Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCa: Flow, Structure, and Acoustic Interactions During Voice Production I

5aSCa6. Acoustic coupling during incomplete glottal closure and its effect on the inverse filtering of oral airflow

Matías Zañartu*, Julio C. Ho, Daryush D. Mehta, Robert E. Hillman and George R. Wodicka

***Corresponding author's address: Department of Electronic Engineering, Universidad Tecnica Federico Santa Maria, Valparaiso, 2390123, Valparaiso, Chile, matias.zanartu@usm.cl**

Inverse filtering of oral airflow using closed-phase linear prediction is expected to preserve the effects of source-filter interactions in the glottal airflow pulse. Under incomplete glottal closure, the glottal airflow estimation is more challenging due to a lowered glottal impedance, increased subglottal coupling, and violated all-pole assumption. To account for these effects, a model-based inverse filtering scheme allowing for coupling between glottis and upper and lower airways was developed. Acoustic transmission in the tracts used a frequency-domain transmission line. A linearized, time-varying expression was used for the glottal impedance, along with a dipole representation. Synthetic vowels sounds and actual recordings were used to evaluate the proposed scheme. Subject-specific model parameters were obtained from simultaneous aerodynamic, acoustic, and high-speed videoendoscopic recordings of normal subjects uttering vowels with various degrees of glottal closure. Results illustrated that, even under incomplete glottal closure, the airflow entering the vocal tract preserved source-filter interactions and was comparable to that obtained using closed-phase linear prediction. The scheme also yielded an uncoupled glottal airflow that exhibited a clear pulse de-skewing, making it proportional to the glottal area. Cases with larger glottal gaps exhibited less pulse skewing, reduced unsteady flow, and a time-varying glottal impedance matching a parallel gap.

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

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

In order to estimate glottal airflow, the filtering effects of the acoustic loads need to be removed via inverse filtering (IF). This is typically performed using source-filter theory principles, i.e., the acoustic loads can be linearly separated from the source and removed via filter estimation techniques. However, current nonlinear principles (Titze, 2008) bring the fundamental separation between source and filter into question. In spite of the presence of source-filter interactions, it has been suggested that separability between source and filter is still possible (Krishnamurthy and Childers, 1986). This separation can be achieved in two ways, by assuming that: 1) the source is independent and that the vocal tract has different formant frequencies and bandwidths during the open and closed phase, or 2) the vocal tract is time-invariant as in the closed phase and that the glottal source contains the formant frequency and bandwidth changes from the previous case. The latter explains the vowel dependent ripples normally observed during the open phase of the glottal airflow (Ananthapadmanabha and Fant*,* 1982).

A widely used scheme for inverse filtering uses autoregressive (AR) parametric modeling, where an all-pole model and a least squared error estimation are utilized to approximate the vocal tract filter (Makhoul, 1975). This method, normally referred to as linear prediction, has been dominant in speech processing due to its computational characteristics, error tractability, stability, and spectral estimation properties (Walker and Murphy, 2007). Two types AR modeling are common: the autocorrelation and covariance methods. The autocorrelation method assures stable solutions, but requires a large number of samples (including several cycles) to yield a reliable solution. The covariance method requires a much smaller number of samples, allowing for estimation even within a single cycle. Due to the source-filter interactions, linear prediction applied to a whole period (or more) will contain formant frequency and bandwidth errors (Krishnamurthy and Childers, 1986). This evidently means that the estimation of the properties of the vocal tract (e.g., formant frequencies and bandwidths) obtained via autocorrelation and covariance methods of linear prediction will differ. In fact, closed phase inverse filtering (CPIF) using a constrained covariance method has been shown to provide better estimates of the glottal waveform than those obtained with the autocorrelation method (Plumpe *et al.*, 1999; Walker and Murphy, 2007; Alku *et al.* 2009). However, it is unclear how the CPIF scheme handles incomplete glottal closure where the glottal impedance is lowered, subglottal coupling is present, and there is no true closed phase—all of which violate the all-pole assumption. Further insights into the performance of this and other IF schemes under incomplete glottal closure are needed, especially as they relate to the clinical assessment of vocal function.

One approach to address this problem is to incorporate the effects of glottal coupling into the inverse filtering scheme. This idea was initially proposed by Rothenberg and Zahorian (1977) in an attempt to extract the glottal area from an oral airflow signal. Their method differs from other inverse filtering schemes in that it attempts to remove not only the vocal tract filtering, but also its interaction with the glottal source. A time-varying Norton equivalent airflow of the glottal source was used along with a feedback procedure. However, the method considered no subglottal coupling and only a single supraglottal formant, required manual intervention, and resulted in a complex nonlinear implementation, and thus the scheme has not been adopted widely.

In this study, the complete voice production system is described as a collection of linear impedances (subglottal, glottal, and supraglottal) to investigate the acoustic coupling between components during incomplete glottal closure and its effects on the inverse filtering of oral airflow. The present study focuses on the air-borne coupling among components, thus neglecting any tissue-borne coupling (Lulich *et al.*, 2009).

METHODS

Proposed Inverse Filtering Scheme

Coupling between the subglottal, supraglottal, and glottal systems was investigated for incomplete glottal closure using human subject recordings. For this purpose, a model-based inverse filtering scheme that accounts for acoustic coupling between tracts and source-filter interactions was proposed. The coupled dipole model shown in Figure 1 is the foundation for the desired inverse filtering scheme, given its direct application to subglottal coupling (Hanson and Stevens, 1995; Chi and Sonderegger, 2007). This simple but powerful model has five basic components: an ideal airflow source (U_o), the flow entering the vocal tract (U_{supra}), and three different lumped terms representing the subglottal (Z_{sub}), glottal (Z_g), and supraglottal (Z_{supra}) acoustic impedances. Inverse filtering based on this representation is referred to as "impedance-based inverse filtering" (IBIF) (Zañartu, 2010).

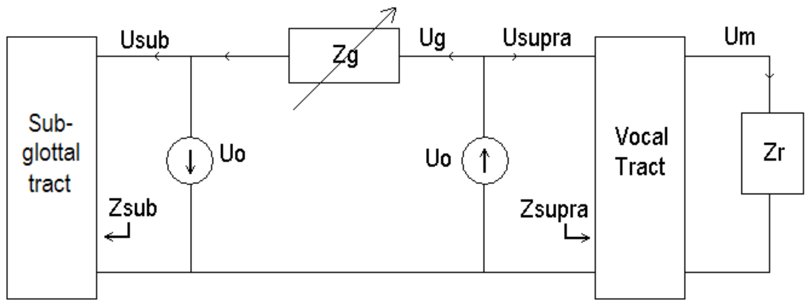

FIGURE 1. Representation of the posterior glottal gap, illustrated in a two-mass model

Each of the system impedances was estimated from experimental data to obtain subject-specific values. In order to estimate the tract impedances, models of acoustic transmission were applied. The proposed frequency-domain transmission model was based on a series of concatenated T-equivalent segments of lumped acoustic elements that related acoustic pressure to volume velocity (Flanagan, 1972; Harper, 2001). The subglottal tract included tube branching down to 24 generations, yielding walls with cartilage and soft tissue, and the standard acoustical representations for losses, elasticity, and inertia in a transmission line model. No nasal coupling was considered for the vocal tract. The transmission line models for the subglottal and supraglottal tracts yielded the driving point impedances as well as transfer functions for any desired location within the tracts. These terms only depended on the tract configuration and its inherent physical properties.

Time-invariant ($\tilde{Z}_g$) and time-varying (Z_g^*) representations of the glottal impedance were investigated. The glottal impedance is a frequency-dependent nonlinear quantity that needs to be linearized to meet the Norton equivalent associated with the dipole circuit representation from Figure 1. Thus, the relation between transglottal pressure (ΔP_g), glottal airflow (U_g), and glottal area (A_g) was rewritten as a Taylor series expansion, resulting in:

$$\tilde{Z}_g = \left.\frac{\partial(\Delta P_g)}{\partial U_g}\right|_{\bar{U}_g,\bar{A}_g} = k_t\frac{\bar{U}_g}{\bar{A}_g^2}. \tag{1}$$

$$Z_{g-kt}^*(t) = \left.\frac{\partial(\Delta P_g)}{\partial U_g}\right|_{\substack{U_g=U_o(t)\\ A_g=A_g(t)}} = kt\frac{U_o(t)}{A_g^2(t)}. \tag{2}$$

The time-invariant impedance was based on mean values for the glottal flow and glottal area, along with an empirical constant (k_t) that depends on the shape of the glottal slit, with an average value of about 1.1 (Titze, 2008). The time-varying representation did not remove the mean values and was evaluated with respect to the time-varying glottal area $A_g(t)$ and time-varying uncoupled airflow $U_o(t)$, which was performed iteratively.

The transfer function between the airflow at the mouth and the airflow entering the vocal tract ($T_1=U_m/U_{supra}$) only depended on the vocal tract (i.e., unaffected by the glottis and subglottal impedances) and was computed using the tract geometry through the aforementioned transmission line scheme. To estimate the coupled airflow, the following relation was considered (Chi and Sonderegger, 2007):

$$T = \frac{U_m}{U_o} = \frac{U_m}{U_{supra}}\cdot\frac{U_{supra}}{U_o} = T_1\cdot T_2, \tag{3}$$

$$T_2 = \frac{U_{supra}}{U_o} = \frac{Z_g}{Z_g + Z_{sub} + Z_{supra}}. \tag{4}$$

The effects of the acoustic coupling on the transfer function are shown in Figure 2 for a time-invariant glottal impedance. Changes in formant frequencies and bandwidths and pole-zero pairs are introduced. To compute the time-varying coupling, a quasi-steady regime was used; i.e., the time-varying case is constructed from a sequence of time-invariant solutions. Since the inertial term of the glottal impedance is negligible (Ananthapadmanabha and Fant*,* 1982), this approach is feasible. Similar assumptions are used in numerical models of phonation (Titze, 1984). Thus, the glottal impedance and coupled transfer functions were estimated and inverse filtered for each sample via the fast Fourier transform and its inverse. Only samples corresponding to associated time steps were retrieved from the set of inverse filtered waveforms.

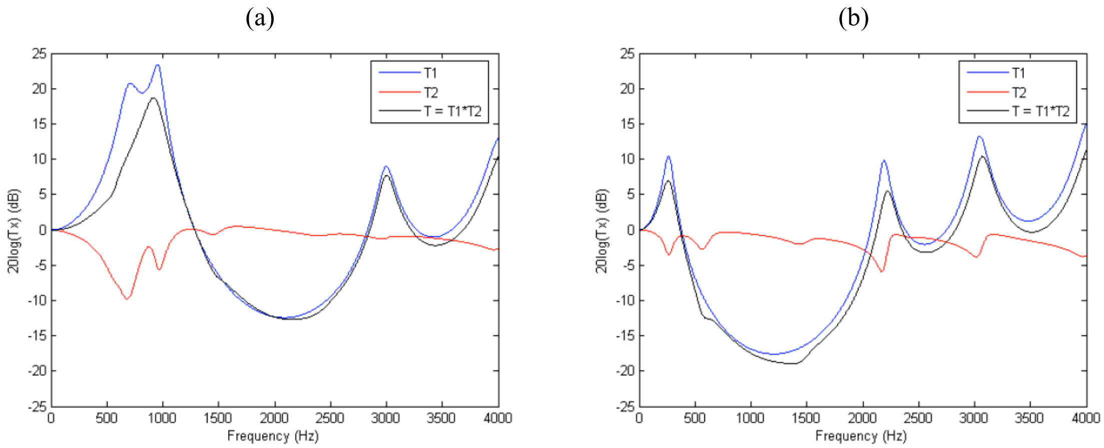

FIGURE 2. Uncoupled (T_1), perturbation function (T_2), and coupled vocal tract transfer function (T) from oral airflow to the glottis for a vowel /a/ and /i/.

Experimental Setup

In order to evaluate the proposed IBIF scheme, a comprehensive set of measurements was obtained. The goal was to obtain estimates of the system behavior through simultaneous recordings of glottal behavior, flow aerodynamics, and acoustic pressure. The experimental setup is described in detail in Zañartu *et al.* (2011) and included synchronous measurements of laryngeal high-speed videoendoscopy (HSV), oral volume velocity (OVV), electroglottography (EGG), radiated acoustic pressure (MIC), and skin surface acceleration (ACC). Two adult subjects (male with no vocal training and female with vocal training) with no history of vocal pathologies uttering different vowels (/a/ and /i/) at different voice qualities (chest, falsetto) were recorded. These vocal gestures imposed distinct acoustic loading conditions and different degrees of incomplete glottal closure.

Glottal area extraction from laryngeal high-speed videoendoscopy was computed using digital image processing, i.e., segmentation and region merging schemes (Mehta *et al.*, 2011). The segmentation yielded a glottal area function in squared pixels that was calibrated to obtain the actual glottal area in absolute units (e.g., square centimeters). The HSV calibration scheme was performed using a laser grid that allowed for quantification of features within the endoscopic view (Kobler *et al.*, 2006) and dedicated software developed for this purpose (Endoview). A laryngeal feature that was adequately visible in the high-speed videos (e.g., a blood vessel on the ventricular folds) was carefully measured in a separate session using the laser grid. This feature was used as a reference to calibrate the spatial units of each HSV recording, where the optic and glottal conditions could change but the dimensions of the laryngeal feature were assumed to remain constant.

In order to estimate the supraglottal impedance, estimates of the vocal tract area functions were obtained from the oral airflow recordings. The method selected for this purpose was based on a systematic variation of high-resolution, MRI-based vocal tract area functions (Story, 2008) using the recursive algorithm described by Story (2006). Subject-specific estimates of subglottal impedance were computed by means of the ACC signal. The acoustical transmission line model of a symmetric branching subglottal representation (Harper, 2000; Harper *et al.*, 2001) was used as default parameter set. This model was then adapted to match subject-specific parameters by adjusting selected parameters to match inverse filtered OVV and ACC signals. This subject-specific calibration scheme is described in detail in Zañartu (2010).

RESULTS

The inverse filtered signals for both vowels and registers using the IBIF scheme are presented in Figure 3. For comparison, the results obtained using CPIF (Alku *et al.*, 2009) are also shown. In this case, the IBIF estimates correspond to those entering the vocal tract (i.e., U_{supra} from Figure 1). As an advantage over CPIF, no closed phase detection was needed to obtain the IBIF estimates. Incomplete glottal closure was observed in all cases and was more significant in the falsetto register, as seen in the DC offset of the glottal airflow signals.

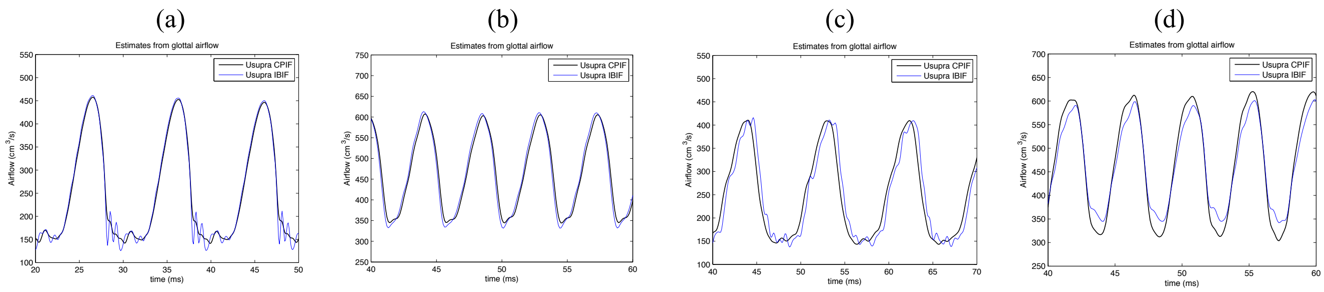

FIGURE 3. Estimates of glottal airflow from CPIF (in black) and IBIF (in blue) for a male subject: (a) Vowel /a/ - chest, (b) Vowel /a/ - falsetto, (c) Vowel /i/ - chest, (d) Vowel /i/ - falsetto

Both inverse filtering schemes yielded almost identical glottal airflow signals for all conditions. It is interesting to note that vowel /i/ has less pulse skewing and a 10% less AC flow than vowel /a/ for the chest register. This indicated that vowel /i/ does impose a different acoustic loading condition for the source. These vowel differences appeared less pronounced when a more abducted glottal condition was presented. Under a more pronounced incomplete glottal closure scenario, the AC flow and pulse skewing are further reduced.

Figure 4 shows the effects of the interaction among all system components on the glottal airflow. The coupled (U_{supra}) and uncoupled (U_o) glottal airflows are compared and contrasted with the glottal area (A_g) for the two types of glottal impedances under study. It can be seen that the time-varying impedance yields an uncoupled glottal flow that is proportional to the glottal area, and thus removes the effects of acoustic coupling. The fixed, time-invariant impedance approach was not capable of achieving this separation.

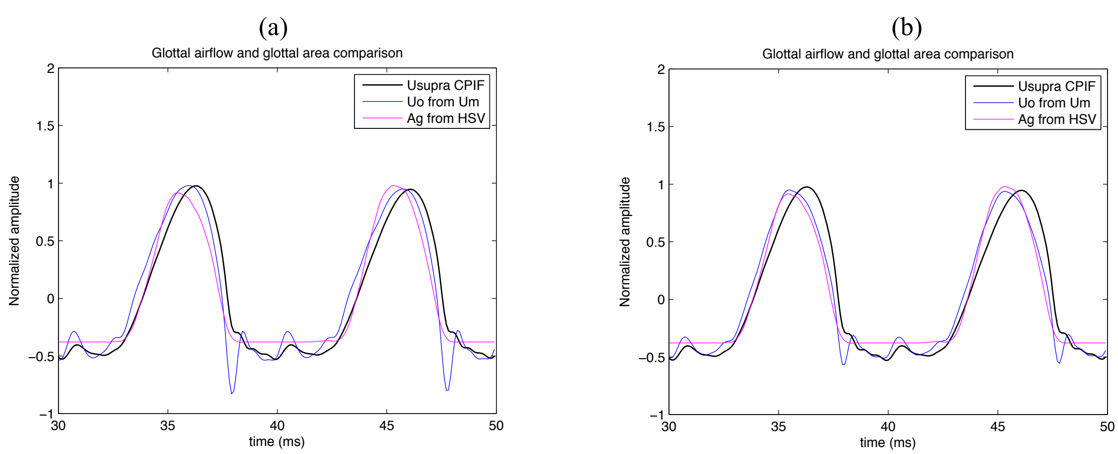

FIGURE 4. Estimates of coupled (U_{supra}, in black) and uncoupled (U_o, in blue) glottal airflow and their contrast with the glottal area waveform (A_g, in magenta) for vowel /a/ in chest register for a male subject: (a) Time-invariant glottal impedance, (b) Time-varying glottal impedance. Amplitudes are normalized with the DC component removed.

As observed in Figure 5, the time-invariant impedance approach yielded a mean glottal impedance of 78 acoustic ohms (cgs), whereas the time-varying one oscillated between 40 and 110 acoustic ohms (cgs). This latter impedance has the structure of a parallel gap, oscillating around the Z_g of the DC airflow during the closed phase, and with a temporal structure that is a minimum at maximum vocal fold excursion. This finding supports the concept that, in this case, the chest voice has an aerodynamic behavior that is equivalent to a posterior glottal gap that was observed from the HSV recording.

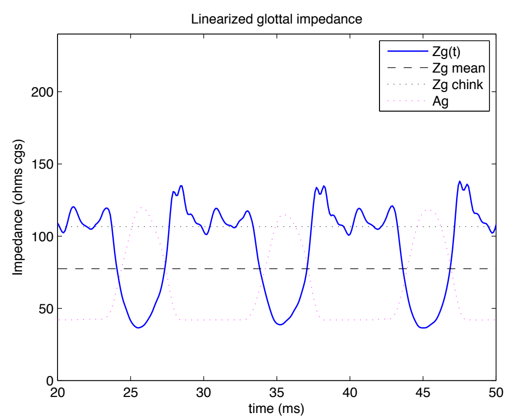

FIGURE 5. Time-invariant and time-varying glottal impedances for a vowel /a/-chest case

DISCUSSION

A transmission line based inverse filtering was used to study the effects of acoustic coupling for different vowels and glottal configurations. The results of this study indicate that glottal and subglottal coupling was not required for the estimation of "true" glottal airflow, i.e., the flow entering the vocal tract (Ananthapadmanabha and Fant, 1982). The standard closed-phase inverse filtering scheme yielded almost identical results, even under incomplete glottal closure scenarios. This finding facilitates the application of the other inverse filtering schemes for general conditions, e.g., a new subglottal inverse-filtering scheme intended for the ambulatory assessment of vocal function (Zañartu, 2010; Mehta *et al.*, 2012; Zañartu *et al.*, submitted).

Exploration of inverse filtering using the coupled version of IBIF indicated that the skewing of the glottal pulse is linked to the effects of acoustic coupling. The results relate the ideal (uncoupled) airflow source to the glottal area, in agreement with previous studies (Rothenberg and Zahorian, 1977). This finding was suggested decades ago through numerical simulations, but had not been properly validated using actual speech recordings. In addition, the proposed IBIF scheme was linear and only based on a quasi-steady assumption. It was also noted that the degree of coupling was not necessary related to the degree of skewing, as cases with incomplete glottal closure exhibited large coupling but less skewing of the glottal pulses and a reduced AC flow, matching prediction from numerical simulations. The time-varying impedance representation was found to better match the expected behavior than the time-invariant one. Even though the time-varying impedance did not explicitly assume a parallel gap configuration, the results are well-aligned with such behavior. This finding is consistent with previous representations of incomplete closure as a parallel impedance (Cranen and Boves, 1987; Cranen and Schroeter, 1995). The uncertainty of the impedance estimates was primarily affected by the degree of glottal exposure during HSV acquisition, as phenomena such as arytenoid hooding and a variable calibration feature may alter the estimated values. Further investigation is needed to improve the estimates of this important parameter from laryngeal HSV.

ACKNOWLEDGMENTS

The work of Matías Zañartu was supported by UTFSM and CONICYT, Grant FONDECYT 11110147. The authors would also like to acknowledge support from the NIH National Institute on Deafness and Other Communication Disorders (T32 DC00038 and R01 DC007640) and the Voice Health Institute.

REFERENCES

Alku, P., Magi, C., Yrttiaho, S., Backstrom, T., and Story, B. **(2009)** "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," J. Acoust. Soc. Am., **125(5)**, 3289–3305.

Ananthapadmanabha, T. V., and Fant, G. **(1982)** "Calculation of true glottal flow and its components," Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm, **23(1)**, 001–030.

Chi, X. and Sonderegger, M. **(2007)**, "Subglottal coupling and its influence on vowel formants," J. Acoust. Soc. Am., **122 (3)**, 1735–1745.

Cranen, B. and Boves, L. **(1987)** "On subglottal formant analysis," J. Acoust. Soc. Am**., 81(3)**, 734–746.

Cranen, B. and Schroeter, J. **(1995)** "Modeling a leaky glottis," J. Phonetics, **23(1-2)**, 165–177.

Flanagan, J. L. **(1972)** *Speech analysis; synthesis and perception*. New York: Springer-Verlag.

Hanson, H. M., and Stevens, K. N. **(1995),** "Subglottal resonances in female speakers and their effect on vowel spectra," in Proceedings of the XIIIth International Congress of Phonetic Sciences, **3**, 182–185.

Harper, P. **(2000),** *Respiratory tract acoustical modeling and measurements*. PhD thesis, Purdue University.

Harper, P., Kraman, S. S., Pasterkamp, H., and Wodicka, G. R. **(2001)** "An acoustic model of the respiratory tract," IEEE Trans. Biomed. Eng., **48**, 543– 550.

Kobler, J. B., Rosen, D. I., Burns, J. A., Akst, L. M., Broadhurst, M. S., Zeitels, S. M., and Hillman, R. E. **(2006)** "Comparison of a flexible laryngoscope with calibrated sizing function to intraoperative measurements," Ann. Otol. Rhinol. Laryngol., **115(10)**, 733–740.

Krishnamurthy, A. and Childers, D. **(1986)** "Two-channel speech analysis," IEEE Trans. Acoust. Speech Sig. Process. **34(4)**, 730–743.

Lulich, S. M., Zañartu, M., Mehta, D. D., and Hillman, R. E. (**2009**). "Source-filter interaction in the opposite direction: Subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase," Proc. Meet. Acoust. **6**, 060007

Makhoul, J. **(1975)** "Linear prediction: A tutorial review," in Proc. IEEE Inst. Electr. Electron. Eng., **63**, 561–580.

Mehta, D. D., Deliyski, D. D., Quatieri, T. F., and Hillman, R. E. (**2011**). "Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings," J. Speech. Lang. Hear. Res. **54**, 47-54.

Mehta, D. D., Zañartu, M., Feng, S. W., Cheyne II, H. A., and Hillman, R. E. (**2012**). "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," IEEE Trans. Biomed. Eng. **59**, 3090-3096.

Plumpe, M. D., Quatieri, T. F., and Reynolds, D. A. **(1999)** "Modeling of the glottal flow derivative waveform with application to speaker identification," IEEE Trans. Speech Audio Process., **7(5)**, 569–586.

Rothenberg, M. and Zahorian, S. **(1977)** "Nonlinear inverse filtering technique for estimating the glottal area waveform," J. Acoust. Soc. Am., **61(4)**, 1063–1070.

Story, B. H. **(2008),** "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," J. Acoust. Soc. Am., **123(1)**, 327–335.

Story, B. H. **(2006),** "Technique for tuning vocal tract area functions based on acoustic sensitivity functions," J. Acoust. Soc. Am., **119(2)**, 715–718.

Titze, I. R. **(1984)** "Parameterization of the glottal area, glottal flow, and vocal fold contact area," J. Acoust. Soc. Am., **75(2)**, 570–580.

Titze, I. R. **(2008)** "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am., **123(5)**, 2733–2749.

Walker, J. and Murphy, P. **(2007)** *A review of glottal waveform analysis*, in Progress in Nonlinear Speech Processing, 1–21, Springer Berlin/Heidelberg.

Zañartu, M. **(2010),** *Acoustic coupling in phonation and its effect on inverse filtering of oral airflow and neck surface acceleration*. PhD thesis, Purdue University.

Zañartu, M., Mehta, D. D., Ho, J. C. Wodicka, G. R., and Hillman, R. E. **(2011)**, "Observation and analysis of *in vivo* vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study," J. Acoust. Soc. Am., **129(1)**, 326–339.

Zañartu, M., Ho, J. C., Mehta, D. D., Hillman, R. E., and Wodicka, G. R., **(submitted)** "Subglottal impedance-based inverse filtering of speech sounds using neck surface acceleration."