Glottal Airflow Estimation Using Neck Surface Acceleration and Low-Order Kalman Smoothing

Arturo Morales, Juan I. Yuz, *Member, IEEE*, Juan P. Cortés, Javier G. Fontanet, and Matías Zañartu, *Senior Member, IEEE*

***Abstract*—The use of non-invasive skin accelerometers placed over the extrathoracic trachea has been proposed in the literature for measuring vocal function. Glottal airflow is estimated using inverse filtering or Bayesian techniques based on a subglottal impedance-based model when utilizing these sensors. However, deviations in glottal airflow estimates can arise due to sensor positioning and model mismatch, and addressing them requires a significant computational load. In this paper, we utilize system identification techniques to obtain a low order state-space representation of the subglottal impedance-based model. We then employ the resulting low order model in a Kalman smoother to estimate the glottal airflow. Our proposed approach reduces the model order by 94% and requires only 1.5% of the computing time compared to previous Bayesian methods in the literature, while achieving slightly better accuracy when correcting for glottal airflow deviations. Additionally, our Kalman smoother approach provides a measure of uncertainty in the airflow estimate, which is valuable when measurements are taken under different conditions. With its comparable accuracy in signal estimation and reduced computational load, the proposed approach has the potential for real-time estimation of glottal airflow and its associated uncertainty in wearable voice ambulatory monitors using neck-surface acceleration.**

***Index Terms*—Vocal folds, vocal hyperfunction, system identification, kalman smoothing.**

I. Introduction

VOICE disorders have a lifetime prevalence of approximately 30% of the adult population in the United States, with an active patient population of approximately 7% every year [1], [2]. One of the most common voice disorders is referred to as vocal hyperfunction (VH), which is associated with excessive perylaryngeal musculoskeletal activity [3]. An updated framework of the etiology and pathophysiology of VH classifies the pathology into two types: Phonotraumatic vocal hyperfunction (PVH) and nonphonotraumatic vocal hyperfunction (NPVH) [4]. PVH is associated with the development of benign vocal fold lesions (nodules and/or polyps) due to persistent tissue inflammation, while NPVH is associated with muscle tension dysphonia and vocal fatigue in the absence of vocal fold tissue trauma [4].

The clinical assessment of VH involves the utilization of various diagnostic techniques, including but not limited to acoustic, aerodynamic, electroglottographic, electromyographic, and laryngeal imaging sensors. The objective is to obtain quantitative measures of vocal function that are capable of identifying a voice problem, assessing its severity, obtaining a diagnosis, and providing suitable treatment [5]. Aerodynamic measures derived from glottal airflow (also known as glottal volume velocity, GVV) and subglottal pressure in patients with PVH and NPVH have been shown to be significantly different with respect to vocally healthy subjects [6], [7], [8], [9]. Normalized parameters obtained from the subglottal pressure and GVV signals, which include the peak-to-peak AC flow, maximum flow declination rate, open quotient, are typically the most salient ones [8], [9]. Obtaining direct measurements of GVV requires the use of invasive techniques, for which efforts have been made to indirectly derive GVV from other signals that can be more easily measured. Inverse filtering (IF) is the most common method to estimate glottal airflow. The IF process usually involves the estimation and removal of vocal tract resonances (i.e., formants) to obtain an estimate of GVV either from an acoustic pressure or oral airflow signal. The estimation of formants requires a time-invariant signal, for which the analysis of sustained vowels is commonly used. There exist many IF techniques in the literature, which include linear prediction [10], iterative adaptation [11], closed phase covariance [12], weighted linear prediction [13], among others. For a comprehensive review of IF algorithms, we refer the reader to relevant review papers of this topic [14], [15].

A different IF approach to estimate GVV is based on a miniature accelerometer attached to the neck-surface between the thyroid prominence and the suprasternal notch. One of the advantages of the neck-surface accelerometer is that the effects of the vocal tract resonances are minimal compared to the resonances of the neck-skin and subglottal system. In addition, the

Manuscript received 25 July 2022; revised 22 February 2023 and 24 April 2023; accepted 5 May 2023. Date of publication 17 May 2023; date of current version 26 May 2023. This work was supported in part by ANID Chile under Grants ANID-Subdirección de Capital Humano/Magíster Nacional/2020-22200156 scholarship, FONDECYT 1230623, and BASAL FB0008, and in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Grant P50DC015446, in part by DGP of the UTFSM under Grant PIIC No. 015/2021. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Juan Ignacio Godino-Llorente. *(Corresponding author: Arturo Morales.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Scientific Ethics Committee of the University of Valparaso (CEC-UV) under Application No. CB057-15, and performed in compliance with the Chilean guidelines for research with human subjects and the Declaration of Helsinki.

The authors are with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile (e-mail: arturo.morales.14@sansano.usm.cl; juan.yuz@usm.cl; juan.cortess@usm.cl; javier.gonzalezf@sansano.usm.cl; matias.zanartu@usm.cl).

Digital Object Identifier 10.1109/TASLP.2023.3277269

subglottal inverse filtering can easily handle running speech scenarios given that these resonances are mostly time-independent. Moreover, the accelerometer is a noise-robust and non-invasive sensor that has been used in ambulatory settings with portable devices, such as pocket PCs [16] and smartphones [17]. The privacy of the user is protected since the sensor does not capture intelligible speech from resonances of the vocal tract [18]. Several works have used this type of ambulatory monitor to study long-term behavior of voice use, including, vocal fatigue [19], occupational voice [20], Lombard effect [21], biofeedback [22], [23], PVH [24], [25], [26], and NPVH [27].

Subglottal IF methods focus on removing both resonances below the glottis and neck-skin effects, thus requiring a per-subject calibration process to account for anatomical differences. One approach is to manually identify poles and zeros of a parametric model that matches experimental observations of the subglottal system [18], [28]. However, the method is limited by impedance-matching through observations, which could be prone to error due to subjective selection of poles and zeros. An extension of the former approach, called Impedance-Based Inverse Filter (IBIF) [29], incorporates an impedance model for the neck-skin [30] coupled with the subglottal system, resulting in a frequency-based filter with the GVV as input and the neck-surface acceleration as output. Features of this deterministic method are its physiological relevance and reduced processing time which makes it suitable for ambulatory and real-time applications [22], [26]. However, the per-subject calibration parameters of the filter is challenging and can exhibit variations due to sensor positioning, across vowels [31], and reading passages [32]. The calibration of the IBIF model currently requires the use of a Rothenberg mask [33] and vocal tract IF processing (e.g., [9]), which adds uncertainty to the calibration and estimation process.

More recently, a state-space model obtained from the IBIF filter transformed to the time-domain has been used in a Kalman filter to account for the uncertainty of the parameter calibration [34]. In that work, the implementation of the Bayesian filter provides glottal airflow estimates from neck-skin accelerometer measurements up to time $t + k$ to estimate at time t, i.e., acting as a smoother [35]. A non-smoother version of the filter was also presented, requiring a *colored* process noise that models a parametric GVV in the frequency domain [36]. However, the approach in [34] relies on large state-space model matrices, since a moving average (MA) model of the subglottal system is used. As a consequence, the large associated computational time implies that the estimation can only be performed offline.

In this article, we propose an approach that improves previous Bayesian estimation methods based on the impedance-based model, for a faster and more reliable subglottal IF. Initially, we employ the prediction error method (PEM) [37] to obtain low-order models that match the frequency response of the IBIF filter. Subsequently, we apply these resulting low-order state-space models in a Kalman smoother to estimate the GVV signal.

The results obtained using sustained vowels indicate that the proposed method offers an improvement over previous efforts in estimating GVV. Our method achieves this by significantly reducing computational time while maintaining the same level of accuracy as well as retaining the advantages of the Bayesian framework, i.e., estimating an uncertainty measure (error covariance) of the glottal airflow estimate. Thus, the proposed approach offers an efficient and accurate Bayesian framework that has the potential to reduce experimental uncertainty and model mismatch in real-time during the ambulatory monitoring of vocal function.

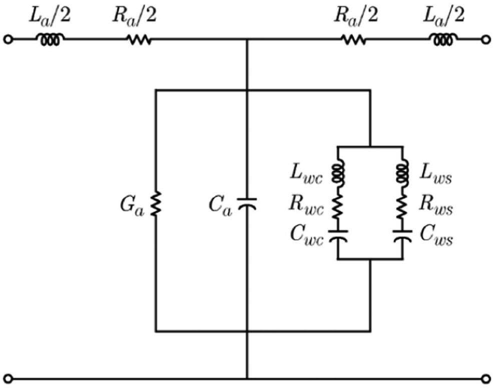

Fig. 1. Representation of the T network used to build the SIB model. The acoustic elements describe the acoustical representations for air viscosity and heat conduction losses (R_a, G_a), elasticity (C_a), inertia (L_a), and the cartilage (R_{wc}, L_{wc}, C_{wc}) and soft tissue (R_{ws}, L_{ws}, C_{ws}) components for the yielding wall. Reprinted with permission from Ref. [29]. Copyright 2013 IEEE.

The outline of the article is as follows: Section II presents a frequency domain model of the subglottal tract that has been successfully utilized in prior research to estimate the glottal airflow using inverse filtering. Section III addresses the problem of obtaining a low-order state-space representation of this model and demonstrates how it can be utilized in a Kalman smoother to obtain a glottal airflow estimate in a Bayesian framework. Section IV outlines the methods used in this study. Section V presents the results of comparing the proposed approach against state-of-the-art subglottal inverse filters. Section VI discusses the results and their implications, as well as future directions. Finally, Section VII concludes this article.

II. Subglottal Impedance-Based Model

To understand how we relate the accelerometer and the glottal airflow signals, we summarize the mechano-acoustic model of the subglottal tract and neck-skin impedance. The model relates glottal airflow and neck-skin acceleration during phonation in the frequency domain, and it has been previously used and inverted in the IBIF framework [29] for sustained vowels [9], [29], [38] and ambulatory recordings [26].

The subglottal impedance-based (SIB) model [29] is the underlying acoustic model of the subglottal system based on mechano-acoustic analogies, transmission line principles, and physiological descriptions. The model is built using T-equivalent segments of lumped acoustic elements that relate acoustic pressure $P(\omega)$ (representing the *voltage*) to airflow volume velocity $U(\omega)$ (representing the *current*), where ω is frequency (see Fig. 1).

Concatenating multiple T-equivalent segments as those shown in Fig. 1, we can model the subglottal system [39], [40], [41],

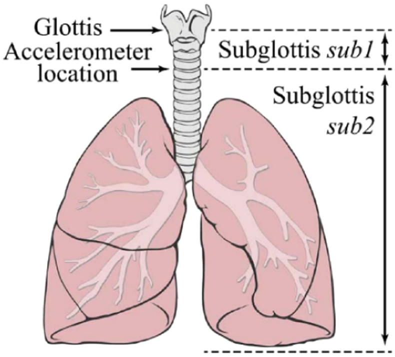

Fig. 2. Physiological representation of the subglottal system. The tract above the accelerometer, which is placed on the skin surface over the extrathoracic trachea and below the glottis, is labeled as $sub1$. The tract below this location until the alvioli is labeled as $sub2$. Reprinted with permission from Ref. [29]. Copyright 2013 IEEE.

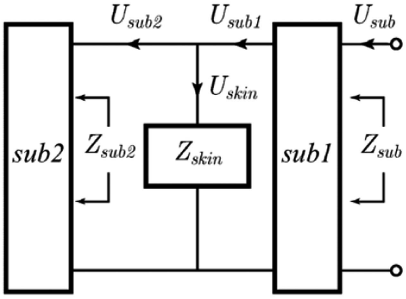

Fig. 3. Electrical representation of the subglottal system. Reprinted with permission from Ref. [29]. Copyright 2013 IEEE.

whose physiological and electrical representations are shown in Figs. 2 and 3, respectively, to describe the SIB model components [29].

The SIB model is a forward model that relates the glottal airflow and the acceleration of the neck-skin in the frequency domain by

$$\dot{U}_{\text{skin}}(\omega) = T_{\text{skin}}(\omega) \cdot U_{\text{sub}}(\omega) \tag{1}$$

where $\dot{U}_{\text{skin}}(\omega)$ represents the volume acceleration at the neck-surface (i.e., linear acceleration signal multiplied by the surface A_{acc} of the accelerometer), $U_{\text{sub}}(\omega)$ the volume velocity entering the subglottal tract (i.e., the sign inverted version of GVV), and $T_{\text{skin}}(\omega)$ the SIB model. $T_{\text{skin}}(\omega)$ is modeled by

$$T_{\text{skin}}(\omega) = \frac{H_{\text{sub1}}(\omega) \cdot Z_{\text{sub2}}(\omega) \cdot H_d(\omega)}{Z_{\text{sub2}}(\omega) + Z_{\text{skin}}(\omega)} \tag{2}$$

where $H_{\text{sub1}}(\omega)$ is the frequency response of the subglottal section from the glottis to the accelerometer location, and $H_d(\omega) = j\omega$ is a derivative filter. $Z_{\text{sub2}}(\omega)$ is a frequency-dependent driving-point impedance representing the subglottal section below the accelerometer position until the alvioli. The neck-skin impedance $Z_{\text{skin}}(\omega)$ is modeled based on a mechanical analogy with a series RLC circuit, i.e.,

$$Z_{\text{skin}}(\omega) = R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) + Z_{rad} \tag{3}$$

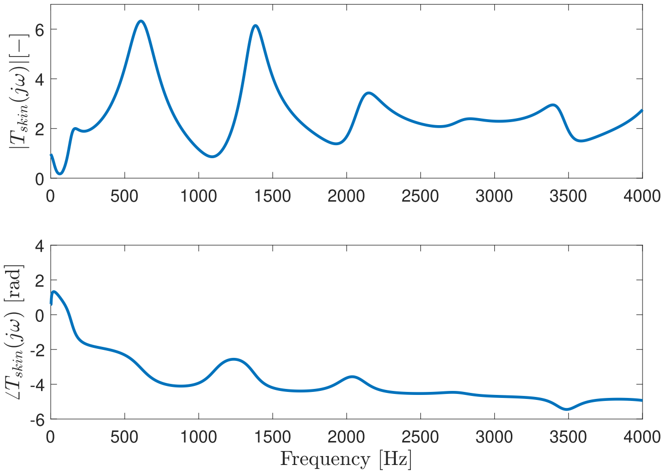

Fig. 4. Example of the magnitude (top) and phase (bottom) of the frequency response $T_{\text{skin}}(\omega)$ of the SIB model corresponding to $\boldsymbol{Q}$ factors of the subject FN001 (see Table I in Section IV).

where R_m, M_m, and K_m are the per-unit-area resistance, inertance, and stiffness of the skin, respectively. The last term (Z_{rad}) corresponds to the inductive radiation impedance due to the accelerometer loading, i.e., the derivative term proportional to the accelerometer mass per-unit-area, such that

$$Z_{rad} = H_d(\omega) \cdot \frac{M_{acc}}{A_{acc}}, \tag{4}$$

where M_{acc} is the accelerometer mass, and A_{acc} the accelerometer surface. The SIB model shown in (1)–(3) is defined by five subject-specific parameters: three related to the mechanical properties of the skin (R_m, M_m, and K_m), and two that consider the length of the trachea and the accelerometer location (which are implicitly included in $Z_{\text{sub2}}(\omega)$ and $H_{\text{sub1}}(\omega)$). Nominal values for the neck-skin impedance are obtained from direct measurements [30], as well as values for the length of the trachea [42] and an estimated position of the sensor with respect to the glottis. Since these nominal values differ per subject, a set of scale factors $\boldsymbol{Q}$ is needed, such that:

$$\boldsymbol{Q} = \{Q_1, Q_2, Q_3, Q_4, Q_5\} \tag{5}$$

$$R_m = 2320 \cdot Q_1 \qquad [\text{g} \cdot s^{-1} \cdot \text{cm}^{-2}] \tag{6}$$

$$M_m = 2.4 \cdot Q_2 \qquad [\text{g} \cdot \text{cm}^{-2}] \tag{7}$$

$$K_m = 491000 \cdot Q_3 \qquad [\text{dyn} \cdot \text{cm}^{-3}] \tag{8}$$

$$L_{\text{trachea}} = 10 \cdot Q_4 \qquad [\text{cm}] \tag{9}$$

$$L_{\text{sub1}} = 5 \cdot Q_5 \qquad [\text{cm}]. \tag{10}$$

In practice, these $\boldsymbol{Q}$ factors can be obtained using Particle Swarm Optimization (PSO) [43] to minimize the root-mean-square-error of the GVV waveform (and its derivative) between the *true* GVV (obtained by inverse filtering oral airflow measurements from a Rothenberg mask [33]) and the GVV obtained from the accelerometer measurements inverting the model, as shown in (1)-(3) and previous studies [9], [29]. Fig. 4 shows an example of the magnitude and phase of $T_{\text{skin}}(\omega)$ for a set of $\boldsymbol{Q}$ factors.

Note that the IBIF scheme inverts the SIB model to obtain the glottal airflow from the acceleration signal. For further details on the SIB model and the IBIF scheme, see [29].

III. Kalman Smoothing for Subglottal IF

To estimate GVV through a Kalman smoother (KS), we first utilize system identification techniques to obtain low order state-space representations of the SIB model. Then, this state-space model is used in a KS to solve the inverse filtering problem. The proposed Bayesian approach provides estimates of the GVV and its associated uncertainty. Also, Kalman smoothing allows to consider process and measurement noise, which is an important consideration when dealing with measurements from patients.

A. State-Space Representation of the SIB Model

The SIB model (1)-(3) is defined in the frequency domain and depends on the $\boldsymbol{Q}$ factors in (5)-(10). In that model, impedance $Z_{\text{sub2}}(\omega)$ is numerically obtained from discrete frequency points based on a physiological description of the lungs structure [44].

We obtain a state-space representation of the SIB model based on the frequency response (1)-(3), given by the matrices ($\boldsymbol{A}_s$, $\boldsymbol{B}_s$, $\boldsymbol{C}_s$, $\boldsymbol{D}_s$) in the following model structure:

$$\boldsymbol{x}_{t+1}^s = \boldsymbol{A}_s \boldsymbol{x}_t^s + \boldsymbol{B}_s u_t^g \tag{11}$$

$$y_t^a = \boldsymbol{C}_s \boldsymbol{x}_t^s + \boldsymbol{D}_s u_t^g, \tag{12}$$

where y_t^a is the accelerometer signal, u_t^g is the glottal airflow, and $\boldsymbol{x}_t^s \in \mathbb{R}^{n\times 1}$ is the state vector.

To find a state-space representation of the SIB model from its frequency response, we apply system identification techniques using the frequency domain input-output data $\{\omega_k, T_{\text{skin}}(\omega_k)\}$ with $\omega_k = 2\pi k f_s/N$ $(k = 0, 1, \ldots, N-1)$. Specifically, we first use the subspace identification N4SID algorithm [45], [46], [47] to obtain initial estimates $(\boldsymbol{A}_{s0}, \boldsymbol{B}_{s0}, \boldsymbol{C}_{s0}, \boldsymbol{D}_{s0})$ of the state-space matrices. Then, we apply the prediction error method [37] in the frequency domain to refine the matrices estimates. In summary, given the order n of the system from (11)–(12), we find the matrices $\boldsymbol{A}_s \in \mathbb{R}^{n\times n}$, $\boldsymbol{B}_s \in \mathbb{R}^{n\times 1}$, $\boldsymbol{C}_s \in \mathbb{R}^{1\times n}$ and $\boldsymbol{D}_s \in \mathbb{R}$, such that the following mean square error is minimized:

$$V(\boldsymbol{A}_s, \boldsymbol{B}_s, \boldsymbol{C}_s, \boldsymbol{D}_s) = \sum_{k=0}^{N-1} |T_{\text{skin}}(\omega_k) - \hat{T}_{\text{skin}}(\omega_k)|^2,$$

with $\hat{T}_{\text{skin}}(\omega_k) = \boldsymbol{C}_s(j\omega_k - \boldsymbol{A}_s)^{-1}\boldsymbol{B}_s + \boldsymbol{D}_s$.

Notice that to estimate the matrices in the state-space description, it is necessary to choose the model order n. In fact, the model order selection defines the trade-off between quality (fitting of the identified model to the data) and complexity (due to the size of the model matrices). The selection of the appropriate state-space model order n will be discussed later in the numerical results presented in Section V.

Note that, since the state-space representation is obtained by system identification techniques, i.e., a fitting process that approximates the frequency response of the SIB model, a physical interpretation is possible for the input and output variables only (i.e., glottal airflow and neck-surface accelerometer, respectively). The state-space variables and matrices $\boldsymbol{A}_s, \boldsymbol{B}_s, \boldsymbol{C}_s$, and $\boldsymbol{D}_s$ are internal components that combine skin effects, subglottal resonances, tracheal yielding walls, etc., to produce an approximation of the SIB model in the frequency domain, and cannot be disentangled for individual physical interpretations.

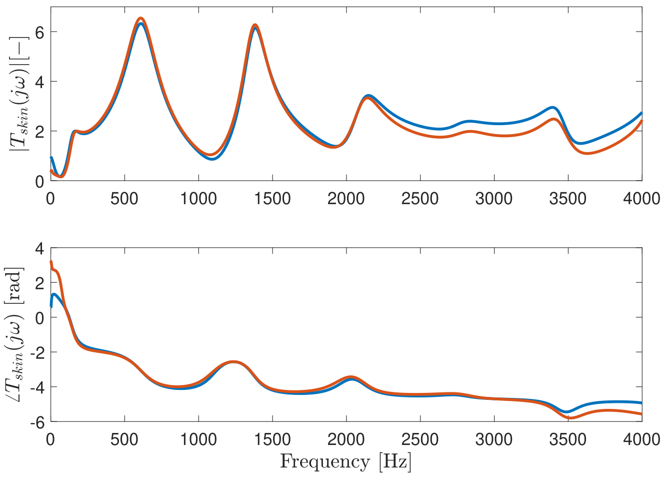

Fig. 5. Magnitude (top) and phase (bottom) of the frequency response $T_{\text{skin}}(\omega)$ of the SIB model (blue solid line) and of a state-space representation of order 20 obtained from it (red solid line).

Fig. 5 shows the numerically obtained frequency response data of $T_{\text{skin}}(\omega)$ compared to the frequency response of the state-space model of order 20 obtained from the data $\{\omega_k, T_{\text{skin}}(\omega_k)\}$ with $\omega_k = 2\pi k f_s/N$ $(k = 0, 1, \ldots, N-1)$ and using the techniques described above. We can notice that both frequency responses are similar, specifically in the bandwidth from 0 and 2000 Hz. The results in Fig. 5 show that it is possible to find low order state-space representation whose frequency response is accurate, in a mean square sense, to the one of the original SIB model. This procedure can be applied to any SIB model and, therefore, to any subject.

B. Inverse Filtering Through a Kalman Smoother

To be able to estimate the glottal airflow through a KS, it is necessary to have a state-space model whose output is the measured signal, i.e. the acceleration on the neck-skin y_t^a, and whose state vector includes the signal to estimate, i.e. the glottal airflow u_t^g.

In the state-space model (11)-(12), we are interested in estimating the GVV input u_t^g which is not directly observed. Therefore, to be able to estimate it using a Kalman filter or smoother, we include this signal in the state vector of the model by assuming that the input u_t^g is the output of a discrete-time filter $H(z)$ whose input is white noise [48]. In the following subsections, we discuss two possibles choices of $H(z)$ to include u_t^g in the state vector $\boldsymbol{x}_t^s$ without significantly increasing the order of the state-space model.

1) Modeling the Input as White Noise (WT): A straightforward model is to consider the input u_t^g as Gaussian white noise $w_t \sim \mathcal{N}(0, \sigma_w^2)$ [49], i.e., we assume a filter $H(z) = 1$. In this way, we do not introduce any prior assumption about the nature of the signal u_t^g. Thus, redefining the state vector as

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{x}_t^s & u_t^g \end{bmatrix}^\top, \tag{13}$$

we have that (11)-(12), can be rewritten as

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}w_t \tag{14}$$

$$y_t^a = \boldsymbol{C}\boldsymbol{x}_t \tag{15}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_s & \boldsymbol{B}_s \\ \boldsymbol{0} & 0 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} \tag{16}$$

$$\boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^\top \in \mathbb{R}^{(n+1)\times 1} \tag{17}$$

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_s & \boldsymbol{D}_s \end{bmatrix} \in \mathbb{R}^{1\times(n+1)} \tag{18}$$

and the input w_t to the augmented model is a Gaussian distributed white noise sequence, i.e., $w_t \sim \mathcal{N}(0, \sigma_w^2)$. Then we have that, in (14), the process noise is given by $\boldsymbol{w}_t = \boldsymbol{B}w_t$ with $\boldsymbol{w}_t \sim \mathcal{N}_{\boldsymbol{w}}(\boldsymbol{0}, \boldsymbol{R})$, where the covariance matrix of the noise process matrix is given by:

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_w^2 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}. \tag{19}$$

This process noise accounts for model mismatch due to sensor position and calibration process with the Rothenberg mask. On the other hand, experimental uncertainty from the accelerometer sensor can be included in (15) by assuming that y_t^a contains an output noise, i.e., an additive Gaussian white noise component $v_t \sim \mathcal{N}_v(0, \sigma_v^2)$. Thus, we have that the deterministic system (14)-(15) becomes a stochastic model (with no input):

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{w}_t \tag{20}$$

$$y_t^a = \boldsymbol{C}\boldsymbol{x}_t + v_t. \tag{21}$$

The model above can then be used in a KS to estimate the glottal airflow from accelerometer measurements.

2) Modeling the Input Using a Butterworth Filter (BW): An alternative model for the filter $H(z)$ is based on the effect of the glottis in the spectral domain [36]. Specifically, the glottal airflow u_t^g can be modeled as the output of a low-pass system excited by an impulse train. This low-pass filter was modeled by Fant [50] in the continuous-time domain, considering four poles on the negative real axis:

$$G(s) = \frac{U_0}{\prod_{i=1}^{4}(1 - s/s_{ri})}, \tag{22}$$

where $|s_{r1}| \simeq |s_{r2}| = 2\pi 100$ rad/s, $|s_{r3}| = 2\pi 2000$ rad/s, $|s_{r4}| = 2\pi 4000$ rad/s and U_0 is a gain factor. The poles of the filter s_{r1} and s_{r2} are chosen to take into account the variability with respect to the speaker. In particular, given that the subjects considered in this article are women, we choose $|s_{r1}| \simeq |s_{r2}| = 2\pi 200$ rad/s. On the other hand, the poles s_{r3} and s_{r4} are not included in the filter since the frequency bandwidth of the signals we deal with is below 2000 Hz. As a result, we consider a second-order Butterworth low-pass filter with a cutoff frequency of 200 Hz corresponding to an estimate of the low range of the fundamental frequency for female voices we used in this study. To use this filter in our discrete-time framework, we transform $G(s)$ to a discrete-time equivalent mapping the poles according to [51]:

$$z_i = e^{s_i T}, \tag{23}$$

where s_i is the continuous-time pole, z_i is the discrete-time pole, and T is the sampling period, which in our case is $T = 50$ μs (equivalent to a sampling frequency of 20 kHz). Hence, we obtain the discrete-time filter:

$$H(z) = \frac{H_0}{(z - z_1)(z - z_2)} \tag{24}$$

$$= \frac{H_0 z^{-2}}{1 - (z_1 + z_2)z^{-1} + (z_1 z_2)z^{-2}}, \tag{25}$$

and, therefore, the glottal airflow is described as the output of a system excited by Gaussian white noise:

$$u_t^g = \frac{H_0 z^{-2}}{1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}} w_t. \tag{26}$$

If we redefine the state vector as:

$$\boldsymbol{x}_t = \begin{bmatrix} \boldsymbol{x}_t^s & u_{t-1}^g & u_t^g \end{bmatrix}^\top \tag{27}$$

then the system (11)-(12) can be rewritten as

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{w}_t \tag{28}$$

$$y_t^a = \boldsymbol{C}\boldsymbol{x}_t + v_t \tag{29}$$

where $\boldsymbol{w}_t \sim \mathcal{N}_{\boldsymbol{w}}(\boldsymbol{0}, \boldsymbol{R})$, $v_t \sim \mathcal{N}_v(0, \sigma_v^2)$ and

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_s & \boldsymbol{B}_s & \boldsymbol{0} \\ \boldsymbol{0} & 0 & 1 \\ \boldsymbol{0} & -\alpha_2 & -\alpha_2 \end{bmatrix} \in \mathbb{R}^{(n+2)\times(n+2)} \tag{30}$$

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_s & 0 & \boldsymbol{D}_s \end{bmatrix} \in \mathbb{R}^{1\times(n+2)} \tag{31}$$

$$\boldsymbol{R} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \sigma_{\tilde{w}}^2 \end{bmatrix} \in \mathbb{R}^{(n+2)\times(n+2)}. \tag{32}$$

In the following, these models will be referred to as the white noise (WT) model and the Butterworth (BW) model, respectively. Also, by model order n we will be referring to the order of the state-space representation of the SIB model (11)-(12) instead of the order of the WT and BW models (which are $n + 1$ and $n + 2$, respectively).

Once we have built the stochastic state-space SIB model as shown in (20)-(21), it is possible to apply a KS [52] (see

Algorithm 1: Kalman Smoother Algorithm.

```
procedure KS$\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{R}, \sigma_v^2, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, y_0^a, \ldots, y_N^a$
  Initialization: $\hat{\boldsymbol{x}}_{0|-} = \boldsymbol{\mu}_0$ and $\boldsymbol{P}_{0|-} = \boldsymbol{\Sigma}_0$
  for $t = 0, \ldots, N$ do
    $\boldsymbol{K}_t \leftarrow \boldsymbol{P}_{t|t-1}\boldsymbol{C}^\top(\boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}^\top + \sigma_v^2)^{-1}$
    $\hat{\boldsymbol{x}}_{t|t} \leftarrow \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t(y_t^a - \boldsymbol{C}\hat{\boldsymbol{x}}_{t|t-1})$
    $\boldsymbol{P}_{t|t} \leftarrow (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_{t|t-1}$
    $\hat{\boldsymbol{x}}_{t+1|t} \leftarrow \boldsymbol{A}\hat{\boldsymbol{x}}_{t|t}$
    $\boldsymbol{P}_{t+1|t} \leftarrow \boldsymbol{A}\boldsymbol{P}_{t|t}\boldsymbol{A}^\top + \boldsymbol{R}$
  end for
  for $t = N, \ldots, 0$ do
    $\boldsymbol{G}_t \leftarrow \boldsymbol{P}_{t|t}\boldsymbol{A}^\top\boldsymbol{P}_{t+1|t}^{-1}$
    $\hat{\boldsymbol{x}}_{t|T} \leftarrow \hat{\boldsymbol{x}}_{t|t} + \boldsymbol{G}_t(\hat{\boldsymbol{x}}_{t+1|T} - \hat{\boldsymbol{x}}_{t+1|t})$
    $\boldsymbol{P}_{t|T} \leftarrow \boldsymbol{P}_{t|t} + \boldsymbol{G}_t(\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1|t})\boldsymbol{G}_t^\top$
  end for
end procedure
```

Algorithm 1) to estimate the state $\boldsymbol{x}_t$ and the GVV signal u_t^g, using time domain data for $t \in \{0, \ldots, N\}$. In Algorithm 1, $\hat{x}_{i|\ell}$ denotes the state estimate at time $t = i$ using the data for $t \in \{0, \ldots, \ell\}$, and $P_{i|\ell}$ denotes the associated estimation error covariance matrix.

IV. Methodology

To describe the procedure used to estimate the glottal airflow through a KS, we first introduce the data considered in this study and later discuss how the $\boldsymbol{Q}$ parameters for the SIB model are found. Subsequently, by applying the methods described in Section III-A, we find the best order of the state-space representation of the SIB model (11)-(12) for all subjects in this study. Then, we select the variances σ_w^2 and σ_v^2 of the process and output noises, respectively, for running the KS algorithm and obtaining the GVV estimates. Finally, we define performance metrics that will help comparing the different approaches.

A. Human Subject Recordings

In this study, we use synchronous signals collected from an accelerometer (ACC) placed over the extrathoracic trachea and from a Rothenberg mask (OVV), corresponding to six adult women uttering sustained vowels /a/ and /i/. The minimum length of the sustained vowels is 12 seconds, and the maximum is 20. Study participants consisted of three female PVH patients diagnosed with vocal fold nodules (noted as FP) and three female participants with no history of voice disorders (noted as FN). This study only used a reduced number of subjects as a proof of concept for the proposed signal processing scheme and it is not intended to show classification between groups. Informed consent was obtained from all subjects participating in this study, and experimental protocols were approved by the Scientific Ethics Committee of the University of Valparaso (CEC-UV) under Application No. CB057-15, in compliance with the Chilean guidelines for research with human subjects and the Declaration of Helsinki.

TABLE I
Resulting $\boldsymbol{Q}$ Parameters for the Subjects Analyzed in This Study

	$\boldsymbol{Q}$ **parameters**				
	$Q_{1,2,3} \in [0.1, 20]$			$Q_{4,5} \in [0.6, 1.2]$	
Subject	Q_1	Q_2	Q_3	Q_4	Q_5
FN001	1.40	2.77	11.59	1.1	0.6
FN003	0.67	4.97	19.72	1.1	0.6
FN006	0.53	3.00	16.64	1.2	0.8
FP003	0.59	3.02	16.45	1.1	0.6
FP004	0.56	2.39	6.23	1.1	0.6
FP008	0.24	1.86	8.45	1.1	0.6

B. Obtaining the Q Factors

The $\boldsymbol{Q}$ factors are the parameters that define the SIB model. Therefore, the first step to apply the proposed approach is to find these factors using the experimental data. Thus, we compute the $\boldsymbol{Q}$ factors for each subject by using a particle swarm optimization (PSO) scheme [26]. The PSO scheme is performed using the glottal airflow signal obtained through an inverse filtering technique that uses the OVV signal as input [33]. This GVV signal is considered the *ground truth* and will be used to compare the different estimation methods. The $\boldsymbol{Q}$ parameters of the six subjects are shown in Table I.

Although the $\boldsymbol{Q}$ parameters and the SIB model are obtained for vowel /a/ to run the calibration scheme, we utilize the data collected for vowel /i/ to test the approaches in a scenario where the pronounced vowel does not match the vowel used to calibrate the model. Vocal tract configuration and laryngeal height are known to be different between these two vowels.

C. Selecting the Order of the State-Space Representation

The order n of the state-space representation (11)-(12) is an important user selected parameter required for the identification algorithm. To choose its value, we consider a metric that measures the quality of the fit in the frequency domain. Specifically, we use the root-mean square error (RMSE) in the frequency response:

$$\text{RMSE}_f = \sqrt{\frac{1}{N}\sum_{k=0}^{N-1}|T_{\text{skin}}(\omega_k) - \hat{T}_{\text{skin}}(\omega_k)|^2} \qquad (33)$$

where T_{skin} is the frequency response of the SIB model (given by the $\boldsymbol{Q}$ parameters) and $\hat{T}_{\text{skin}}$ the one of the state-space model.

To obtain the best model orders, we identify a state-space model as detailed in Section III-A considering model orders from 5 up to 100 and we compute the RMSE_f for each one of them. The resulting RMSE in the frequency domain are shown in Fig. 6, as *box plots* considering the different subjects when comparing the frequency responses of the SIB model with the frequency responses of the identified model, for the different model orders. It can be noticed that a minimum RMSE is obtained for orders greater or equal to 15. Also, if we observe the

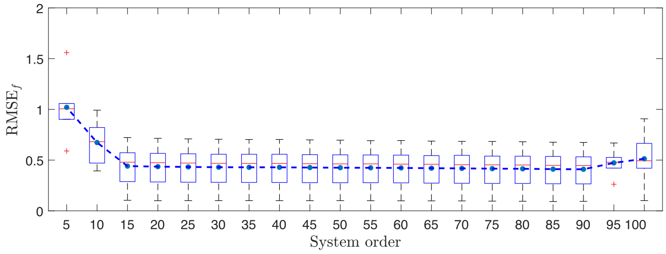

Fig. 6. Box plots of root-mean square error in frequency domain (RMSE$_f$) considering different model orders and all subjects. Mean for each order is shown as a blue dot.

models of order 95 and 100 we note that the RMSE increases, which can be associated with overfitting.

In the following, we consider only state-space models of orders 15, 20 and 25, since, even though they have almost the same frequency response, numerical stability problems or different performance in the estimation step may arise.

D. Choosing the Variances of the Process and Output Noises

The KS approach presented in Section III-B requires to choose the variances σ_w^2 and σ_v^2 of the process and output noises, respectively. To obtain the best pair (σ_w^2, σ_v^2), we quantify the absolute error in both the glottal airflow and its derivative (since both signals are used to calculate the aerodynamic measures) using as metric the following weighted mean absolute error (WMAE), given by:

$$\mathrm{WMAE}(\sigma_w^2, \sigma_v^2) = \frac{1}{2}\sum_{i=1}^{2}\left(\frac{1}{N}\sum_{t=1}^{N}\left|\Delta^{(i-1)}\tilde{u}_t^g - \Delta^{(i-1)}\hat{u}_t^g\right|\right) \tag{34}$$

where $\tilde{u}_t^g$ is the reference signal (i.e. the *true* glottal airflow obtained from OVV), $\hat{u}_t^g$ is the estimated glottal airflow by the KS (using the accelerometer measurements only) synchronized with $\tilde{u}_t^g$, and $\Delta^{(i-1)}$ is the approximate derivative operator of $(i-1)$-th order.

Fig. 7 shows an example of the obtained WMAE, in dB, for different pairs (σ_w^2, σ_v^2), considering subject FN001, model WT, and order 20. It can be noticed that the minimum WMAE is obtained in the dark blue area, and that it remains constant along in the diagonals. The latter highlights the fact that the WMAE depends on the quotient between the two variances, rather than their particular variance values.

Table II shows the ratios σ_w^2/σ_v^2 which gives the minimum WMAE for each subject, for different model orders and for the two different input models for u_t^g proposed in Section III-B.

Once the noise variances are chosen (in what follows we choose $\sigma_v^2 = 10^2$, and σ_w^2 is obtained according Table II), it is possible to apply the Kalman smoother to estimate the GVV signal.

E. Performance Metrics

A simple metric used to quantify the quality of the estimates in the time domain is the RMSE resulting from comparing the reference signal with the estimated glottal airflow:

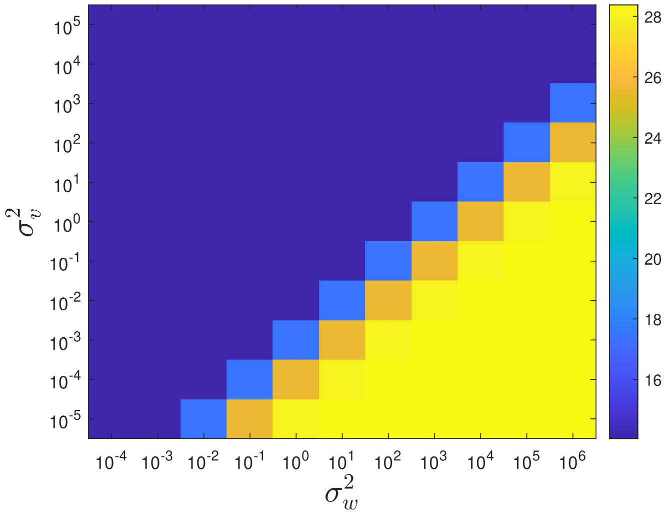

Fig. 7. Example of the weighted mean absolute error (WMAE), in dB, when comparing the estimated signal against the reference one, for different pairs (σ_w^2, σ_v^2), and considering subject FN001, model WT, and order 20.

TABLE II
RATIO BETWEEN THE DESIGN PARAMETERS σ_w^2 AND σ_v^2 OF THE KALMAN SMOOTHER CHOSEN BASED ON THE SMALLEST WMAE FOR THE /A/ VOWEL

Parameters ratio σ_w^2/σ_v^2							
Model	Model order n	Subject					
		FN001	FN003	FN006	FP003	FP004	FP008
WT	15	10^2	10^2	10^2	10^3	10^1	10^1
	20	10^1	10^4	10^1	10^3	10^1	10^1
	25	10^1	10^2	10^0	10^2	10^1	10^1
BW	15	10^1	10^1	10^1	10^2	10^1	10^1
	20	10^1	10^1	10^1	10^2	10^1	10^1
	25	10^1	10^1	10^0	10^1	10^1	10^1

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum\nolimits_{t=1}^{N}(\tilde{u}_t^g - \hat{u}_t^g)^2} \tag{35}$$

where $\tilde{u}_t^g$ is the reference signal (i.e. the *true* glottal airflow obtained from OVV) and $\hat{u}_t^g$ is the estimated glottal airflow.

An alternative metric to be used is the normalized RMSE, or NRMSE:

$$\mathrm{NRMSE} = \frac{1}{\tilde{u}_{\max}^g - \tilde{u}_{\min}^g}\sqrt{\frac{1}{N}\sum\nolimits_{t=1}^{N}(\tilde{u}_t^g - \hat{u}_t^g)^2} \tag{36}$$

which quantifies the error with respect to the *peak-to-peak* value of the reference signal.

Finally, we define a metric to quantify the quality in the estimation of the aerodynamic characteristics. This metric is the relative error given by

$$e_r = \left|\frac{\tilde{y} - \hat{y}}{\tilde{y}}\right| \quad (\%) \tag{37}$$

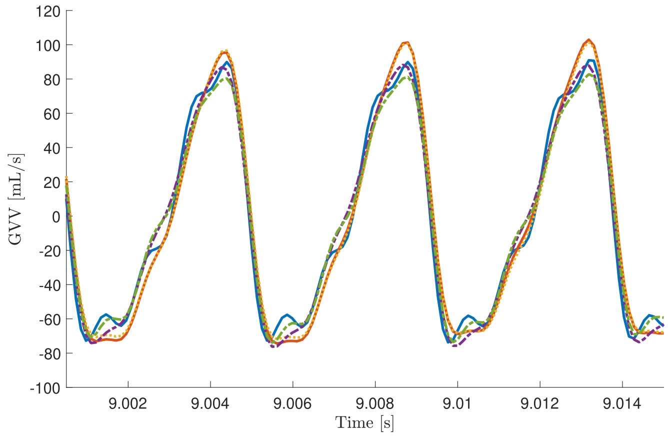

Fig. 8. Glottal airflow (GVV) estimated from different approaches: IBIF (red solid line), KF (yellow dotted line), KS-WT (purple dash-dotted line), and KS-BW (green dash-dotted line). As a reference, the glottal airflow (blue solid line) obtained through inverse filtering of the oral airflow is shown.

where $\tilde{y}$ is the value of the aerodynamic characteristic obtained from the reference signal and $\hat{y}$ is the one obtained from the estimated signal.

V. Results

To assess the performance of the proposed method, we conduct a comparative analysis with previously employed techniques such as IBIF [29] and high-order Kalman filtering [34], using the GVV signal from the IF oral airflow as reference. The evaluation process encompasses three essential parameters: (1) Time-domain analysis, where the RMSE and NRMSE are computed against the reference GVV signal. (2) Aerodynamic analysis, where selected aerodynamic measures are contrasted against those from the reference GVV signal, and (3) Computational cost analysis, where the processing times required for each approach are contrasted.

A. Estimating the Glottal Airflow

After we find the $\boldsymbol{Q}$ factors, select the model order n, and choose the variances of the process and output noises, we run Algorithm 1 to estimate the glottal airflow through the Kalman smoother.

As a qualitative result, Fig. 8 shows the estimation of the glottal airflow (GVV) by the low-order KS strategy proposed in this article. Specifically, we illustrate the estimation obtained by using the Gaussian white noise (KS-WT) and the Butterworth model (KS-BW) assumptions. As a reference, we show the *true* glottal airflow (i.e., GVV obtained from the inverse filtering of the oral airflow measurements). For comparison, we show other estimates of GVV: one obtained by using IBIF [29] and one obtained by the high-order Kalman filtering (KF) [34]. Fig. 9 shows similar plots for the derivative of the glottal airflow (dGVV). It can be noticed in both figures that the estimates of GVV and its derivative obtained from the proposed KS approach are similar to the reference signal.

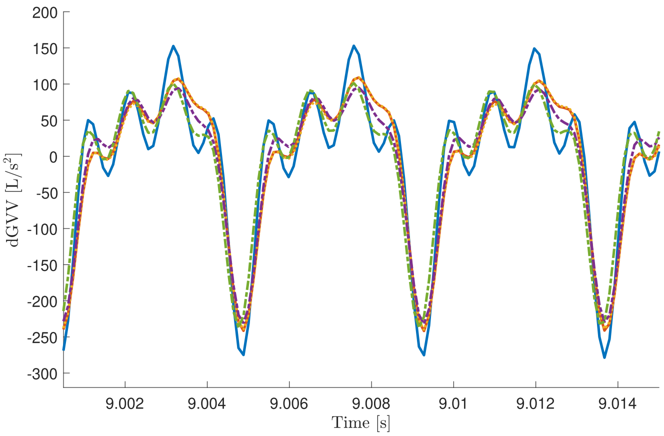

Fig. 9. Glottal airflow derivative (dGVV) estimated from different approaches: IBIF (red solid line), KF (yellow dotted line), KS-WT (purple dash-dotted line), and KS-BW (green dash-dotted line). As a reference, the glottal airflow (blue solid line) obtained through inverse filtering of the oral airflow is shown.

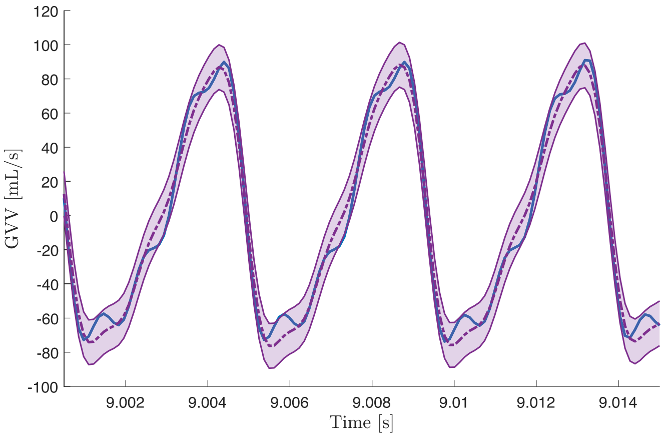

Fig. 10. Glottal airflow (GVV) estimated from a Kalman smoother by modeling it as Gaussian white noise (purple dash-dotted line), with its 95% confidence interval (purple area). As a reference, the glottal airflow (blue solid line) obtained through inverse filtering of the oral airflow is shown.

A key advantage of a Bayesian approach is that a probability function of the estimated signal is obtained. The Kalman smoother provides a point estimate of the signal, given by the expectation, $\hat{\boldsymbol{x}}_t = \mathbb{E}\{\boldsymbol{x}_t\}$, and also the variance of the estimation error, that can be used as a metric to quantify the estimation uncertainty.

Figs. 10 and 11 show the glottal airflow estimate and the associated uncertainty using the Gaussian white noise assumption and the Butterworth model, respectively. The figures also show the GVV signal obtained by oral airflow as a reference. The purple dash-dotted line represents the most probable value of the glottal airflow according to the Kalman smoothing approach. In the figures, the $\pm 2\sigma$ (or 95%) probability band is shown in purple.

As a quantitative result, Table III shows the mean RMSE and NRMSE obtained when estimating the glottal airflow from the phonemes /a/ and /i/ by using the different approaches and

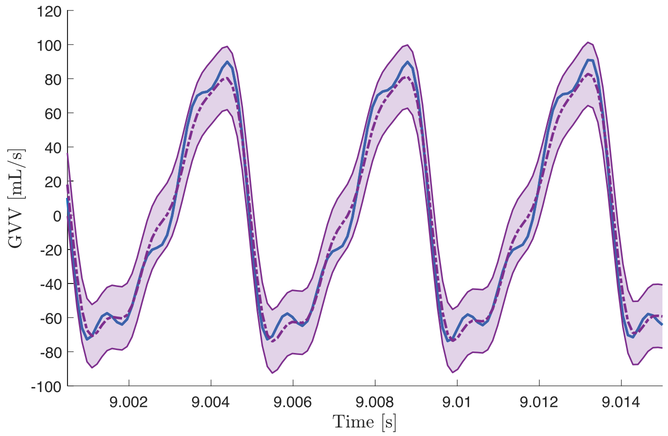

Fig. 11. Glottal airflow (GVV) estimated from a Kalman smoother by modeling it with a Butterworth filter (purple dash-dotted line), with its 95% confidence interval (purple area). As a reference, the glottal airflow (blue solid line) obtained through inverse filtering of the oral airflow is shown.

TABLE III
Root-Mean-Square Error (RMSE) and Normalized RMSE (NRMSE) Obtained From the Comparison in the Time Domain Between the Reference Signal and the Estimated Glottal Airflow

RMSE and NRMSE - Vowels: /a/ and /i/

Method	Model order n	/a/ $\overline{\text{RMSE}}$ (mL/s)	/a/ $\overline{\text{NRMSE}}$ (%)	/i/ $\overline{\text{RMSE}}$ (mL/s)	/i/ $\overline{\text{NRMSE}}$ (%)
IBIF		11.7 ± 7.4	5.8 ± 1.0	18.4 ± 11.4	12.4 ± 4.4
KF	350	11.6 ± 7.4	5.8 ± 0.9	17.8 ± 11.0	11.9 ± 4.1
KS-WT	20	11.7 ± 7.9	5.7 ± 1.1	17.1 ± 10.8	11.4 ± 4.1
KS-BW	20	10.9 ± 7.6	5.3 ± 1.1	16.5 ± 9.7	11.2 ± 3.9

considering all the subjects in Table I. For the Kalman smoothing approach, we have selected the order with the best performance according to the RMSE and NRMSE. It can be noticed that all approaches lead to similar RMSE and NRMSE for the vowel /a/, however, the best results are obtained for the Kalman smoother method with a Butterworth input model (KS-BW) proposed in this article. The similarity in the error measures can be explained because all the approaches use the same model as the basis to obtain the estimates, and this model was obtained by using the /a/ vowel.

On the other hand, we can notice that for the vowel /i/ the Bayesian approaches show better performance when compared to the IBIF method, and the best results are obtained again for the KS-BW scheme proposed in this article. The results also confirm one of the advantages of the Bayesian approaches that, by including process noise, may be less sensitive to errors in the model used to obtain the estimates.

B. Aerodynamic Measures

Aerodynamic measures obtained from the glottal airflow signal have been used to assess vocal hyperfunction [4], [9]. Therefore, an important issue to consider is to explore the capability of the proposed approach to correctly estimate such aerodynamic characteristics. Thus, in this section we compare the relative error (37) resulting of estimating these aerodynamic features, specifically (i) the peak-to-peak glottal airflow (ACFL), (ii) the negative peak of the first derivative of the glottal waveform (MFDR) [53], (iii) the difference, in dB, between the magnitude of the first two harmonics (H1-H2) [54], and (iv) the normalized amplitude quotient (NAQ) [55].

TABLE IV
Relative Error in Aerodynamic Measures Obtained From the Reference Signal and Different Approaches Considering the Vowel /a/

Relative error in aerodynamic measures - Vowel: /a/

Method	Model order n	ACFL (%)	MFDR (%)	H1-H2 (%)	NAQ (%)
IBIF		3.2 ± 2.3	12.5 ± 5.4	18.1 ± 20.6	14.0 ± 7.2
KF	350	4.6 ± 2.7	14.1 ± 5.2	20.4 ± 20.1	13.2 ± 6.7
KS-WT	20	5.9 ± 3.9	15.7 ± 5.3	17.9 ± 21.7	11.0 ± 12.9
KS-BW	20	8.0 ± 3.9	13.5 ± 4.7	16.9 ± 21.3	5.5 ± 11.7

TABLE V
Relative Error in Aerodynamic Measures Obtained From the Reference Signal and Different Approaches Considering the Vowel /i/

Relative error in aerodynamic measures - Vowel: /i/

Method	Model order n	ACFL (%)	MFDR (%)	H1-H2 (%)	NAQ (%)
IBIF		39.7 ± 26.8	45.7 ± 28.9	25.1 ± 14.2	6.5 ± 10.7
KF	350	39.1 ± 27.0	45.7 ± 29.2	27.2 ± 14.9	7.3 ± 10.8
KS-WT	20	33.8 ± 25.6	39.3 ± 27.8	24.4 ± 13.3	6.4 ± 11.5
KS-BW	20	32.0 ± 25.7	44.2 ± 33.1	25.2 ± 12.6	8.8 ± 16.5

Tables IV and V show the mean and standard deviation of the relative errors obtained when estimating the different aerodynamic features for the phonemes /a/ and /i/ for all the subjects in Table I, respectively. For the vowel /a/ it can be observed that, on average, IBIF provides better results for ACFL and MFDR, compared to the Kalman-based approaches. On the other hand, for H1-H2 and NAQ, the lowest errors are obtained with the Kalman smoothing approach proposed in the article. Considering that the state-space model used for glottal airflow was obtained from frequency domain data, it may be expected that the lowest error is obtained for the latter features with the proposed approach.

For the case of vowel /i/ (Table V) the best results are obtained for most aerodynamic features with the proposed Kalman smoothing approach, using white noise as input model (KS-WT). The only exception is for ACFL, where KS-BW leads

TABLE VI
COMPUTATION TIME IN ESTIMATING THE GLOTTAL AIRFLOW USING DIFFERENT APPROACHES

Computation time (in seconds)			
Statistics	IBIF	KF $n = 350$	KS $n = 20$
Minimum	0.014	247.9	3.6
Average	0.017	298.7	4.6
Maximum	0.022	418.9	6.8

to lower relative error. This confirms again that the adaptive characteristic from the Bayesian approach is an advantage to the estimation of GVV measures across vowels.

C. Computing Time

In ambulatory voice monitoring, glottal airflow estimation may need to be implemented in real time in portable devices. For instance, ambulatory voice treatments may require real-time biofeedback from aerodynamic signals that could detect voice misuse [22], [23], [56]. Thus, we here compare the computational cost associated to each scheme. Our main interest is to compare the proposed low-order Kalman smoother approach to a previous Kalman filter strategy proposed in [34], since both are Bayesian schemes implemented recursively. The algorithms were run in a standard laptop computer with MATLAB 2021, 16 GB of RAM, and 512 GB of SSD.

The results in Table VI show that the computation time when using the proposed Kalman smoothing approach is reduced around two orders of magnitude, from 300 to 5 seconds, approximately, on average. Moreover, if we consider that the data processed was, on average, a time interval of 20 seconds, the KS approach may well be implemented for on-line glottal airflow estimation. This is not feasible with the previous IBIF Kalman filtering approach [34], which requires approximately 20 times the length of the data to obtain the estimates.

VI. Discussion

The proposed Kalman smoother approach offers a new and enhanced Bayesian solution for estimating glottal airflow using a neck-surface accelerometer. The proposed method yields time-domain and aerodynamic performance metrics that are comparable to or surpass existing ambulatory techniques. Moreover, it provides a significantly more efficient option in terms of memory and computation time. The proposed approach exhibits comparable results to the original IBIF scheme [29] when the calibrated SIB model is used for the same vowel. However, it outperforms IBIF when the vowel is different from the calibrated one. For instance, in the case of vowel /i/, the scheme achieves a 7%-10% decrease in RMSE and a 1.5%-14.9% reduction in relative error in aerodynamic measures. This illustrates the adaptability of the Bayesian framework, which is particularly useful for analyzing long-term ambulatory data where model and sensor positioning fluctuations are expected [34]. Additionally, the impact of model mismatch can be further minimized by selecting appropriate process and output noise covariance values in the Kalman smoother. In addition, the proposed scheme provides a distinct interpretation, where we estimate the *expected* value of the glottal airflow with an associated level of uncertainty. This aspect of the Bayesian approach provides valuable insights for assessing the confidence of estimates, which is not available in the IBIF scheme and can very valuable in ambulatory data analysis.

If we consider the Kalman filter approach in [34], the main advantage of the proposed scheme is related to the model complexity and, as a consequence, the computational cost. First, the proposed Kalman smoother uses a model of order 21-22, instead of a model of order 350, which means a reduction of more than one order of magnitude (94%). This reduction is increased by a power of two when considering memory aspects due to new state-space model matrices. Also, the computation time needed by the Kalman smoother is reduced approximately 100 times, i.e., to 1.5% of the time needed by the Kalman filter. This is a significant improvement to implement a real-time Bayesian solution to estimate glottal airflow, which would be a crucial step forward in biofeedback applications for assisting patients and clinicians in the ambulatory assessment of VH [23].

It is important to emphasize that any novel technique for subglottal inverse filtering should be compared with other methods that estimate glottal airflow using different types of sensors, such as acoustic microphones or oral airflow masks. Contrasting against these standard reference signals enables an effective comparison between inverse filtering methods. It is also possible that some resulting estimates may exceed the accuracy of these standard reference signals. High precision comparisons using other sensing capabilities (e.g., hot-wire anemometry or PIV) and benchmark conditions (e.g., silicone vocal folds) would be needed to address the limitations of standard reference signals.

In this study, we tested the proposed approach with sustained vowels. Due to the adaptive capabilities of the proposed scheme, improvements are expected when assessing this scheme for running speech signals. In this regard, other considerations such as the length of the data window for the smoothing algorithm, need to be further studied for a successful real-time implementation of the proposed approach.

VII. Conclusion

An efficient Bayesian approach using a low-order Kalman smoother is proposed to estimate glottal airflow from a neck-skin acceleration signal. Low order state-space representations of an impedance-based model of the subglottal system are obtained from the frequency response of the system. This approach is shown to provide an accurate model for the system, with a significant reduction to only 6% of the model order (22/350) and to 1.5% of the computing time, when compared to previous Kalman filtering strategies for glottal airflow estimation. Given its reduced computational load and state-of-the-art accuracy, the proposed approach enables real-time estimation of glottal airflow (and its associated uncertainty) using a neck-surface accelerometer. Consequently, this approach offers novel processing capabilities for wearable voice ambulatory monitors.

References

[1] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, 2005.

[2] N. Bhattacharyya, "The prevalence of voice problems among adults in the United States," *Laryngoscope*, vol. 124, no. 10, pp. 2359–2362, 2014.

[3] J. Oates and A. Winkworth, "Current knowledge, controversies and future directions in hyperfunctional voice disorders," *Int. J. Speech- Lang. Pathol.*, vol. 10, pp. 267–277, 2008.

[4] R. E. Hillman, C. E. Stepp, J. H. Van Stan, M. Zañartu, and D. D. Mehta, "An updated theoretical framework for vocal hyperfunction," *Amer. J. Speech- Lang. Pathol.*, vol. 29, no. 4, pp. 2254–2260, 2020.

[5] J. C. Stemple, N. Roy, and B. K. Klaben, *Clinical Voice Pathology: Theory and Management*, 6th ed. San Diego, CA, USA: Plural Publishing, 2018.

[6] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Objective assessment of vocal hyperfunction," *J. Speech, Lang., Hear. Res.*, vol. 32, no. 2, pp. 373–392, 1989.

[7] E. B. Holmberg, P. Doyle, J. S. Perkell, B. Hammarberg, and R. E. Hillman, "Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy," *J. Voice*, vol. 17, no. 3, pp. 269–282, 2003.

[8] V. M. Espinoza, M. Zañartu, J. H. Van Stan, D. D. Mehta, and R. E. Hillman, "Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction," *J. Speech, Lang., Hear. Res.*, vol. 60, no. 8, pp. 2159–2169, 2017.

[9] V. M. Espinoza, D. D. Mehta, J. H. Van Stan, R. E. Hillman, and M. Zañartu, "Glottal aerodynamics estimated from neck-surface vibration in women with phonotraumatic and nonphonotraumatic vocal hyperfunction," *J. Speech, Lang., Hear. Res.*, vol. 63, no. 9, pp. 2861–2869, Sep. 2020.

[10] J. Markel and A. Gray, *Linear Prediction of Speech*. Berlin, Germany: Springer, 1976.

[11] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Commun.*, vol. 11, pp. 109–118, 1992.

[12] P. Alku, C. Magi, S. Yrttiaho, T. Bäckström, and B. Story, "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," *J. Acoust. Soc. Amer.*, vol. 125, pp. 3289–3305, 2009.

[13] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilised weighted linear prediction," *Speech Commun.*, vol. 51, no. 5, pp. 401–411, 2009.

[14] P. Alku, "Glottal inverse filtering analysis of human voice production: A review of estimation and parameterization methods of the glottal excitation and their applications," *SADHANA - Acad. Proc. Eng. Sci.*, vol. 36, pp. 623–650, 2011.

[15] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," *Comput. Speech Lang.*, vol. 28, pp. 1117–1138, 2014.

[16] P. S. Popolo, J. G. Švec, and I. R. Titze, "Adaptation of a pocket PC for use as a wearable voice dosimeter," *J. Speech, Lang., Hear. Res.*, vol. 48, pp. 780–791, 2005.

[17] D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne II, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3090–3096, Nov. 2012.

[18] H. Cheyne II, "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Ph.D. dissertation, Harvard-MIT, Division Health Sci. Technol. Speech Hear. Biosciences Technol. Prog., Cambridge, MA, USA, 2002.

[19] Z. Lei, L. Fasanella, L. Martignetti, N. Y. K. Li-Jessen, and L. Mongeau, "Investigation of vocal fatigue using a dose-based vocal loading task," *Appl. Sci.*, vol. 10, 2020, Art. no. 1192.

[20] E. J. Hunter and I. R. Titze, "Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings," *J. Speech, Lang., Hear. Res.*, vol. 53, pp. 862–875, 2010.

[21] T. H. Whittico et al., "Ambulatory monitoring of lombard-related vocal characteristics in vocally healthy female speakers," *J. Acoust. Soc. Amer.*, vol. 147, pp. EL552–EL558, 2020.

[22] A. F. Llico et al., "Real-time estimation of aerodynamic features for ambulatory voice biofeedback," *J. Acoust. Soc. Amer.*, vol. 138, pp. EL14–EL19, 2015.

[23] J. H. V. Stan et al., "Integration of motor learning principles into real-time ambulatory voice biofeedback and example implementation via a clinical case study with vocal fold nodules," *Amer. J. Speech- Lang. Pathol.*, vol. 26, no. 1, pp. 1–10, 2017.

[24] M. Ghassemi et al., "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1668–1675, Jun. 2014.

[25] J. H. V. Stan et al., "Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls," *J. Speech, Lang. Hear. Res.*, vol. 63, pp. 372–384, 2020.

[26] J. P. Cortés et al., "Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration," *PLoS One*, vol. 13, no. 12, pp. 1–22, 2018.

[27] J. H. V. Stan et al., "Differences in daily voice use measures between female patients with nonphonotraumatic vocal hyperfunction and matched controls," *Res. Article J. Speech, Lang., Hear. Res.*, vol. 64, pp. 1457–1470, 2021.

[28] K. Ishizaka, M. Matsudaira, and T. Kaneko, "Input acoustic-impedance measurement of the subglottal system," *J. Acoust. Soc. Amer.*, vol. 60, no. 1, pp. 190–197, 1976.

[29] M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, "Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 9, pp. 1929–1939, Sep. 2013.

[30] K. Ishizaka, J. C. French, and J. L. Flanagan, "Direct determination of vocal tract wall impedance," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 23, no. 4, pp. 370–373, Aug. 1975.

[31] R. Manriquez, V. Espinoza, C. Castro, J. Cortés, and M. Zañartu, "Parameter analysis and uncertainties of impedance-based inverse filtering from neck surface acceleration," in *Proc. 14th Int. Conf. Adv. Quantitative Laryngol., Voice Speech Res.*, 2021, p 26.

[32] V. Espinoza, "Stationary and dynamic aerodynamic assessment of vocal hyperfunction using enhanced supraglottal and subglottal inverse filtering methods," Ph.D. dissertation, Universidad Técnica Federico Santa María, Valpara, Chile, 2018.

[33] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *J. Acoust. Soc. Amer.*, vol. 53, no. 6, pp. 1632–1645, 1973.

[34] J. P. Cortés et al., "Kalman filter implementation of subglottal impedance-based inverse filtering to estimate glottal airflow during phonation," *Appl. Sci.*, vol. 12, no. 1, 2022, Art. no. 401.

[35] S. Sarkka, *Bayesian Filtering and Smoothing*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[36] B. Doval, C. Alessandro, and N. H. Bernardoni, "The spectrum of glottal flow models," *Acta Acustica United Acustica*, vol. 92, no. 6, pp. 1026–1046, 2006.

[37] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ, USA: Prentice Hall, 1999.

[38] J. Z. Lin, V. M. Espinoza, K. L. Marks, M. Zañartu, and D. D. Mehta, "Improved subglottal pressure estimation from neck-surface vibration in healthy speakers producing non-modal phonation," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 449–460, Feb. 2020.

[39] G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, and D. C. Shannon, "A model of acoustic transmission in the respiratory system," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 9, pp. 925–934, Sep. 1989.

[40] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, "An acoustic model of the respiratory tract," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 5, pp. 543–550, May 2001.

[41] J. L. Flanagan, *Speech Analysis, Synthesis, and Perception*, Berlin, Germany: Springer, 1972.

[42] K. N. Stevens, *Acoustic Phonetics*. Cambridge, MA, USA: MIT Press, 2000.

[43] M. Clerc, *Particle Swarm Optimization*, vol. 93. Hoboken, NJ, USA: Wiley, 2010, .

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

[45] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.

[46] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*. Norwell, MA, USA: Kluwer, 1996.

[47] T. Katayama, *Subspace Methods for System Identification*. Berlin, Germany: Springer, 2004.

[48] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[49] J. Glover, "The linear estimation of completely unknown signals," *IEEE Trans. Autom. Control*, vol. 14, no. 6, pp. 766–767, Dec. 1969.

[50] G. Fant, *Acoustic Theory of Speech Production*. Berlin, Germany: Walter de Gruyter, 1970.

[51] K. J. Åström and B. Wittenmark, *Computer Controlled Systems*. Englewood Cliffs, NJ, USA: Prentice Hall, 1997.

[52] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, 1965.

[53] E. B. Holmberg, R. E. Hillman, and J. S. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch," *J. Voice*, vol. 3, pp. 294–305, 1989.

[54] J. Kreiman, B. R. Gerratt, and N. Antoñanzas-Barroso, "Measures of the glottal source spectrum," *J. Speech, Lang., Hear. Res.*, vol. 50, pp. 595–610, 2007.

[55] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parametrization of the glottal flow," *J. Acoust. Soc. Amer.*, vol. 112, pp. 701–710, 2002.

[56] J. H. V. Stan, D. D. Mehta, and R. E. Hillman, "The effect of voice ambulatory biofeedback on the daily performance and retention of a modified vocal motor behavior in participants with normal voices," *J. Speech, Lang., Hear. Res.*, vol. 58, pp. 713–721, 2015.

Arturo Morales received the Ingeniero Civil Electrónico and M.Sc. degrees in electronic engineering from Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile, in 2021. He is currently working as an Advanced Process Control Engineer with Honeywell Chile, Santiago, Chile, where he provides support and continuous improvement services to advanced control applications in the mining industry. His research interests include Bayesian estimation, system identification, their applications to biomedical systems, robotics, and reinforcement learning. He was the recipient of the Magíster Nacional 2020 Scholarship from ANID (Chile) and the Distinción Académica Federico Santa María 2021 Award from UTFSM.

Juan I. Yuz (Member, IEEE) received the Ingeniero Civil Electrónico and M.Sc. degrees in electronic engineering from Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile, in 2001, and the Ph.D. degree in electrical engineering from The University of Newcastle, Newcastle, NSW, Australia, in 2006. From 2015 to 2019, he was the Director of the Advanced Center for Electrical and Electronic Engineering (AC3E - UTFSM). He is currently a Full Professor with the Departamento de Electrónica, UTFSM. He is coauthor of the book *Sampled-data Models for Linear and Nonlinear Systems* (Springer, 2014). His research interests include control and identification of sampled-data systems, and their applications to power electronics and biomedical systems.

Juan P. Cortés received the B.S. and M.S. degrees in electrical engineering from University of California, Los Angeles, Los Angeles, CA, USA, in 2010 and 2013, respectively, and the Ph.D. degree in electronic engineering from Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile, in 2020. He is currently Senior Research Scientist with Lanek SPA, Valparaíso, Chile, where he leads research and development on ambulatory voice monitors for clinical applications.During 2020 –2021, he was a Postdoctoral Fellow with the MGH Center for Laryngeal Surgery and Voice Rehabilitation, Boston, MA, USA . His interests include the development of digital signal processing and machine learning algorithms from physiological and neural models of voice/speech production with emphasis to the assessment of vocal function and related pathologies. Dr. Cortés is an Eta Kappa Nu Member. He was the recipient of a Qualcomm Q Award of Excellence, and Best Student Poster Award at the 49th Voice Foundation Conference.

Javier G. Fontanet received the Engineer and M.Sc. degrees in automatic from the Universidad de Oriente, Santiago de Cuba, Cuba, in 2009 and 2014, respectively. He is currently working toward the Ph.D. degree in electronic engineering with Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile. He works as a part-time Professor with the Department of Electronic Engineering, UTFSM. His research interests include system identification and its application to biomedical systems, electronics, and electrical networks.

Matías Zañartu (Senior Member, IEEE) received the B.S. degree in acoustical engineering from Universidad Tecnológica Vicente Pérez Rosales, Santiago, Chile, and the Ph.D. and M.S. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA. He is currently an Associate Professor with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile, and Director of the Advanced Center for Electrical and Electronic Engineering from the same institution, where he also leads the biomedical engineering research and development. His interests include the development of digital signal processing, system modeling, and biomedical engineering tools that involve speech, hearing, and acoustics. His research efforts have revolved around developing quantitative models of human voice production and applying these physiological descriptions for the development of clinical technologies. He is an Associate Editor for IEEE Transactions on Neural Systems and Rehabilitation Engineering and *JASA Express Letters*, Member of the Technical Committee on Speech Communication Speech Communication of the Acoustical Society of America, and Fulbright Fellow.