Subglottal Impedance-Based Inverse Filtering of Voiced Sounds Using Neck Surface Acceleration

Matías Zañartu [Member, IEEE],
Department of Electronic Engineering, Universidad Técnica Federico Santa Maíia, Valparaíso, Chile

Julio C. Ho,
Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, 47901 USA

Daryush D. Mehta [Member, IEEE],
School of Engineering and Applied Sciences, Harvard University and the Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital, Boston MA 02114 USA

Robert E. Hillman, and
Center for Laryngeal Surgery & Voice Rehabilitation at Massachusetts General Hospital, Harvard Medical School, and Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts 02114

George R. Wodicka [Fellow, IEEE]
Weldon School of Biomedical Engineering and the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47901 USA

Matías Zañartu: matias.zanartu@usm.cl; Julio C. Ho: hoj@purdue.edu; Daryush D. Mehta: dmehta@seas.harvard.edu; Robert E. Hillman: hillman.robert@mgh.harvard.edu; George R. Wodicka: wodicka@purdue.edu

Abstract

A model-based inverse filtering scheme is proposed for an accurate, non-invasive estimation of the aerodynamic source of voiced sounds at the glottis. The approach, referred to as subglottal impedance-based inverse filtering (IBIF), takes as input the signal from a lightweight accelerometer placed on the skin over the extrathoracic trachea and yields estimates of glottal airflow and its time derivative, offering important advantages over traditional methods that deal with the supraglottal vocal tract. The proposed scheme is based on mechano-acoustic impedance representations from a physiologically-based transmission line model and a lumped skin surface representation. A subject-specific calibration protocol is used to account for individual adjustments of subglottal impedance parameters and mechanical properties of the skin. Preliminary results for sustained vowels with various voice qualities show that the subglottal IBIF scheme yields comparable estimates with respect to current aerodynamics-based methods of clinical vocal assessment. A mean absolute error of less than 10% was observed for two glottal airflow measures –maximum flow declination rate and amplitude of the modulation component– that have been associated with the pathophysiology of some common voice disorders caused by faulty and/or abusive patterns of vocal behavior (i.e., vocal hyperfunction). The proposed method further advances the ambulatory assessment of vocal function based on the neck acceleration signal, that previously have been limited to the estimation of phonation duration, loudness, and

pitch. Subglottal IBIF is also suitable for other ambulatory applications in speech communication, in which further evaluation is underway.

Index Terms

Inverse filtering; glottal airflow; ambulatory monitoring; voice production; vocal folds; accelerometer; neck vibration; glottal source

I. Introduction

Inverse filtering (IF) of voiced speech sounds is used to estimate the source of excitation at the glottis by removing the effects of the acoustic loads from an output signal. This technique is primarily performed for the vocal tract using recordings of sustained vowels obtained from either mouth-radiated acoustic pressure or airflow. A wide range of applications make use of IF, including clinical assessment of vocal function [1]–[4], speech synthesis and coding [5], [6], assessment of voice quality [7], [8], estimation of vocal tract area functions [9], speech enhancement [10], and speaker identification [11].

Most IF algorithms are aimed at removing vocal tract resonances and have included a variety of approaches such as filter banks [1], parametric estimation [12], nonlinear feedback [13], multi-channel assistance [14], iterative schemes [15], [16], model optimization [17], [18], closed-phase covariance [11], [19], spectral methods [8], weighted linear prediction [20], and hidden Markov models [6]. The primary difficulties in current speech IF are due to the effects of nonlinear source-filter interaction, inaccurate performance over a full range of fundamental frequencies, and lack of robustness for continuous speech analysis [8], [21], [22].

A substantially different IF approach consists of removing subglottal resonances and skin effects from measurements made with a surface bioacoustic sensor placed near the suprasternal notch. Subglottal (e.g., tracheal and bronchial) resonances are relatively stable and do not exhibit the same complex temporal patterns as those from the vocal tract [23]–[25]. In addition, neck surface recordings provide relatively strong signals that are less sensitive to background acoustic noise and yield unintelligible speech patterns that preserve confidentiality needed in some applications [26], [27]. These factors make a subglottal IF method attractive for the task of estimating the source of excitation at the glottis, especially in continuous speech.

Subglottal IF requires a different mathematical approach than that used for the vocal tract, as strong zeros in the neck-to-glottis transfer function are present, thus making previous vocal tract–based methods inapplicable. Subglottal IF was first attempted for sustained vowels using a parametric pole-zero representation [23] to match experimental observations of the subglottal impedance [28]. When used in the context of a vocal system model, this approach yielded better estimates of sound pressure level than the glottal parameters of interest [23], [29]. This result was attributed to the prevailing notions about the underlying physical phenomena affecting the glottal estimates. For instance, the effects of source-filter interaction were thought to distort the estimates of glottal airflow, particularly during

segments exhibiting incomplete glottal closure [23]. However, recent work illustrates that the "true" glottal airflow (i.e., glottal airflow with all the effects of source-filter interactions) could be accurately estimated without the need for modeling glottal coupling, even under an incomplete glottal closure scenario [30], [31]. This means that the acoustic coupling between tracts through the glottis is embedded in the resulting dipole source and should not be compensated for.

The concept of obtaining estimates of glottal parameters via neck surface recordings has been initially motivated by the desire to perform ambulatory assessment of vocal function for clinical and research purposes and has thus been applied to the design of portable monitoring equipment, e.g., [32]–[35]. The ultimate goal of ambulatory assessment is to obtain measures of vocal function that are essentially equivalent to what can be acquired in the clinical laboratory setting, including an important combination of acoustic and aerodynamic parameters (for details on traditional assessment, see [36]), but to do so as individuals engage in their typical daily activities. The overriding assumption is that ambulatory assessment will more accurately characterize vocal behaviors associated with voice use–related disorders (e.g., vocal hyperfunction). However, current ambulatory methods only extract estimates of fundamental frequency and sound pressure level, along with voice-use parameters (e.g., vocal dose measures) that are derived from these basic acoustic measures [32], [37]. Although still considered to have clinical value, the diagnostic capabilities of this basic battery of measures is considered to be inherently limited because it does not include aerodynamic parameters that have been shown in laboratory studies to provide important insights into the underlying pathophysiology of some of the most common types of voice disorders that are caused by faulty and/or abusive patterns of vocal behavior (i.e., vocal hyperfunction) [2], [27]. A recently developed ambulatory voice monitoring system provides recordings of the raw neck surface acceleration signal that can be used for subglottal IF purposes [33].

In this study, a biologically-inspired acoustic model builds upon initial efforts [23], [29] for an enhanced estimation of glottal airflow from neck surface acceleration. The subglottal IF scheme is evaluated under controlled scenarios that represent different quantifiable glottal configurations during sustained phonation. The objective is to address the primary challenges to typical IF, namely source-filter interaction, performance over a full range of pitch frequencies, and robustness of automatic IF for continuous speech. Although the application that is investigated in this study is associated with the ambulatory assessment of vocal function, other potential applications include biofeedback to help facilitate behavioral voice therapy for common hyperfunctional voice disorders, robust speech enhancement, speaker normalization, automatic speech recognition, and speaker identification.

The outline of this paper is as follows. Section II introduces the proposed subglottal impedance-based IF (IBIF) scheme, Section III explains the experimental setup used to test the proposed scheme, Section IV presents key results contrasting traditional IF with the subglottal IBIF method, and Section V discusses the implications of these results and future directions. Finally, Section VI provides a summary of the study.

II. Subglottal impedance-based inverse filtering

A model-based inverse filtering scheme based on mechano-acoustic analogies, transmission line principles, and physiological descriptions is proposed. The method follows a lumped impedance parameter representation in the frequency domain that has been proven useful to model sound propagation in the subglottal tract [38]–[41] and vocal tract [42]–[45] and to explore source-filter interactions [46]–[49]. Time-domain techniques such as the wave reflection analog [50]–[52] and loop-equations [53]–[55] can accomplish similar goals in numerical simulations, but are typically less attractive for the development of an on-board signal processing scheme. Although the proposed method can incorporate multiple levels of system interactions, the present study is focused on the subglottal tract.

A. General principles

The proposed scheme is based on a series of concatenated T-equivalent segments of lumped acoustic elements that relate acoustic pressure $P(\omega)$ to airflow volume velocity $U(\omega)$ [38], [41]–[44], where ω represents frequency. A representation of the basic T section is depicted in Fig. 1. Yielding wall parameters include cartilage (L_{wc}, R_{wc}, C_{wc}) and soft tissue (L_{ws}, R_{ws}, C_{ws}) components. The remaining lumped acoustic elements in Fig. 1 describe the relatively standard acoustical representations for air viscosity and heat conduction losses (R_a, G_a), elasticity (C_a), and inertia (L_a).

In this representation, a cascade connection is used to account for the acoustic transmission matrix associated with each section represented by the two-port T network. This approach provides relations for both $P(\omega)$ and $U(\omega)$, so that the frequency response of the flow-flow transfer function $H(\omega)$ and driving-point input impedance $Z(\omega)$ can be computed. The notation for frequency dependence (ω) is omitted in some cases for simplicity. If the equivalent impedance of the shunt terms in Fig. 1(a) is denoted as Z_b, and that of the series term on each side as Z_a, then the symmetric transmission matrix that relates two neighboring T sections has the following ABCD network structure:

$$\begin{bmatrix} P_1(\omega) \\ U_1(\omega) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P_2(\omega) \\ -U_2(\omega) \end{bmatrix}, \quad (1)$$

where both flows are considered to enter the T section, so that

$$A = (Z_a + Z_b)Z_b^{-1}, \quad (2)$$

$$B = (Z_a + 2Z_b)Z_a Z_b^{-1}, \quad (3)$$

$$C = Z_b^{-1}, \quad (4)$$

$$D = A. \quad (5)$$

Thus, the frequency response of the flow transfer function $H(\omega) = -U_2/U_1$ is given by

$$H(\omega) = \frac{1}{CZ_2(\omega)+D} \quad (6)$$

and the driving-point impedance, or input impedance, $Z_1(\omega)$ from the first section by

$$Z_1(\omega) = \frac{AZ_2(\omega)+B}{CZ_2(\omega)+D}, \quad (7)$$

where $Z_2(\omega)$ acts as the effective load impedance for the two-port network. As either cascade or branching configurations are commonly encountered in the subglottal tract, the network is solved by carrying through the equivalent driving-point impedance of previous tracts, starting with a terminal lung impedance and ending at the glottis. This allows for the inclusion of subglottal branching without increasing the complexity of the overall approach [41], [55].

B. Subglottal tract module and its transfer function

It is desired to estimate glottal airflow from measurements of neck surface acceleration on the skin overlying the extrathoracic trachea at the level of the suprasternal notch. For this purpose, the transmission line model is used as a tool to obtain a transfer function between neck surface acceleration and glottal airflow. Therefore, a neck skin impedance (Z_{skin}) is introduced to account for the combined effects of mechanical skin properties and accelerometer loading and to incorporate the neck skin velocity (U_{skin}) and its time derivative ($\dot{U}_{skin}$) into the model.

This module takes as input an acceleration signal $\dot{U}_{skin}$ and as output the airflow U_{sub} just below the glottis. The glottal volume velocity (GVV) is thus defined as the negative of the subglottal airflow (i.e., $-U_{sub}$). The frequency response of the transfer function $T_{skin} = U_{skin}/U_{sub}$ is obtained through the subglottal tract module and then inverted to estimate the glottal airflow from neck surface acceleration.

The subglottal tract module and its physiological basis are illustrated in Figs. 2 and 3, respectively. The subglottal tract is composed of two subglottal sections, $sub1$ and $sub2$, that represent the portion of the extrathoracic trachea above and below the accelerometer location, respectively. The neck skin impedance Z_{skin} is located in parallel between sections $sub1$ and $sub2$. In addition, Z_{sub1} and Z_{sub2} are the frequency-dependent driving-point impedances of the corresponding tract subsections. The volume velocity U_{skin} flowing through Z_{skin} is expressed as

$$U_{skin}(\omega) = U_{sub1} \frac{Z_{sub2}}{Z_{sub2}+Z_{skin}}, \quad (8)$$

where Z_{skin} is determined as the mechanical impedance of the skin (Z_m) in series with the radiation impedance due to the accelerometer loading (Z_{rad}). Thus,

$$Z_{skin}(\omega) = Z_m(\omega) + Z_{rad}(\omega), \quad (9)$$

$$Z_m(\omega) = R_m + j(\omega M_m - \frac{K_m}{\omega}), \quad (10)$$

where R_m, M_m, and K_m are, respectively, the per-unit-area resistance, inertance, and stiffness of the skin. In addition, the radiation impedance due to accelerometer loading is

$$Z_{rad}(\omega) = \frac{j\omega M_{acc}}{A_{acc}}, \quad (11)$$

where M_{acc} and A_{acc} are, respectively, the per-unit-area mass and surface of the accelerometer and any coating or mounting disk attached to it.

The skin volume velocity U_{skin} is differentiated to obtain the neck surface acceleration signal $\dot{U}_{skin}$. The frequency response T_{skin} of the transfer function between U_{sub} and $\dot{U}_{skin}$ is expressed as:

$$T_{skin}(\omega) = \frac{\dot{U}_{skin}}{U_{sub}} = \frac{H_{sub1} \cdot Z_{sub2} \cdot H_d}{Z_{sub2} + Z_{skin}}, \quad (12)$$

where $H_{sub1} = U_{sub1}/U_{sub}$ is the frequency response of the transfer function of the subglottal section $sub1$ from the glottis to the accelerometer location, and $H_d = j\omega$ is the ideal derivative filter. It is convenient to estimate directly U_{supra}, the airflow entering the vocal tract, which is derived from the subglottal airflow using $U_{supra} = -U_{sub}$. The linear magnitude of T_{skin} for one of the subjects in this study is shown in Fig. 4.

The inverse filtering process was performed digitally in the frequency domain using the fast Fourier transform (FFT) and the inverse FFT. Thus, the frequency response $T_{skin}(\omega)$ was sampled and became $T_{skin}(k)$ with $k = 0, 1, \ldots, N-1$, where N is the number of FFT points. Reconstruction with a real-valued output was achieved by setting N to be at least the number of samples in $\dot{U}_{skin}$ and forcing $T_{skin}(k)$ to be conjugate symmetric, i.e., $T_{skin}(k) = T_{skin}^*(N-k)$.

Estimation of the airflow entering the vocal tract requires inverting the subglottal frequency response (i.e., $U_{supra} = -\dot{U}_{skin}/T_{skin}$). Note that the subglottal impedance Z_{sub2} has a zero on the imaginary axis in the s-plane at 0 Hz that yields large amplitudes in the inverse transfer function of T_{skin} at low frequencies, which in turn drifts the neck surface acceleration signal significantly. Thus, to avoid this undesirable effect, the magnitude $|T_{skin}(k)|$ must be constrained to be larger than or equal to one in the low-frequency range. This could be performed by either directly reassigning the magnitude of the corresponding FFT bins (with zero phase, if desired) to a unitary gain, or utilizing a properly designed high-pass filter. In this study, a low-frequency correction with a unitary magnitude and zero phase was used for this purpose.

C. Subject-specific calibration

A calibration scheme was implemented to account for variations among subjects by assigning subject-specific parameters to the subglottal tract model. Specifically, the parameter space given by the skin model and tracheal geometry was adjusted to minimize the error between oral-based and neck-based GVV waveforms. These glottal waveforms were aligned, and model properties were obtained via a constrained multivariate optimization. Five different model parameters were adjusted in this process: tracheal length $L_{trachea}$, accelerometer position L_{sub1}, skin inertance M_m, skin resistance R_m, and skin stiffness K_m. A default parameter set was allowed to vary using a vector of scaling factors $\mathbf{Q}$ = $\{Q_i\}_{i=1, ..., 5}$. These scaling factors defined the subject specific parameters and were sought within the multivariate optimization scheme. The scaling factors are defined as

$$\mathbf{Q}=\{Q_1, Q_2, Q_3, Q_4, Q_5\}, \quad (13)$$

$$R_m = 2320 \cdot Q_1 \quad [\mathrm{g} \cdot \mathrm{s}^{-1} \cdot \mathrm{cm}^{-2}], \quad (14)$$

$$M_m = 2.4 \cdot Q_2 \quad [g \cdot \mathrm{cm}^{-2}], \quad (15)$$

$$K_m = 491\,000 \cdot Q_3 \quad [\mathrm{dyn} \cdot \mathrm{cm}^{-3}], \quad (16)$$

$$L_{trachea} = 10 \cdot Q_4 \quad [\mathrm{cm}], \quad (17)$$

$$L_{sub1} = 5 \cdot Q_5 \quad [\mathrm{cm}]. \quad (18)$$

Calibration of the model can be expressed in the form of a standard optimization problem, as described in Eq. 19, where the solution $\mathbf{Q}^*$ yields the subject-specific parameter values. In this study, the fitness function $f(\cdot) : \mathbb{R}^5 \to \mathbb{R}$ minimized the mean-square error incorporating GVV signals and waveform features that are described further in Section III. Thus,

$$\mathbf{Q}^* = \arg \min_{\mathbf{Q}} f(\mathbf{Q}), \text{ subject to } \mathbf{Q} \in \mathbf{D}, \quad (19)$$

where $\mathbf{D} = \{D_i\}_{i=1, ..., 5}$ is the constricted search domain and each D_i defines the subrange of the corresponding Q_i. This search domain was used to avoid model overfitting and to keep parameter values physiologically meaningful [28], [57]–[59]. Each scaling factor was restricted to yield parameter values within the following ranges:

$$D_i = \begin{cases} [\ 0.1, & 10\ ], & i=1, 2, 3, \\ [\ 0.8, & 1.2\ ], & i=4, 5. \end{cases} \quad (20)$$

While the proposed calibration does not require a specific multivariate optimization scheme to compute $\mathbf{Q}^*$ or IF mechanism to obtain the oral-based GVV, recommendations in this regard are presented in Section III.

Default mechanical properties for the neck skin were taken from early studies [57], [60]. Mechanical properties for the accelerometer loading were based on the accelerometer described in Section III, with a mass per unit area of $M_{acc}/A_{acc} = 0.26 \text{ g} \cdot \text{cm}^{-2}$. The length of the trachea was considered to be the primary anatomical difference between subjects in the lower airways [59], and tracheal scaling has been used to correct anatomical differences when modeling subglottal impedances [28]. The default tracheal length was 10 cm [58]. The default accelerometer location was 5 cm below the glottis. The acoustical transmission line model of a symmetric branching subglottal representation from previous studies [40], [41] was used. In particular, symmetric anatomical descriptions for an average male were used [58], since they yielded a response similar to that reported experimentally [28], [55].

Note that the selected mechanical properties of the skin likely capture other anatomical differences (e.g., skin fat) and are assumed to be constant for each subject. Some of these skin properties may account for other components in the system, such as tracheal diameter and losses in the subglottal system, for which their resulting values need to be carefully interpreted. Given the assumption of constant skin properties, this calibration protocol is expected to be performed once for each subject. However, changes in head position and other factors affecting the skin properties were re-calibrated in order to reduce initial uncertainties. The robustness of the skin parameter calibration is a topic of current research.

III. Experimental Methods

An experimental evaluation was used to evaluate the sub-glottal IBIF scheme with a protocol that included synchronous measurements of neck skin surface acceleration (ACC), oral airflow volume velocity (OVV), electroglottography (EGG), and radiated acoustic pressure (MIC). The primary goal was to obtain estimates of GVV from the ACC signal and contrast them with those estimated from the OVV signal. The remaining signals were used as references to facilitate the process.

The OVV waveform was obtained through a circumferentially-vented mask (model MA-1L, Glottal Enterprises). Calibration of the OVV signal used an airflow calibration unit (model MCU-4, Glottal Enterprises) after each recording session. The ACC signal was obtained using a lightweight accelerometer (model BU-27135, Knowles) attached to the skin overlying the suprasternal notch using double sided-tape (No. 2181, 3M), as seen in Fig. 5. The accelerometer at this location provides good tissue-borne sensitivity and is minimally affected by typical background noise [26]. The accelerometer was calibrated using a laser vibrometer as described in [23], where a sensitivity of 90 dB $\pm$ 3 dB re 1 cm/s^2/volt in the 70–3000 Hz range was found. As in that study, the nominal value of 88.121 dB re 1 cm/s^2/volt was used for the calibration. A custom signal conditioner supplied power to the ACC, removed its DC component, and provided a gain of 1.2 dB $\pm$ 0.1 dB [23].

The experimental protocol included two sustained vowels (/a/ and /i/) at different vocal conditions (modal, modal-loud, breathy, falsetto). Amplitude of vibration, degree of

incomplete glottal closure, vibratory mode, and fundamental frequency were expected to change among these cases [61], [62]. The goal was to explore various glottal conditions (different fundamental frequencies, amplitude levels, and degrees of source-filter interaction) and amplitude levels to evaluate the linearity and performance of the subglottal IBIF approach. Seven adult subjects (four males and three females) with no history of vocal pathologies yielded 33 recordings for this study. Table I lists the characteristics of the recordings made from the subjects and indicates which recordings were used for subject-specific calibration (CAL column). All of the recordings were performed in an acoustically insulated and absorptive recording booth at the Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital, Boston, MA.

The scheme's performance was evaluated in two ways. First, the root mean-square error (RMSE) metric was used on the time–aligned GVV signal and its time derivative (dGVV). The derivative of the airflow signal it is of interest since clinically-significant parameters are derived from it, while also preserving the same second-order nature as the acceleration signal. Second, measures of glottal behavior were selected to explore the ability of the approach to correctly estimate key characteristics of the glottal source, as proposed by [19]. The selected source-related measures included the difference between the first two harmonics (H1–H2), harmonic richness factor (HRF), peak-to-peak amplitude of the unsteady glottal airflow (AC flow), and maximum flow declination rate (MFDR). The harmonic measures (H1–H2 and HRF) are associated with the spectral slope of the glottal source, AC flow is related to the glottal area, and MFDR is derived from the flow derivative and linked to collision forces in the vocal folds. For more details on these measures, refer to [2], [19], [63]. Each of these measures was taken from at least 20 successive cycles for each recording. Estimation errors are presented with respect to the parameters estimated from the oral-based GVV signal in terms of absolute magnitude differences and percentages with respect to mean parameter values. The percent error calculation removes the dependence on physical units for each parameter and facilitates error interpretation across parameters.

A particle swarm optimization scheme [64] was used for the constrained optimization of the model parameters with a fitness function given by the mean-square error of the aforementioned measures (with equal weighting) between the oral-based and acceleration-based GVV signals. A combination of all measures in the fitness function provided a robust estimate that minimized model overfitting (when using RMSE measures only) and signal distortion (when using glottal measures only). Calibration was performed once on each subject, unless changes in head position or other factors affecting the skin model were identified to require an additional calibration. For the recordings considered in this study, 10 out of 33 recordings were used for subject-specific model calibration (CAL column of Table I).

The oral-based GVV was obtained using closed-phase inverse filtering (CPIF) [19] from the OVV signal, with the aid of the EGG signal to identify instants of glottal closure. CPIF is one of the most accepted vocal tract IF schemes to estimate glottal airflow from oral airflow or radiated acoustic pressure, particularly for applications related to the assessment of vocal function [19], [22], [65]. Although having a smaller bandwidth (0–4 kHz) than a microphone signal, the OVV signal provides advantages and unique insight for the clinical

assessment of vocal function since it preserves the DC characteristic of the airflow and is less affected by the acoustic environment [1].

IV. Results

Estimates of GVV and its time derivative (dGVV) were obtained from the ACC signal and compared with those acquired by inverse filtering the OVV signal. Examples of raw waveforms are presented in Fig. 6 for three different vocal conditions (normal, loud, and breathy). It is noted from Fig. 6 that the ACC-based waveforms were very similar to the OVV-based ones, with an error that was lower during the open phase portion of the cycle for all cases. This is an initial indication that the scheme is capable of retrieving key components of the glottal source.

The measures extracted for all cases and subjects under evaluation are presented in Table I. It was observed that for the modal vowel /a/, the measures were within the range for male and female subjects expected from previous studies [61], [62]. The vowel /i/ has not been consistently studied and thus has no previous reference for comparison.

The absolute differences between the OVV- and ACC-based estimates for each parameter in Table I are shown in the third column of Table II. The relative (percent) errors of these differences with respect to the values from the oral flow–based GVV signal are shown in the fourth column of Table II. For the non-harmonic measures (AC Flow and MFDR), the error and its variations are considered sufficiently low (a mean percent error less than 10%) to allow for the proposed scheme to be considered for clinical use. Significantly higher than normal values for these non-harmonic measures have been associated with vocal hyperfunction (e.g., by increments larger than 200% with respect to the normal population) [2], [31], [66]. Thus, the subglottal IBIF method produces estimates for these two ACC-based measures that are considered to be accurate enough for use in assessing the pathophysiology of common hyperfunctional voice disorders. The performance of the subglottal IBIF scheme for other applications is a topic of current research.

More details on the performance of the subglottal IBIF algorithm can be obtained from the scatter plots shown in Fig. 7. As noted before, the AC Flow and MFDR measures are well represented with the proposed scheme, as they exhibit a near 1:1 relation with those from the inverse-filtered OVV signal. The variance of the estimates across different amplitudes is smaller for the AC Flow than the variance of the MFDR estimates, indicating that the former has a more robust and linear behavior. On the other hand, the harmonic measures (H1–H2 and HRF) exhibit a larger spread around the 1:1 line and across different values, indicating that these measure estimates were consistently less accurate. The coefficients of determination R^2 reported in Fig. 7 quantify these trends, meaning that they are higher for the non-harmonic measures.

Similar comparisons between MFDR obtained via ACC and OVV as those shown in Fig. 7b and Table II were obtained in [23], [29]. The best possible scenario in those studies yielded $R^2 = 0.86$ (using a logarithmic scale re 1 L/s^2), and a mean error of 1.5 dB $\pm$ 3.7 dB (i.e., percent error of 19 % $\pm$ 53 pp, respectively). In contrast, MFDR estimation by the subglottal IBIF scheme exhibits $R^2 = 0.97$ with a mean absolute error of 9.3% $\pm$ 7.5 pp, indicating that

the proposed scheme provides a significant enhancement in accuracy and robustness for the MFDR estimates. Previous studies did not consider estimates of AC Flow or harmonic measures, preventing prior comparison.

The large error observed for the RMSE values and harmonic measures (H1–H2 and HRF) may be attributable to the more important differences observed during the closed phase portion of the cycle, which has an effect on the spectral tilt estimation. This is another indication that the scheme captures the temporal structure of the open phase portion of the glottal cycle better than it does during the closed-phase. In addition, the fitness function used to optimize the model parameters was more sensitive to the RMSE, and thus indirectly yielded better non-harmonic estimates. Although this behavior could be altered simply by modifying the fitness function weighting, this does not appear to be critical at this point since the most clinically relevant parameters are sufficiently well represented.

V. Discussion

The subglottal IBIF scheme provided a simple and accurate method to estimate glottal airflow and aerodynamic parameters from a neck skin acceleration signal. The scheme yielded comparable estimates with respect to the current criterion standard obtained from inverse filtering the oral airflow, although being indirect aerodynamic estimates with no DC component. The proposed method was capable of more precisely retrieving the signal structure during the open portion of the glottal cycle. Thus, those measures that are extracted during the open portion of the glottal waveform (e.g., AC Flow and MFDR) were more accurately estimated. The capability of retrieving these particular measures provides evidence for the scheme's potential in clinical applications that seek to assess the pathophysiology associated with common hyperfunctional voice disorders [2]. The subglottal IBIF estimates are not expected to surpass the capabilities of traditional oral airflow–based estimation techniques, but they may complement them especially providing a simple and robust manner in which to obtain glottal airflow estimates for ambulatory voice monitoring applications.

Comparing the results of this study with previous efforts [23], [29] elucidated the benefits of incorporating a transmission line model of the complete subglottal system. This description was believed to more accurately account for the effects of accelerometer placement and subject-specific characterizations. The addition of a skin compliance and a direct calibration with the inverse-filtered oral airflow via waveform matching allowed for an improved estimation of glottal parameters. However, this study entailed a relatively favorable scenario since subjects remained still during the recordings and any potential change in the skin model (e.g., due to head movements) was compensated for. Current research efforts are devoted to explore the robustness of the skin parameters under ambulatory conditions for long-term recordings.

Although the results presented are promising for sustained vowels, further investigations are needed to assess the method on running speech. The key challenges are to evaluate the effects of possible changes of the skin properties due to neck movements, temporal variations in the subglottal resonances due to laryngeal height changes, and effects of co-

articulation on the glottal source. Topics for investigation include quantifying and potentially defining an optimization scheme to account for this variability. Other components for future investigation include evaluating the subglottal IBIF module for pathological cases and determining whether measures believed to reflect vocal hyperfunction are still viable and valid when collected during ambulatory assessment. Efforts to implement the sub-glottal IBIF scheme and address these research problems using an ambulatory voice monitoring system in a large cohort of normal and voice-disordered subjects are currently underway [33]. Implementing the proposed subglottal IBIF algorithm in an ambulatory voice monitoring system should not only enhance clinical evaluation of vocal function, but it may make possible new types of biofeedback that can also be used in ambulatory/wearable systems to facilitate/enhance behavioral treatment of commonly occurring voice disorders.

VI. Conclusions

A model-based inverse filtering approach was used to estimate glottal airflow and glottal parameters from non-invasive measurements of neck surface acceleration. The scheme uses a subglottal approach, wherein resonances are relatively stable and do not exhibit the same complex temporal patterns observed for the vocal tract. In an evaluation using sustained vowels, the results indicate that accurate estimates of glottal airflow and related parameters can be obtained. The scheme surpassed the accuracy of previous efforts and provided reliable estimates of salient clinical measures that have shown sensitivity to common hyperfunctional voice disorders. The proposed subglottal inverse filtering scheme is potentially suitable not only for ambulatory assessment, but also for use in biofeedback approaches that seek to facilitate and enhance behavioral voice therapy and other ambulatory applications in speech communication where robust estimation of the source of excitation is desired.

Acknowledgments

Biographies

Matías Zañartu (S'08–M'11) received the Ph.D. and M.S. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2010 and 2006,

respectively, and his professional title and B.S. degree in acoustical engineering from Universidad Tecnológica Vicente Pérez Rosales, Santiago, Chile, in 1999 and 1996, respectively.

He is currently an Academic Research Associate at the Department of Electronic Engineering from Universidad Técnica Federico Santa Maria, Val-paraíso, Chile. His research interests include digital signal processing, nonlinear dynamic systems, acoustic modeling, speech/audio/biomedical signal processing, speech recognition, and acoustic biosensors.

Dr. Zañartu was the recipient of a Fulbright Scholarship, an Institute of International Education IIE-Barsa Scholarship, a Qualcomm Q Award of Excellence, and the Best Student Paper in Speech Communication in the 157th meeting of the Acoustical Society of America.

Julio C. Ho received the Ph.D degree in biomedical engineering and the B.S. degree (with distinction) in electrical and computer engineering from Purdue University, West Lafayette, IN, in 2011 and 2006, respectively.

Dr. Ho was the recipient of the A. F. Welch Memorial Scholarship in 2003 and 2004, the Kimberly-Clark Industrial Affiliates Scholarship in 2005, and the Schlumberger Industrial Affiliates Scholarship in 2006. His research interests include biomedical acoustics and biomedical signal and image processing.

Daryush D. Mehta (S'01–M'11) received the B.S. degree in electrical engineering (summa cum laude) from University of Florida, Gainesville, in 2003, the S.M. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 2006, and the Ph.D. degree from MIT in speech and hearing bioscience

and technology in the Harvard–MIT Division of Health Sciences and Technology, Cambridge, in 2010.

He currently holds appointments at Harvard University (Research Associate in the School of Engineering and Applied Sciences), Massachusetts General Hospital (Assistant Biomedical Engineer in the Department of Surgery), and Harvard Medical School (Instructor in Surgery), Boston. He is also an Honorary Senior Fellow in the Department of Otolaryngology, University of Melbourne, in Australia.

Robert E. Hillman received the B.S. and M.S. degrees in speech pathology from Pennsylvania State University, University Park, in 1974 and 1975, respectively, and the Ph.D. degree in speech science from Purdue University, West Lafayette, IN, in 1980.

He is currently Co-Director/Research Director of the MGH Center for Laryngeal Surgery and Voice Rehabilitation, Associate Professor of Surgery and Health Sciences and Technology at Harvard Medical School, and Professor and Associate Provost for Research at the MGH Institute of Health Professions, Boston, MA. His research has been funded by both governmental and private agencies since 1981, and he has over 100 publications on normal and disordered voice.

Prof. Hillman is a Fellow of the American Speech-Language-Hearing Association (also receiving Honors of the Association, ASHA's highest honor) and the American Laryngological Association.

George R. Wodicka (S'81–M'82–SM'93–F'02) received the B.E.S. degree in biomedical engineering from Johns Hopkins University, Baltimore, MD, in 1982, and the S.M. degree in electrical engineering and computer science and the Ph.D. degree in medical engineering from Massachusetts Institute of Technology, Cambridge, MA, in 1985 and 1989, respectively.

He is currently a Professor and the Head of the Weldon School of Biomedical Engineering and a Professor of electrical and computer engineering at Purdue University, West Lafayette, IN. His current research interests include biomedical acoustics, acoustical modeling, and signal processing.

Prof. Wodicka was the recipient of the National Science Foundation Young Investigator Award. He is a Fellow of the Institute of Electrical and Electronics Engineers, the American Institute for Medical and Biological Engineering, and a Guggenheim Fellow.

References

1. Rothenberg M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J Acoust Soc Am. 1973; 53(6):1632–1645. [PubMed: 4719255]

2. Hillman RE, Holmberg EB, Perkell JS, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: An experimental framework and initial results. J Speech Hear Res. 1989; 32(2):373–392. [PubMed: 2739390]

3. Perkell JS, Holmberg EB, Hillman RE. A system for signal processing and data extraction from aerodynamic, acoustic, and electroglottographic signals in the study of voice production. J Acoust Soc Am. 1991; 89(4):1777–1781. [PubMed: 2045586]

4. Mehta DD, Hillman RE. Use of aerodynamic measures in clinical voice assessment. Perspectives on Voice and Voice Disorders. 2007; 17(3):14–18.

5. Childers DG, Hu HT. Speech synthesis by glottal excited linear prediction. J Acoust Soc Am. 1994; 96(4):2026–2036. [PubMed: 7963019]

6. Raitio T, Suni A, Yamagishi J, Pulakka H, Nurminen J, Vainio M, Alku P. HMM-based speech synthesis utilizing glottal inverse filtering. IEEE Trans Audio Speech Lang Process. 2011; 19(1):153–165.

7. Titze IR, Sundberg J. Vocal intensity in speakers and singers. J Acoust Soc Am. 1992; 91(5):2936–2946. [PubMed: 1629486]

8. Arroabarren I, Carlosena A. Inverse filtering in singing voice: a critical analysis. IEEE Trans Audio Speech Lang Process. Jul; 2006 14(4):1422–1431.

9. Deng H, Ward R, Beddoes M, Hodgson M. A new method for obtaining accurate estimates of vocal-tract filters and glottal waves from vowel sounds. IEEE Trans Audio Speech Lang Process. 2006; 14(2):445–455.

10. Quatieri TF, McAulay RJ. Shape invariant of time-scale and pitch modification of speech. IEEE Trans Audio Speech Lang Process. 1992; 40(3):497–510.

11. Plumpe MD, Quatieri TF, Reynolds DA. Modeling of the glottal flow derivative waveform with application to speaker identifica-tion. IEEE Trans Speech Audio Proces. 1999; 7(5):569–586.

12. Makhoul J. Linear prediction: A tutorial review. Proc of the IEEE. 1975; 63(4):561–580.

13. Rothenberg M, Zahorian S. Nonlinear inverse filtering technique for estimating the glottal-area waveform. J Acoust Soc Am. 1977; 61(4):1063–1070. [PubMed: 864095]

14. Childers D, Chieteuk A. Modeling the glottal volume-velocity waveform for three voice types. J Acoust Soc Am. 1995; 97(1):505–519. [PubMed: 7860829]

15. Alku P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Communication. 1992; 11(2-3):109–118.

16. Alku, P. Visual representations of speech signals. Chichester, UK: Wiley; 1993. p. 139-146.ch. "Estimation of the glottal excitation of speech with pitch-synchronous iterative adaptive inverse filtering"

17. Frohlich M, Michaelis D, Strube HW. SIM–simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals. J Acoust Soc Am. 2001; 110(1):479–488. [PubMed: 11508972]

18. Fu Q, Murphy P. Robust glottal source estimation based on joint source-filter model optimization. IEEE Trans Audio Speech Lang Process. 2006; 14(2):492–501.

19. Alku P, Magi C, Yrttiaho S, Bäckström T, Story B. Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering. J Acoust Soc Am. 2009; 125(5):3289–3305. [PubMed: 19425671]

20. Kafentzis GP, Stylianou Y, Alku P. Glottal inverse filtering using stabilised weighted linear prediction. Proc IEEE ICASSP. 2011:5408–5411.

21. Walker, J.; Murphy, P. Advanced methods for glottal wave extraction. In: Faundez-Zanuy, M., editor. Nonlinear Analyses and Algorithms for Speech Processing. Springer; Berlin/Heidelberg: 2005. p. 139-149.

22. Walker, J.; Murphy, P. A review of glottal waveform analysis. In: Stylianou, Y., editor. Progress in Nonlinear Speech Processing. Springer; Berlin/Heidelberg: 2007. p. 1-21.

23. Cheyne, HA. PhD dissertation. Harvard-MIT Division of Health Sciences and Technology; 2002. Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck.

24. Lulich, SM. PhD dissertation. Harvard University–MIT Division of Health Sciences and Technology; 2006. The role of lower airway resonances in defining vowel feature contrasts.

25. Quatieri TF, Brady K, Messing D, Campbell JP, Campbell WM, Brandstein MS, Weinstein CJ, Tardelli JD, Gatewood PD. Exploiting nonacoustic sensors for speech encoding. IEEE Trans Audio Speech Lang Process. 2006; 14(2):533–544.

26. Zañartu M, Ho JC, Kraman SS, Pasterkamp H, Huber JE, Wodicka GR. Air-borne and tissue-borne sensitivities of acoustic sensors used on the skin surface. IEEE Trans Biomed Eng. 2009; 56(2):443–451. [PubMed: 19272887]

27. Hillman RE, Mehta DD. Ambulatory monitoring of disordered voices. Perspectives on Voice and Voice Disorders. 2011; 21(2):56–61.

28. Ishizaka K, Matsudaira M, Kaneko T. Input acoustic-impedance measurement of the subglottal system. J Acoust Soc Am. 1976; 60(1):190–197. [PubMed: 956526]

29. Cheyne, H. Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck. Proc. 3rd. IEEE/EMBS Int. Summer School Symp. Med. Dev. Biosens.; 2006. p. 118-121.

30. Zañartu M, Mehta DD, Ho JC, Hillman RE, Wodicka GR. An impedance-based inverse filtering scheme with glottal coupling. J Acoust Soc Am. 125(4):2009, 2638–2638.

31. Zañartu, M. PhD dissertation. School of Electrical and Computer Engineering, Purdue University; 2010. Acoustic coupling in phonation and its effect on inverse filtering of oral airflow and neck surface acceleration.

32. Cheyne HA, Hanson HM, Genereux RP, Stevens KN, Hillman RE. Development and testing of a portable vocal accumulator. J Speech Lang Hear Res. 2003; 46(6):1457–1467. [PubMed: 14700368]

33. Mehta DD, Zañartu M, Feng SW, Cheyne HA, Hillman RE. Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. IEEE Trans Biomed Eng. 2012 in press.

34. Popolo PS, Svec JG, Titze IR. Adaptation of a pocket pc for use as a wearable voice dosimeter. J Speech Lang Hear Res. 2005; 48(4):780–791. [PubMed: 16378473]

35. Lindstrom F, Ren K, Li H, Waye KP. Comparison of two methods of voice activity detection in field studies. J Speech Hear Res. 2009; 52(6):1658–1663.

36. Mehta DD, Hillman RE. Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. Curr Opin Otolaryngol Head Neck Surg. 2008; 16(3):211–215. [PubMed: 18475073]

37. Titze IR, Svec JG, Popolo PS. Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. J Speech Lang Hear Res. 2006; 46(4):919–932. [PubMed: 12959470]

38. Wodicka GR, Stevens KN, Golub HL, Cravalho EG, Shannon DC. A model of acoustic transmission in the respiratory system. IEEE Trans Biomed Eng. 1989; 36(9):925–934. [PubMed: 2777281]

39. Wodicka GR, Stevens KN, Golub HL, Shannon DC. Spectral characteristics of sound transmission in the human respiratory system. IEEE Trans Biomed Eng. 1990; 37(12):1130–1135. [PubMed: 2289787]

40. Harper, P. PhD dissertation. School of Electrical and Computer Engineering, Purdue University; 2000. Respiratory tract acoustical modeling and measurements.

41. Harper P, Kraman SS, Pasterkamp H, Wodicka GR. An acoustic model of the respiratory tract. IEEE Trans Biomed Eng. May; 2001 48(5):543–550. [PubMed: 11341528]

42. Flanagan, JL. Speech analysis; synthesis and perception. 2. New York: Springer-Verlag; 1972.

43. Sondhi MM. Model for wave propagation in a lossy vocal tract. J Acoust Soc Am. 1974; 55(5): 1070–1075. [PubMed: 4833700]

44. Stevens, KN. Acoustic phonetics. 1. Cambridge, Mass: MIT Press; 1998.

45. Story BH, Laukkanen AM, Titze IR. Acoustic impedance of an artificially lengthened and constricted vocal tract. J Voice. 2000; 14(4):455–469. [PubMed: 11130104]

46. Ishizaka K, Flanagan JL. Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Syst Tech J. 1972; J51:1233–1268.

47. Rothenberg, M. Speech Transmission Laboratory Quarterly Progress and Status Report. Vol. 22. Royal Institute of Technology; Stockholm: 1981. An interactive model for the voice source; p. 001-017.

48. Ananthapadmanabha, TV.; Fant, G. Speech Transmission Laboratory Quarterly Progress and Status Report. Vol. 23. Royal Institute of Technology; Stockholm: 1982. Calculation of true glottal flow and its components; p. 001-030.

49. Titze IR. Nonlinear source-filter coupling in phonation: Theory. J Acoust Soc Am. 2008; 123(5): 2733–2749. [PubMed: 18529191]

50. Kelly, JL.; Lochbaum, CC. Speech synthesis. Proceedings of the Fourth International Congress on Acoustics; Copenhagen. 1962. p. 1-4.

51. Liljencrants, J. PhD dissertation. Dept. of Speech Commun. and Music Acoust., Royal Inst. of Tech; Stockholm, Sweden: 1985. Speech synthesis with a reflection-type line analog.

52. Story, BH. PhD dissertation. University of Iowa; 1995. Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract.

53. Maeda S. A digital simulation method of the vocal-tract system. Speech Commun. 1982; 1(3–4): 199–229.

54. Mokhtari P, Takemoto H, Kitamura T. Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches. Speech Commun. 2008; 50(3):179–190.

55. Ho JC, Zañartu M, Wodicka GR. An anatomically-based, time-domain acoustic model of the subglottal system for speech production. J Acoust Soc Am. 2011; 129(3):1531–1547. [PubMed: 21428517]

56. Lynch, PJ. Creative Commons Attribution 2.5 License 2006. Dec. 2006 Lungs-simple diagram of lungs and trachea.

57. Ishizaka K, Flanagan JL. Direct determination of vocal-tract wall impedance. IEEE Trans Acoust Speech Sig Process. 1975; ASSP-23:370–373.

58. Weibel, ER. Morphometry of the Human Lung. 1. New York: Springer; 1963.

59. Sanchez I, Pasterkamp H. Tracheal sound spectra depend on body height. Am Rev Respir Dis. 1993; 148(4–1):1083–1087. [PubMed: 8214929]

60. Franke EK. Mechanical impedance of the surface of the human body. J Appl Physiol. 1951; 3(10): 582–590. [PubMed: 14824041]

61. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. J Acoust Soc Am. 1988; 84(2):511–529. [PubMed: 3170944]

62. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. J Voice. 1989; 3:294–305.

63. Hillman RE, Heaton JT, Masaki A, Zeitels SM, Cheyne HA. Ambulatory monitoring of disordered voices. Ann Otol Rhinol Laryngol. 2006; 115(11):795–801. [PubMed: 17165660]

64. Poli R, Kennedy J, Blackwell T. Particle swarm optimization: An overview. Swarm Intelligence. 2007; 1:33–57.

65. Quatieri, TF. Discrete-time speech signal processing: Principles and practice. Upper Saddle River, NJ: Prentice Hall; 2002. ser. Prentice-Hall signal processing series

66. Holmberg EB, Doyle P, Perkell JS, Hammarberg B, Hillman RE. Aerodynamic and acoustic voice measurements of patients with vocal nodules–variation in baseline and changes across voice therapy. J Voice. 2003; 17(3):269–282. [PubMed: 14513951]
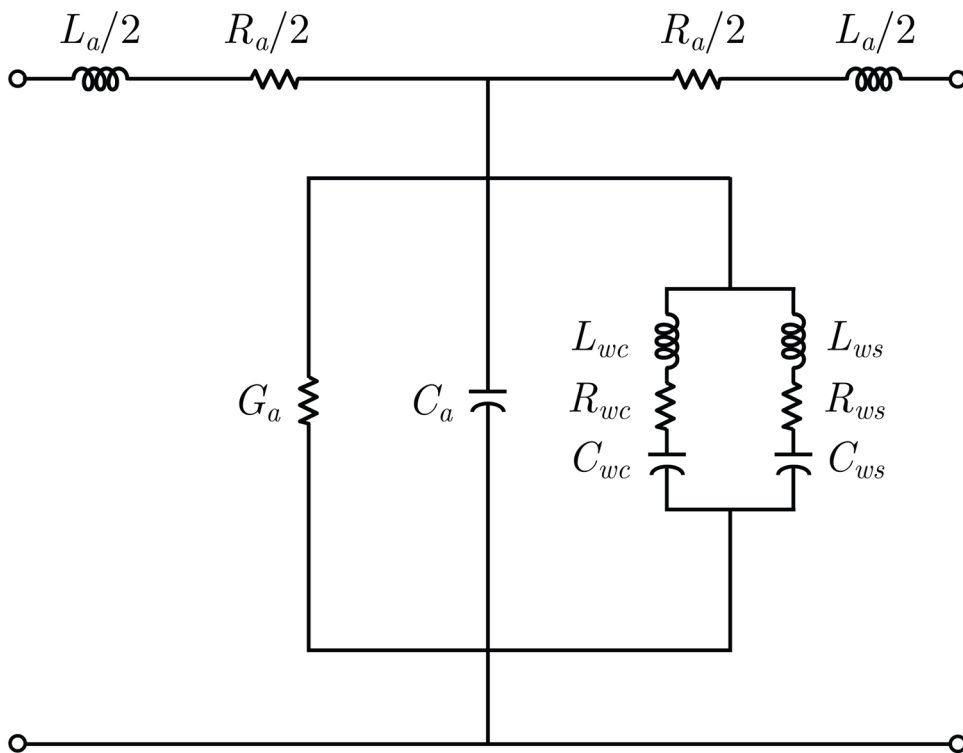
(a)
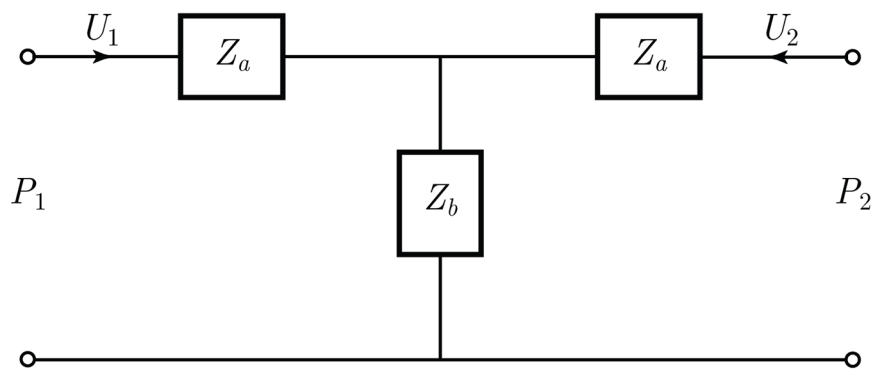
(b)

Fig. 1.
Representation of the T network used for the subglottal model. (a) Acoustical representation of losses, elasticity, inertia, and yielding walls. (b) Simplified two-port symmetric representation used to compute the ABCD transmission line parameters.
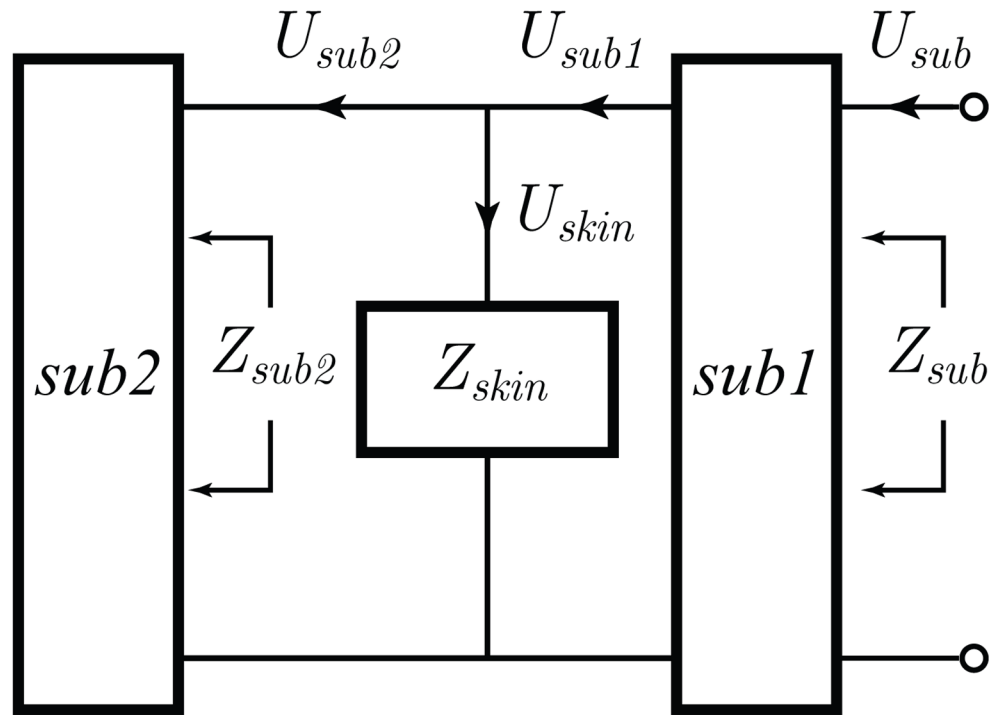
Fig. 2.
Representation of the subglottal IBIF module. An electrical representation of the neck skin and accelerometer loading was included in Z_{skin}. See Fig. 3 for anatomical details.
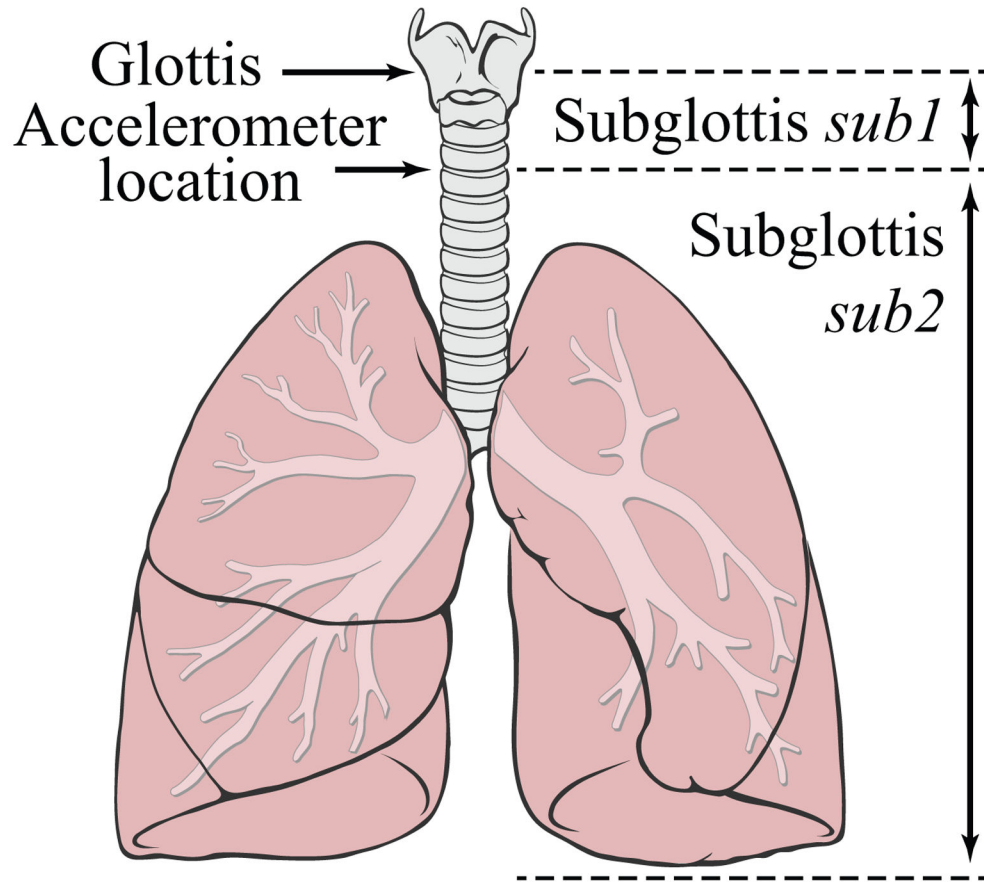
Fig. 3.
Representation of the subglottal system. The accelerometer is placed on the skin surface overlying the suprasternal notch at approximately 5 cm below the glottis. The tracts above and below this location are labeled *sub*1 and *sub*2, respectively. Figure adapted from [56].
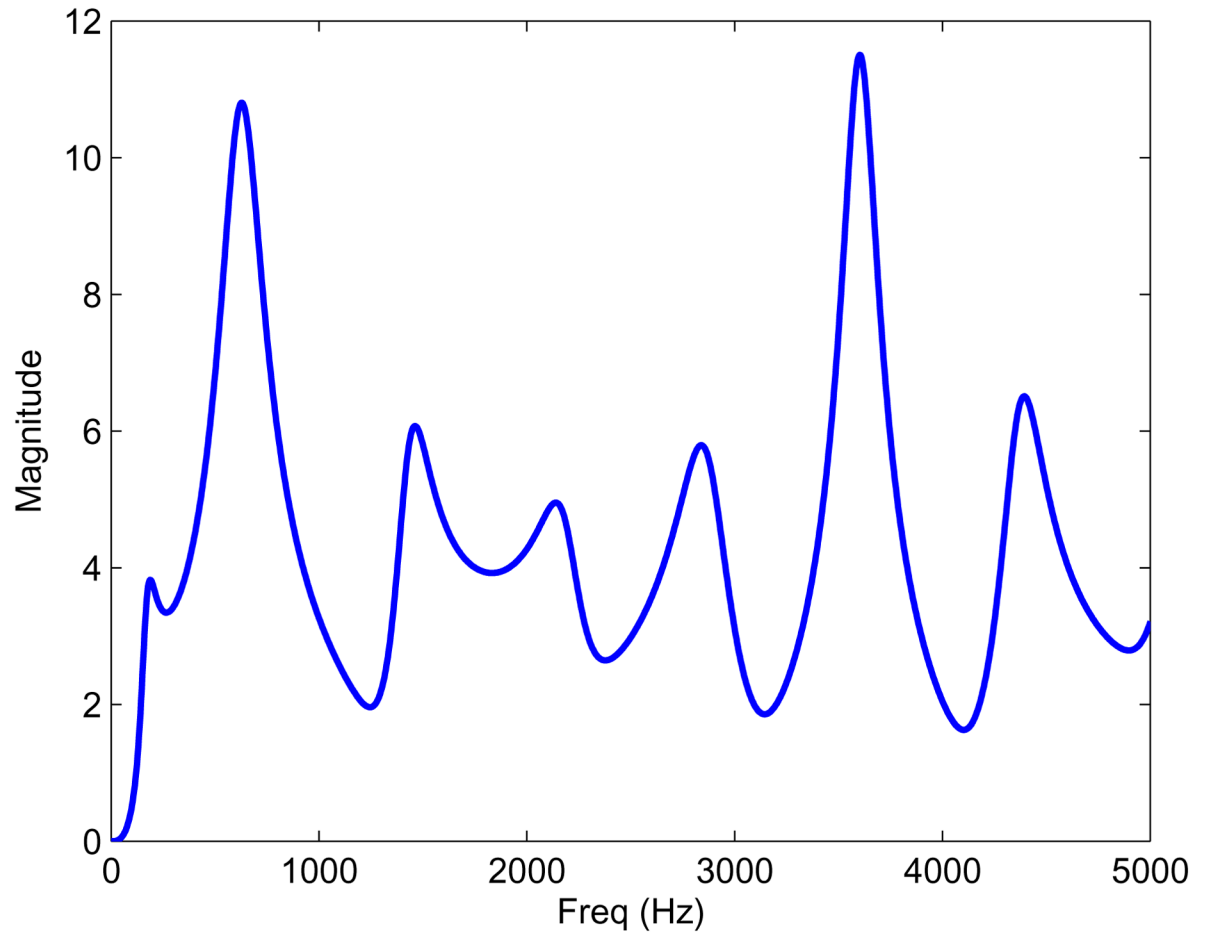
Fig. 4.
Example of the linear magnitude of the frequency response $T_{skin}(\omega)$ obtained from the subglottal module.

Fig. 5.
(Color online) The accelerometer sensor mounted subglottally on the neck skin with double-sided tape.
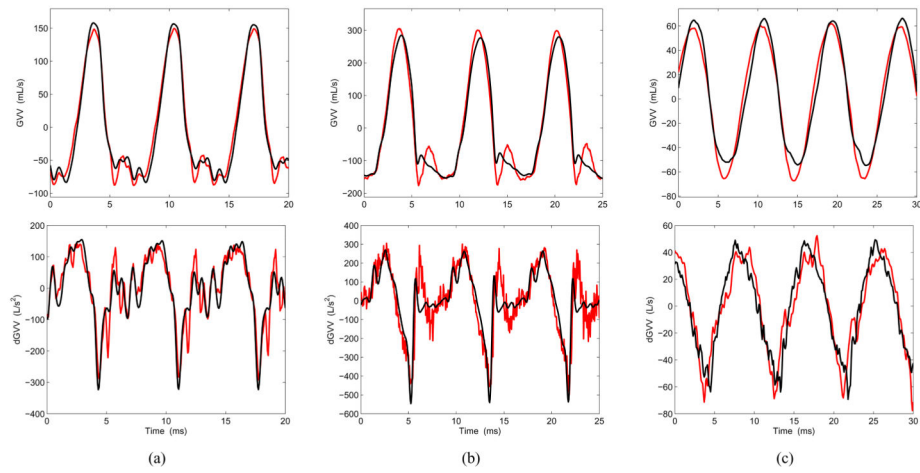
Fig. 6.
(Color online) Examples of estimates of glottal airflow (GVV, top) and flow derivative (dGVV, bottom) for three vocal conditions: (a) modal, (b) modal-loud, (c) breathy. The estimates obtained from OVV are shown in black and those from ACC are shown in red. The selected cases are taken from Table I (denoted with an asterisk).

Fig. 7.
(Color online) Scatter plots contrasting measures obtained from ACC (y axis) vs. OVV (x axis) for each condition shown in Table I: (a) AC Flow, (b) MFDR, (c) H1–H2, (d) HRF. The 1:1 lines are shown in black and the coefficient of determination R^2 for each relation is reported.

TABLE I

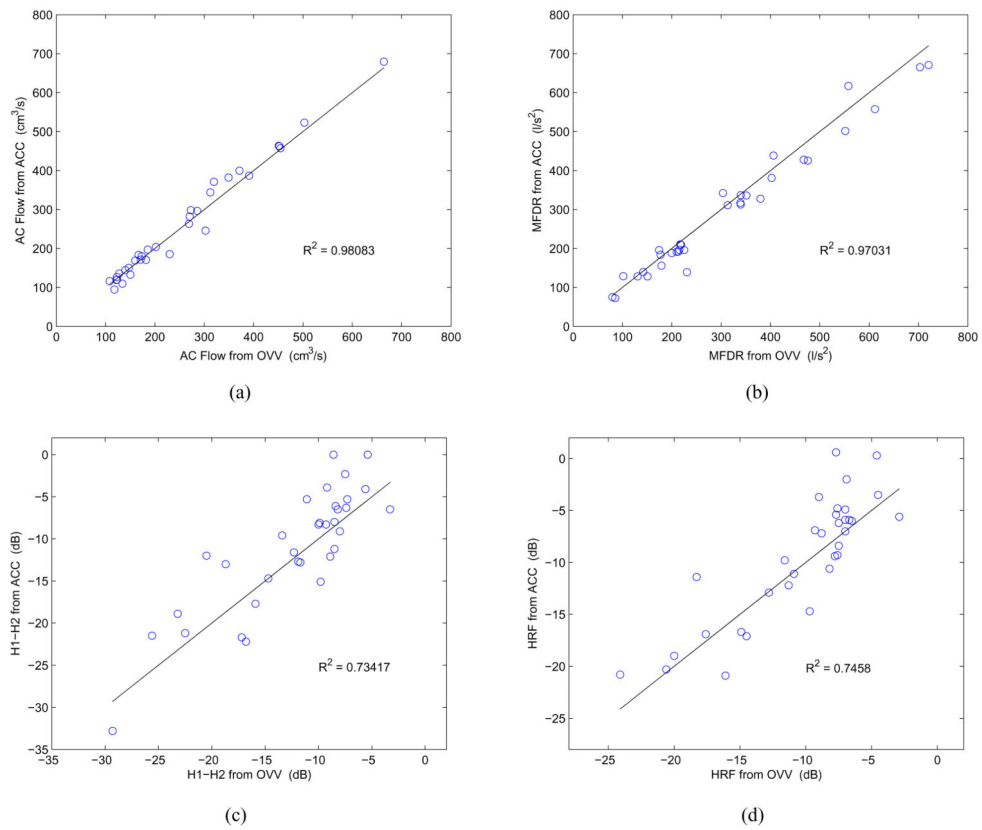Measures of glottal behavior extracted from GVV estimates obtained from OVV and ACC signals. SID indicates the subject ID number and CAL is a flag that indicates if calibration of skin parameters was performed. Source-related measures are the difference between the first two harmonics (H1–H2), harmonic richness factor (HRF), maximum flow declination rate (MFDR), and peak-to-peak amplitude of the unsteady glottal airflow (AC flow). The fundamental frequency (F0) was observed to be the same for the OVV and ACC signals, so only the latter is presented.

SID	CAL	Voice Quality	Vowel	F0 ACC (Hz)	H1–H2 (dB) OVV	H1–H2 (dB) ACC	HRF (dB) OVV	HRF (dB) ACC	MFDR (L/s²) OVV	MFDR (L/s²) ACC	AC Flow (mL/s) OVV	AC Flow (mL/s) ACC
M1	Y	Modal	/a/	114.7	12.3	11.6	−11.6	−9.8	102.0	129.0	147.2	150.4
	N	Modal	/i/	117.2	9.9	8.1	−9.3	−6.9	213.6	192.1	229.7	185.5
	N	Breathy	/a/	119.6	22.5	21.2	−20.6	−20.3	85.4	72.5	150.3	132.8
	N	Breathy*	/i/	117.2	17.2	21.7	−16.1	−20.9	79.7	75.5	127.8	135.9
	N	Modal-loud	/a/	102.5	8.9	12.1	−8.2	−10.6	351.2	336.2	312.3	343.8
	N	Modal-loud	/i/	107.4	11.7	12.8	−10.9	−11.1	174.5	195.8	269.1	263.5
	Y	Falsetto	/a/	227.1	15.9	17.7	−14.5	−17.1	379.8	327.9	270.4	282.0
	N	Falsetto	/i/	224.6	20.5	12.0	−18.3	−11.4	339.8	336.8	302.5	246.0
M2	Y	Modal*	/a/	148.4	8.0	9.1	−7.5	−8.4	210.1	191.5	173.4	180.3
	N	Modal	/i/	149.4	8.2	6.5	−6.5	−6.0	130.9	128.4	123.1	119.9
	N	Modal-loud	/a/	181.6	5.6	4.1	−4.5	−3.5	720.5	670.9	349.2	382.2
	N	Modal-loud	/i/	182.6	11.1	5.3	−7.6	−4.8	551.5	501.5	272.7	298.3
M3	Y	Modal	/a/	98.9	7.3	5.3	−7.0	−4.9	402.2	381.0	452.0	462.5
	N	Modal	/i/	98.7	7.4	6.3	−7.0	−5.9	339.2	317.2	451.0	463.6
	N	Modal-loud	/i/	100.9	7.5	2.3	−6.9	−2.0	611.9	557.7	663.8	679.5
	N	Breathy	/a/	85.3	29.3	32.8	−20.0	−19.0	218.6	209.7	454.5	457.5
	N	Breathy	/i/	85.0	16.8	22.2	−14.9	−16.7	217.4	210.5	390.8	387.1
M4	Y	Modal-loud	/a/	122.1	10.0	8.3	−8.8	−7.2	703.0	665.0	502.8	522.9
	N	Modal-loud*	/i/	120.8	8.4	6.1	−7.7	−5.4	475.9	425.8	371.4	399.7
	N	Breathy	/a/	124.5	23.2	18.9	−17.6	−16.9	150.6	128.4	166.8	183.5
	N	Breathy	/i/	122.1	18.7	13.0	−7.0	−7.0	225.3	196.2	185.6	197.1
F1	Y	Modal-loud	/a/	224.6	14.7	14.7	−12.8	−12.9	467.3	428.0	285.7	296.6
	N	Modal-loud	/i/	229.5	8.5	11.2	−7.8	−9.4	558.0	617.1	319.6	371.2

SID	CAL	Voice Quality	Vowel	F0 (Hz) ACC	H1–H2 (dB) OVV	ACC	HRF (dB) OVV	ACC	MFDR (L/s²) OVV	ACC	AC Flow (mL/s) OVV	ACC
	N	Breathy	/a/	229.5	25.6	21.5	−24.1	−20.8	177.4	183.9	202.1	203.9
	N	Breathy	/i/	236.8	9.8	15.1	−9.7	−14.7	142.3	139.9	121.9	119.2
	Y	Falsetto	/a/	488.3	9.2	3.9	−9.0	−3.7	303.8	342.4	122.9	126.9
	N	Falsetto	/i/	481.0	5.4	0.0	−4.6	0.3	406.1	438.6	140.1	144.4
F2	Y	Modal-loud	/a/	224.6	9.3	8.3	−7.5	−6.2	339.7	313.1	171.2	171.4
	N	Modal-loud	/i/	207.5	8.5	8.0	−6.7	−5.9	313.4	311.2	182.0	170.8
	N	Breathy	/a/	201.4	13.4	9.6	−7.6	−9.3	179.2	155.8	108.2	116.3
F3	Y	Modal-loud	/a/	188.0	8.6	0.0	−7.7	0.6	200.1	188.3	134.0	109.8
	Y	Breathy	/a/	195.3	11.9	12.7	−11.3	−12.2	215.4	196.6	160.1	169.2
	N	Breathy	/i/	201.4	3.3	6.5	−2.9	−5.6	230.9	139.1	118.0	94.6

TABLE II

Estimation error of ACC-based measures of GVV with OVV-based measures as reference. pp = percentage points.

Measures	Units	Absolute error Mean ± Stdv	Relative error Mean ± Stdv
H1–H2	dB	3.2 ± 2.2	31.1% ± 27.3 pp
HRF	dB	2.4 ± 2.0	28.4% ± 29.1 pp
AC Flow	mL/s	15.8 ± 14.4	6.8% ± 5.8 pp
MFDR	L/s^2	26.8 ± 20.6	9.3% ± 7.5 pp
RMSE GVV	mL/s	25.0 ± 14.1	10.4% ± 4.0 pp
RMSE dGVV	L/s^2	74.2 ± 49.1	25.7% ± 10.3 pp