginous glottal impedances using Equations 5 and 6. The vocal fold ('direct path') impedance was calculated using Equation 7, with $A_{vf} = \pi r^2 - (A_{mb} + A_{cg})$, as described above. We have neglected the radiation impedance in this simulation.

B. Results

Figure 4 shows the results of this initial simulation for the frequency $f = 1500Hz$. In panel A, one period of the glottal area function, $A_{mg} + A_{cg}$, is plotted as a function of time. The DC offset is due to the (constant) area of the cartilaginous glottis. The dots on the line indicate specific time points at which the three laryngeal impedances are calculated. The second row shows the magnitude (panel B), the real part (panel C), and the imaginary part (panel D) of the three individual laryngeal impedances. The bottom row shows the magnitude (panel E), the real part (panel F), and the imaginary part (panel G) of various combinations of the three laryngeal impedances in parallel. (The notation in the legend, e.g. '$Z_{mg}||Z_{vf}$', indicates that these impedances were combined in parallel, for instance, $Z_{mg}Z_{vf}/(Z_{mg} + Z_{vf})$.) During the maximally open phase, the membranous glottis dominates the total laryngeal impedance. The difference between the membranous glottal impedance and the total laryngeal impedance does, however, approach several dB during the early part of the opening phase and the late part of the closing phase. The difference between the magnitude of the total impedance during the closed phase and the maximally open phase is on the order of $20dB$, and the closed phase total impedance is dominated by the cartilaginous glottal impedance rather than the vocal fold impedance (although the vocal fold impedance does contribute to decreasing the total parallel impedance by approximately $3dB$), as indicated by the lower impedance magnitude of the curve labeled $Z_{mg} + Z_{cg}$ than of the curve labeled $Z_{mg} + Z_{vf}$. If the cartilaginous glottis is closed so that its impedance is infinite, the laryngeal impedance during the closed phase is approximately $11dB$ higher than in the case of an open cartilaginous glottis, but it is still finite (contrary to the typical assumption that the laryngeal impedance is infinite during complete closure of the glottis).

The input impedance of the subglottal airways, Z_{sg}, the input impedances of the vocal tract for the vowels [i], [e], [a], [u], (Z^i_{vt}, Z^e_{vt}, Z^a_{vt}, Z^u_{vt}, respectively), and the laryngeal impedance, Z_{lar}, during the closed and (maximally) open phase are shown in Figure 5. The four vocal tract input impedances are based on the area functions provided by Story (2005). The subglottal input impedance is based on the model described by Lulich (2006, 2009).

The transfer functions, $T = U_o/U_s$, from the sound source to the lips for each of the four vowels under three laryngeal coupling conditions are shown in Figure 6. The three coupling conditions are: 1) no coupling (i.e. the laryngeal impedance is infinite); 2) laryngeal coupling during the peak opening of the glottis, that is, when $A_{mg} = A_{max}$; 3) laryngeal coupling during the closed phase through the vocal fold tissue and cartilaginous glottis. Figure 7 shows the movement of the first three formant frequencies and amplitudes throughout the glottal cycle for each vowel, when the laryngeal impedance is calculated using the full model (Equations 5-7). The difference between the maximum and minimum values of the frequency and amplitude of each formant are given in Table II for each vowel, along with the differences between the closed phase frequencies and amplitudes and the frequencies and amplitudes in the case with no coupling.

Note that the formants consistently rise in frequency as coupling becomes stronger (i.e. as the laryngeal impedance decreases; see Lulich, 2009, for a further description of this). The amplitudes of the formants consistently decrease as coupling becomes stronger, due primarily to the transfer of energy to the vibrating vocal folds and losses at the glottis.

Depending on the vowel and the formant number, the change in frequency or amplitude over the course of a period with coupling throughout may be small (e.g. F2 in [i]) or quite large (e.g. F1 in [e] and A2 in [a]). The mean increase in formant frequency is $\overline{\Delta f} = 86Hz$, and the mean decrease in amplitude is $\overline{\Delta A} = 9{,}6dB$, averaged over all formants and all vowels. It is clear from Figure 6 that the difference is even greater when comparing the maximum open phase

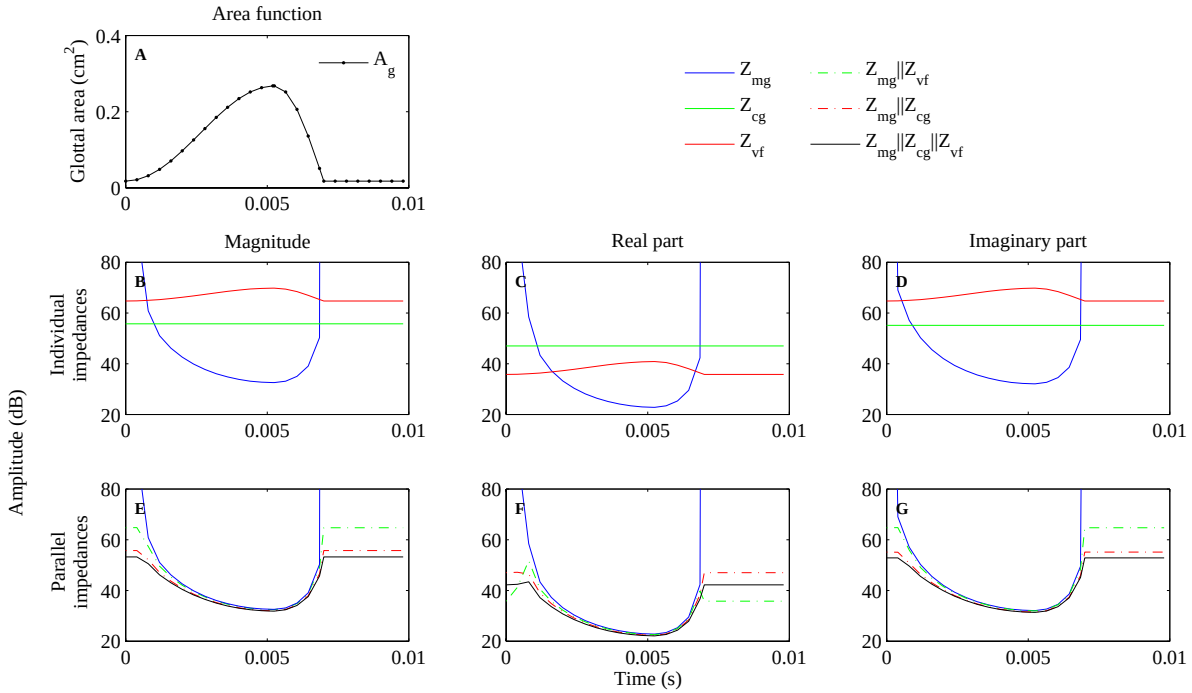

Figura 4. Time-domain simulation of glottal opening (panel A) and laryngeal impedances (panels B-G) at the frequency $1500Hz$. As the glottis opens, the membranous glottal impedance decreases while the cartilaginous glottal impedance remains constant and the vocal fold tissue impedance increases (panels B-D). The three impedances are summed in parallel (in three different configurations, plus the lone membranous glottal impedance) to yield the total laryngeal impedance (panels E-G). Panels B and E are the impedance magnitudes, panels C and F are the real parts, and panels D and G are the imaginary parts.

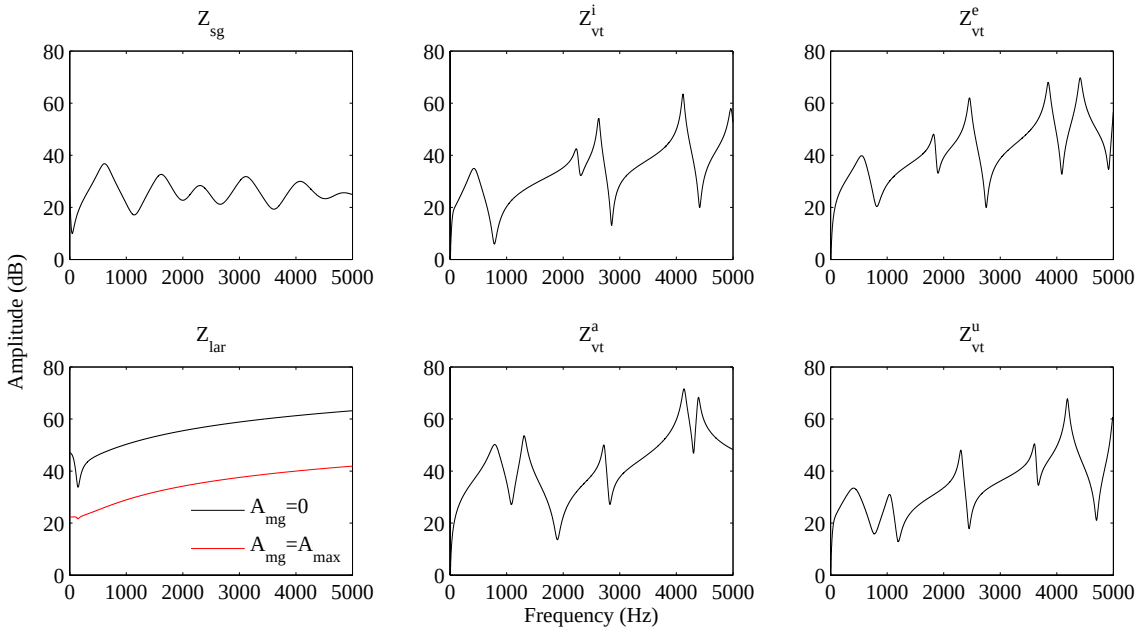

Figura 5. The input impedance of the subglottal airways and the vocal tract for 4 vowels, and the impedance of the larynx during complete closure and maximum opening.

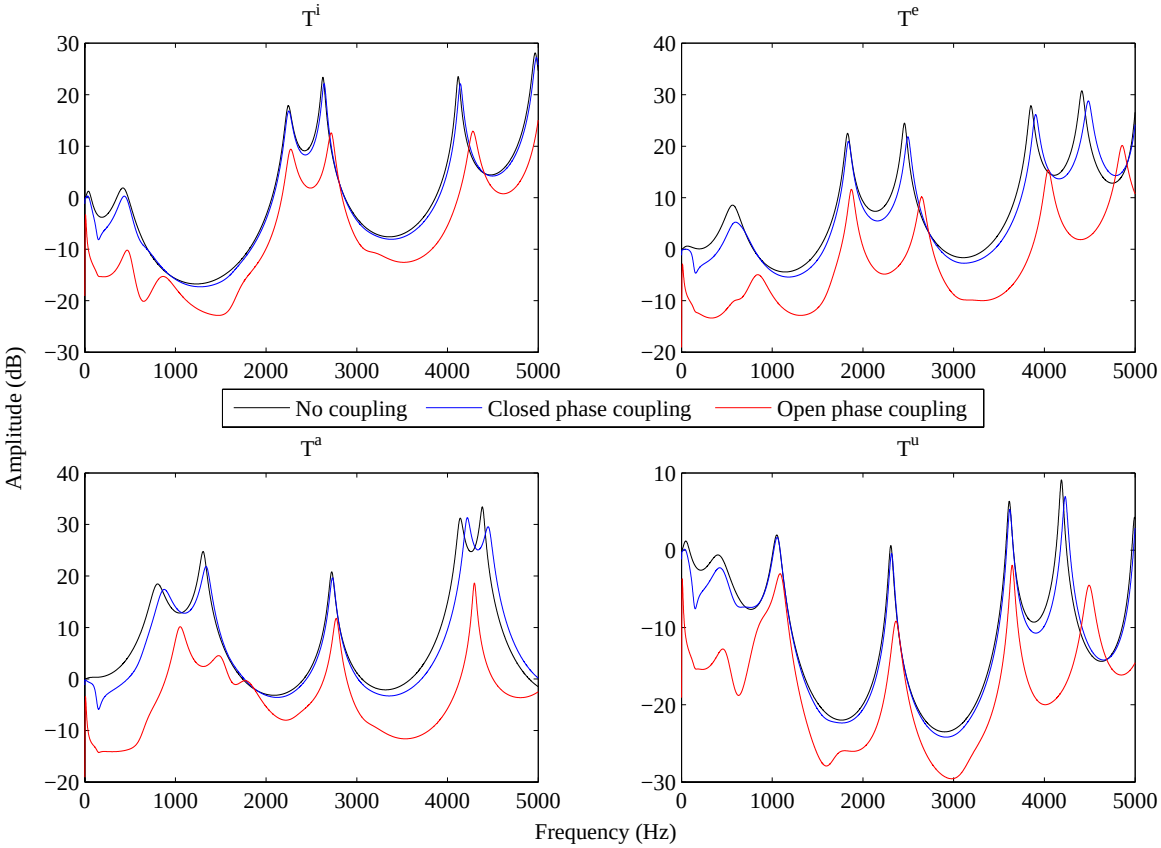

Figura 6. Transfer functions for each vowel under the conditions of no coupling, closed phase coupling, and maximally open phase coupling.

Cuadro II. Range of formant frequencies, Δf, and amplitudes, ΔA, during a single glottal cycle for each vowel, as well as the difference between the frequencies, Δf^0, and amplitudes, ΔA^0, in the closed phase and in the case with no coupling.

	Formant	Δf (Hz)	ΔA (dB)	Δf^0 (Hz)	ΔA^0 (dB)
[i]	F1	31	10.5	16	1.6
	F2	22	7.4	4	1.0
	F3	79	9.5	12	1.3
[e]	F1	245	10.2	32	3.3
	F2	34	9.3	9	1.6
	F3	152	11.7	38	2.6
[a]	F1	176	7.3	73	1.1
	F2	139	17.3	33	2.9
	F3	41	7.9	9	1.1
[u]	F1	34	10.5	20	1.7
	F2	42	4.7	4	0.3
	F3	48	8.8	8	1.0

with the case in which no coupling is allowed. The mean difference in frequencies and amplitudes between the closed phase with coupling and the case in which there is no coupling is $\overline{\Delta f^0} = 21{,}5Hz$ and $\overline{\Delta A^0} = 1{,}6dB$, respectively. In all cases, the difference in frequency, Δf, between the closed and maximally open phases with coupling throughout is greater than the difference, Δf^0, between the closed phase and the case with no coupling. However, for the high vowels [i] and [u], Δf^0 for F1 approaches Δf.

In addition to the formants, there are several formant-like peaks in the transfer functions (see Figure 6). These additional peaks are due to the resonances of the subglottal system. The subglottal input impedance has several

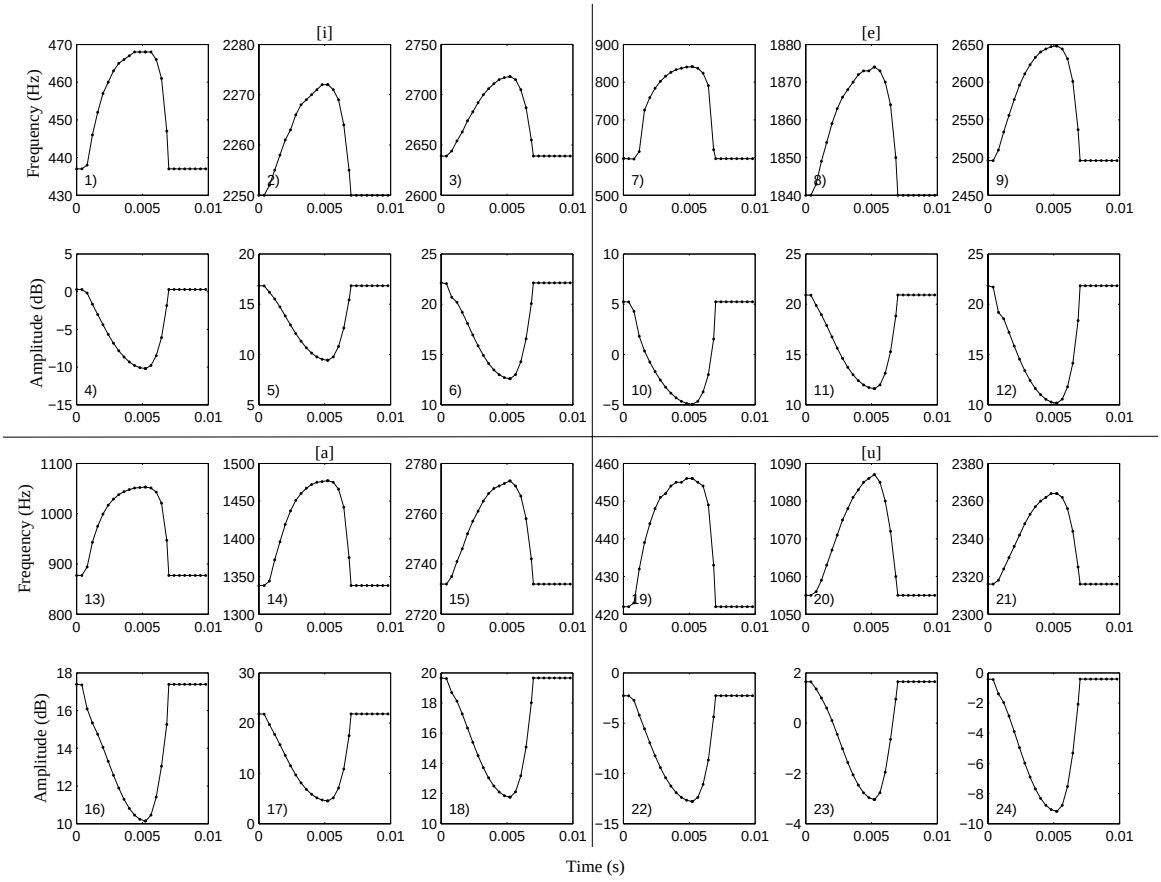

Figura 7. Time course of formant frequencies and amplitudes during a single glottal cycle using the full laryngeal impedance for the vowels [i], [e], [a], and [u]. Panels 1-3 show F1, F2, and F3 frequencies, respectively, of the vowel [i], and Panels 4-6 show the corresponding amplitudes. Panels 7-9, 13-15, and 19-21 similarly show F1, F2, and F3 frequencies of [e], [a], and [u], respectively, and Panels 10-12, 16-18, and 22-24 show the corresponding amplitudes.

resonances which each correspond to a pole-zero pair in the vowel transfer function. The pole is at a somewhat higher frequency than the zero (see Lulich, 2009, for a discussion of this), and it is the pole that appears as an additional peak in the transfer function. The peak near $900Hz$ in the vowel [i] (most clearly visible during the maximal open phase), and again in the vowel [u], is due to the first subglottal resonance, Sg1. Less pronounced shoulders (rather than peaks) can be seen in both the vowels [e] and [a] as well. The peak near $1800Hz$ in the vowels [a] and [u] (most clearly visible during the maximal open phase) are due to the second subglottal resonance, Sg2. It also produces less pronounced shoulders in the vowels [i] and [e]. Effects of other subglottal resonances are present around $2500Hz$ (Sg3) and $3100Hz$ (Sg4). Higher subglottal resonances do not contribute significantly to the transfer function because the laryngeal impedance becomes too great.

In general it appears that subglottal resonances are more likely to appear as peaks in the transfer function if the nearest formant is at a lower frequency than the subglottal resonance. We predict therefore that Sg1 is more likely to be observed in non-low vowels than in low vowels; and similarly Sg2 is more likely to be observed in back vowels than in front vowels. The example of [e] in these simulations does not follow this prediction in that it has F1 at a higher frequency than Sg1 and therefore Sg1 is difficult to observe. This may reflect more on the difficulty of determining a steady-state vocal tract area function for the (strongly diphthongized) vowel [e] in American English than on the predictions made here.

The pronounced dip in the transfer functions near $150Hz$ is due to the natural frequency of vocal fold oscillation, $F0$. At this frequency, acoustic energy is transferred to the vibrating vocal folds rather than transmitted along the vocal tract to the lips. We have seen that the amplitudes of the formants decrease by $9{,}6dB$, on average, from the closed phase to the maximally open phase. The decrease in amplitude at the $F0$ frequency is on the order of 7 or $8dB$. The subglottal resonances contribute pole-zero pairs to the spectrum, which could affect the amplitude of the spectrum, but since each pole lies at a higher frequency than the corresponding zero, each pole-zero pair should

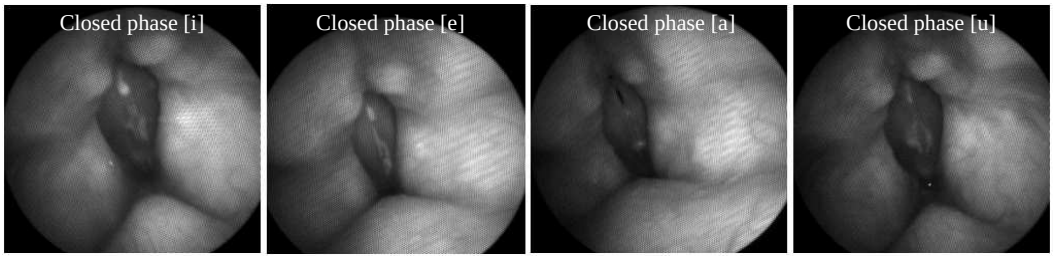

Figura 8. Images from high-speed video showing that the glottis is completely closed in each of the vowels (except for a small opening in the vowel [a]).

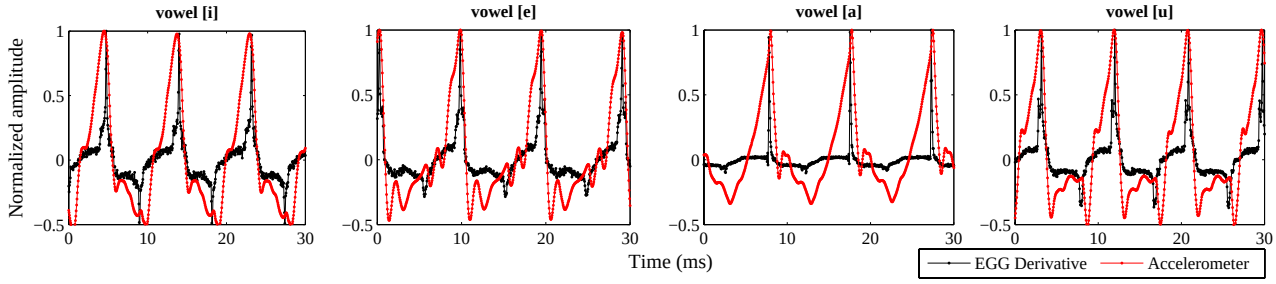

Figura 9. EGG derivative (solid black) and accelerometer (dotted red) signals during portions of each vowel in which the glottis closes completely during the closed phase.

produce a net increase in amplitude at higher frequencies. Subglottal resonances therefore cannot be responsible for the remaining 2 or $3dB$ of amplitude change seen here. We suggest that this amplitude change is due to the resistance of the glottis.

Finally, it is worth noting that the quasi-steady, instantaneous natural frequency of vocal fold vibration changes throughout the cycle, as the source-tract system changes and the effective mass, stiffness, and damping of the vocal folds change.

IV. ANALYSIS OF HUMAN SPEECH

A. Methods

In addition to performing numerical simulations, we analyzed the speech of an adult male native speaker of Spanish sustaining the vowels [i], [e], [a], and [u]. Microphone recordings were made time-synchronously with high-speed videoendoscopy (HSV, achieved using a nasal endoscope), electroglottography (EGG), and recordings from an accelerometer attached to the skin of the neck just above the sternal notch. The microphone, accelerometer, and EGG signals were digitized at a 120 kHz sampling rate, and the HSV signal was digitized at a rate of 4000 frames per second. Portions of the HSV signal of each vowel were identified in which the vocal fold vibration cycle visibly showed complete closure during the closed phase, including a negligible (if any) posterior glottal opening. Figure 8 shows an image from the high-speed video during the closed phase for each of the four vowels.

For each vowel the closed phase portion of the accelerometer signal was determined by inspecting the HSV record and the EGG signal. The (solid black) EGG derivative and (dotted red) accelerometer signals are shown together in Figure 9 for each vowel. In the EGG derivative signal, the beginning of the closed phase is marked by a large, abrupt positive excursion. The end of the closed phase is marked by a smaller but similarly abrupt negative excursion (Childers and Larar, 1984).

The autoregressive (AR) based poles of the microphone ('mic') and accelerometer ('acc') signals were computed to estimate the poles of the supraglottal and subglottal transfer function, respectively. The relation between these signals and transfer functions has been established in previous studies (cf. Cheyne *et al.*, 2003). The closed phase pole frequencies given in Table III were averaged over three consecutive cycles for each vowel. The closed phase of each cycle was obtained by selecting the portion between the large positive and negative peaks of the EGG derivative signal and considering a propagation delay associated with the location of each sensor. This was performed to minimize the error

Cuadro III. Pole frequencies (Hz) from the linear prediction analysis during the closed phase for both the microphone ('mic') and accelerometer ('accel') signals. Bold black type indicates F1; bold red type indicates Sg1; underlined black type indicates F2; underlined red type indicates Sg2; other poles are in italic type.

vowel	[i]	vowel	[e]	vowel	[a]	vowel	[u]
mic	accel	mic	accel	mic	accel	mic	accel
244	**414**	**433**	*161*	**551**	**528**	**247**	**254**
538	**662**	**601**	**549**	**684**	**726**	**583**	**749**
1424	1356	1896	1354	1132	1310	*1517*	1349
2230	*1766*	*2203*	1686	*1756*	*1705*	*2117*	*1950*
2728	*2499*	*2405*	*2103*	*2443*	*2492*	*2430*	*2570*
3166	*3215*	*3095*	*3338*	*3480*	*3476*	*3212*	*3271*

Cuadro IV. Formant and subglottal resonances frequencies from the long-window linear prediction analysis of the microphone ('mic') and accelerometer ('accel') signals, respectively. Only the first two formants and subglottal resonances are reported for comparison with the data in Table III. The second subglottal resonance for the vowel [e] was not detected. The font type and color are the same as in Table III.

vowel	[i]	vowel	[e]	vowel	[a]	vowel	[u]
mic	accel	mic	accel	mic	accel	mic	accel
212	**529**	**393**	**606**	**773**	**590**	**208**	**573**
2374	1503	1857	—	1168	1490	644	1589

that is introduced when incorrectly estimating the boundaries of the closed phase portion (Alku *et al.*, 2009). The poles were obtained via the covariance method of linear prediction, using a rectangular window of length given by the entire closed phase portion (always larger than twice the AR model order for this case). The order of the AR analysis was computed as suggested by Rabiner and Schafer (1978), resulting in an order $p = 18$. Only the first six formant frequencies are presented in Table III for simplicity. Finally, we also performed a long-window linear predictive coding (LPC) analysis (window longer than 1 second) of both the microphone and accelerometer signals.

B. Results

Comparison of the accelerometer waveforms in Figure 9 shows clear differences between the four vowels during both the closed and open phases. There may be many reasons for these differences. One possibility is that the vocal folds do not have the same posture in all four vowels, so that the subglottis is elongated or compressed or otherwise shaped differently. However, the general posture of the supraglottal larynx appears to be roughly the same in all four vowels (see Figure 8), and there is no further positive evidence that there are significant differences in vocal fold posture. A second possibility is that effects of subglottal-supraglottal coupling during the open phase are simply carrying over into the closed phase in the form of phase differences (i.e. since the boundary conditions at the time of closure are different, the transient response to the change in the system from open to closed glottis is different for each vowel, even though the steady-state response might presumably be identical). We cannot entirely rule this possibility out at this time. A third possibility is that subglottal-supraglottal coupling is occurring during the closed phase as well as during the open phase. There may be other possible explanations for these differences, but these appear to us to be the three most likely.

The lowest six pole frequencies from our short-window LPC analysis of the closed phase accelerometer and microphone spectra are reported in Table III, and the results of the long-window LPC anlaysis are reported in Table IV. We have indicated in Table III a possible, straight-forward interpretation of these poles in terms of both the formants and the subglottal resonances (compare with Table IV). In all four vowels, a pole near $550Hz$ (in bold, red type) is identified in both the microphone and accelerometer signals. This is in the expected frequency range for Sg1. For the vowel [a], a higher frequency pole near $700Hz$ corresponds to F1, whereas for the other vowels F1 is lower in frequency than Sg1 (between $244Hz$ and $433Hz$; in bold, black type). F1 is detected for each of the vowels in the microphone signal, and for all but the vowel [e] in the accelerometer signal. A pole near $1350Hz$ is identified in the accelerometer signal of each vowel (in underlined, red type), and in the microphone signal of the vowel [i]. This is the frequency range expected for Sg2. F2 is identified in all but the vowel [u] in the microphone signal, and in the accelerometer signal of the vowel [e] (in underlined, black type).

It is especially noteworthy that in the accelerometer signal, the linear prediction analysis resulted in poles corresponding to Sg1 and Sg2 in each of the vowels, and these poles did not vary significantly from vowel to vowel. The analysis of the accelerometer data also resulted in a pole corresponding to F1 in each of the vowels. This pole

varied according to the expectation that F1 is lowest for high vowels, [i] and [u], highest for low vowels, [a], and intermediate for mid vowels, [e]. Furthermore, the absolute values of F1 obtained from this analysis are appropriate for the individual vowels. This indicates that the supraglottal first formant, F1, was represented in the subglottal accelerometer signal, and supports the hypothesis that subglottal-supraglottal coupling may occur during the closed phase. F2 was generally not identified by the analysis of the accelerometer signal. The pole at $2230Hz$ for the vowel [i] is approximately what we would expect, but a similar pole appeared in the other vowels as well and may be an artefact of the analysis procedure.

V. DISCUSSION AND CONCLUSIONS

We have briefly outlined a theory of subglottal-supraglottal coupling which includes sound transmission not only through the membranous and cartilaginous glottis, but also through the vocal fold tissue. This theory was shown to be equivalent to the theory developed by Titze (1988).

Based on a numerical simulation, we found that although coupling during the open phase is stronger than coupling during the closed phase, the effect of closed-phase coupling on vowel spectra is not negligible. The dominant effects of open-phase coupling are to decrease the overall amplitude of the spectrum and shift the formants to higher frequencies. We also found that subglottal poles are more visible in the vowel spectrum when the nearest formant is at a lower frequency.

In a pilot study of human speech, we carried out linear prediction analyses of an accelerometer signal (measured above the sternal notch) and a microphone signal during four vowels. The results suggest that vowel-specific formants appear as poles in the accelerometer signal even when the vocal folds are completely closed. Although more data is needed to support the proposed theory, our preliminary observations promote the hypothesis that subglottal-supraglottal coupling does occur during the closed phase via the vocal fold tissue.

Acknowledgments

This work was supported in part by NSF grant number 0828903 and NIH-NIDCD grant number T32 DC00038, and by the Institute for Laryngology and Voice Restoration. Special thanks to Dimitar D. Deliyski for facilitating the use of his laboratory's high-speed video camera, which was possible through the NIH grant R01 DC007640-01A2.

REFERENCES

Alku, P., Magi, C., Yrttiaho, S., Bäckström, T., and Story, B. (**2009**), "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering", The Journal of the Acoustical Society of America **125**, 3289-3305.

Ananthapadmanabha, T. V. and Fant, G. (**1982**). "Calculation of true glottal flow and its components", Speech Communication **1**, 167-184.

Becker, S, Kniesburges, S., Müller, S.,Delgado, A.,Link, G.,Kaltenbacher, M., and Döllinger, M. (**2009**). "Flow-structure-acoustic interaction in a human voice model", The Journal of the Acoustical Society of America **125**, 1351-1361.

Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., and Hillman, R. E. (**2003**). "Development and testing of a portable vocal accumulator", Journal of Speech Language and Hearing Research **46**, 1457-1467.

Chi, X. and Sonderegger, M. (**2007**). "Subglottal coupling and its influence on vowel formants", The Journal of the Acoustical Society of America **122**, 1735-1745.

Childers, D. G. and Larar, J. N. (**1984**). "Electroglottography for laryngeal function assessment and speech analysis", IEEE Transactions on Biomedical Engineering **31**, 807-817.

Dickson, D. R. and Maue-Dickson, W. (**1982**). *Anatomical and Physiological Bases of Speech*, PRO-ED: Austin, TX.

Fant, G., Ishizaka, K., Lindqvist, J., and Sundberg, J. (**1972**). "Subglottal formants", STL-QPSR **1**, 1-12.

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

Hanson, H. M. and Stevens, K. N. (**1995**). "Sub-glottal resonances in female speakers and their effect on vowel spectra", Proceedings of the XIIIth international congress of phonetic sciences, Stockholm, **3**, 182-185.

Klatt, D. H. and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers", The Journal of the Acoustical Society of America **87**, 820-857.

Lulich, S. M. (**2006**). "The role of lower airway resonances in defining vowel feature contrasts", Ph.D. thesis, MIT.

Lulich, S. M. (**2009**). "Subglottal resonances and distinctive features", Journal of Phonetics, doi:10.1016/j.wocn.2008.10.006.

Rabiner, L. R. and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* Prentice-Hall: Englewood Cliffs, N.J.

Stevens, K. N. (**1998**). *Acoustic Phonetics* MIT Press: Cambridge, MA.

Story, B. H. (**2005**). "Synergistic modes of vocal tract articulation for English vowels", Journal of the Acoustical Society of America **118**, 3834-3859.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds", The Journal of the Acoustical Society of America **83**, 1536-1552.

Titze, I. R. (**2006**). *The Myoelastic-Aerodynamic Theory of Phonation* National Center for Voice and Speech: Denver, CO.

Titze, I. R. (**2008**). "Nonlinear sourcefilter coupling in phonation: Theory", The Journal of the Acoustical Society of America **123**, 27332749.

Titze, I. R., Riede, T., and Popolo, P. (**2008**). "Nonlinear sourcefilter coupling in phonation: Vocal exercises", The Journal of the Acoustical Society of America **123**, 1902-1915.

Walker, J. and Murphy, P., (**2007**). "A Review of Glottal Waveform Analysis", In *Progress in Nonlinear Speech Processing* (Stylianou, Y., Faundez-Zanuy, M., and Eposito, A., Eds.), pp. 1-21, Springer: Berlin.

Weibel, E. R. (**1963**). *Morphometry of the Human Lung* Springer: Berlin.

Zañartu, M., Ho, J. C., Mehta, D. D., Hillman, R. E., and Wodicka, G. R. (**2009**). "An impedance-based inverse filtering scheme with glottal coupling", Journal of the Acoustical Society of America **125**, 2638.

Zañartu, M., Mongeau, L., and Wodicka, G. R. (**2007**). "Influence of acoustic loading on an effective single mass model of the vocal folds", The Journal of the Acoustical Society of America **121**, 11191129.

Zhang, Z., Neubauer, J., and Berry, D. A. (**2006**). "The influence of subglottal acoustics on laboratory models of phonation", The Journal of the Acoustical Society of America **120**, 1558-1569.