Continuous-time model identification of the subglottal system

Javier G. Fontanet [a],[*], Juan I. Yuz [a], Hugues Garnier [b], Arturo Morales [a], Juan Pablo Cortés [a], Matías Zañartu [a]

[a] *Department of Electronic Engineering, Universidad Técnica Federico Santa María, Av. España 1680, Valparaíso, Chile*
[b] *Université de Lorraine, CNRS, CRAN, F-54000, Nancy, France*

ARTICLE INFO

ABSTRACT

Mathematical models that accurately simulate the physiological systems of the human body serve as cornerstone instruments for advancing medical science and facilitating innovative clinical interventions. One application is the modeling of the subglottal tract and neck skin properties for its use in the ambulatory assessment of vocal function, by enabling non-invasive monitoring of glottal airflow via a neck surface accelerometer. For the technique to be effective, the development of an accurate building block model for the subglottal tract is required. Such a model is expected to utilize glottal volume velocity as the input parameter and yield neck skin acceleration as the corresponding output. In contrast to preceding efforts that employed frequency-domain methods, the present paper leverages system identification techniques to derive a parsimonious continuous-time model of the subglottal tract using time-domain data samples. Additionally, an examination of the model order is conducted through the application of various information criteria. Once a low-order model is successfully fitted, an inverse filter based on a Kalman smoother is utilized for the estimation of glottal volume velocity and related aerodynamic metrics, thereby constituting the most efficient execution of these estimates thus far. Anticipated reductions in computational time and complexity due to the lower order of the subglottal model hold particular relevance for real-time monitoring. Simultaneously, the methodology proves efficient in generating a spectrum of aerodynamic features essential for ambulatory vocal function assessment.

1. Introduction

Voice production is a product of the intricate interplay between airflow and the various structures of the phonatory system. For voiced sounds, the pressure from the lungs provides sufficient energy to induce self-sustained oscillations of the vocal folds (VFs) that result in the main aeroacoustic sound source at the glottis. These sound waves are propagated above and below the glottis through the supraglottal and subglottal systems, respectively. The supraglottal system, also commonly referred to as the vocal tract, provides the main filtering effects that are associated with speech articulation [1,2], while the role of the subglottal system is less well understood. Previous efforts have found that the subglottal system plays a key role in the complex interactions between the resulting sound waves and the airflow at the glottis that alter voice quality [3–6], trigger bifurcations and voice breaks [7,8], and yield natural quantal differences for the vowel space [2,9,10].

The study of voice aerodynamics, particularly its interaction with the supraglottal and subglottal systems, has been crucial in advancing our understanding of clinical issues, such as assessing vocal hyperfunction. Vocal hyperfunction (VH) is a type of voice disorder that

is associated with abuse and misuse of voicing [11], and affects approximately 6.6% of the adult population with a lifetime prevalence of 30% [12,13]. Prior studies have shown significant differences between normal vocal function and VH patients by measuring aerodynamic measures extracted from the glottal airflow signal, also referred to as glottal volume velocity (GVV) [14,15] from recordings of oral airflow (or oral volume velocity, OVV) with a pneumotachograph mask [16]. Notably, the GVV signal can also be obtained using subglottal inverse filtering methods using models of the subglottal system [17,18] for a neck-skin accelerometer (ACC) signal. The non-invasive, noise-robust, and portable nature of the ACC signal has resulted in significant interest in subglottal inverse filtering for studying VH [15,19,20] through ambulatory voice monitoring devices (e.g Mehta et al. [21]).

Mathematical representations of the subglottal system [22–24] have played an important role in the development of the signal processing components needed for subglottal inverse filtering. A transmission line model that represents the subglottal system was proposed by Zañartu et al. [18] to physiologically relate the ACC and GVV signals, in an approach referred to as impedance-based inverse filtering (IBIF).

* Corresponding author.
E-mail address: javier.gonzalezf@sansano.usm.cl (J.G. Fontanet).

As a result, aerodynamic features can be obtained from GVV such as peak-to-peak AC flow, open quotient, and maximum flow declination rate [25–27]. These features can then be used to identify phonatory mechanisms associated with VH, which are supported by glottal aerodynamic measures of subglottal air pressure and glottal airflow [15]. To achieve this, researchers have utilized a transfer function impedance-based model to obtain aerodynamic parameters using a neck surface accelerometer for evaluating, monitoring, and comparing differences between VH and healthy controls [18–20]. However, a significant challenge in these studies is how to accurately estimate subject-specific parameters related to the subglottal system and neck-surface mechanical properties.

The study of voice pathologies has been grounded in mathematically modeled first-principles based on physical principles [23,28,29]. Alternatively, a data-driven approach may be followed when only partial or no knowledge of the physical principles acting upon the system is used. In this case, a "grey box" approach can be applied when only certain parameters are unknown or physical knowledge is used to define the model structure (where then the parameters are estimated). Also, a "black box" modeling can be applied where both the structure and the parameters are chosen based solely on the available data [30–32]. In this paper, GVV and ACC are respectively defined as the input and output signals of the system to be identified, and then a "black box" system identification approach is applied.

System identification aims to derive a mathematical model, whether linear or nonlinear, that effectively captures the dynamic behavior of a system and exhibits a high level of accuracy. This identification process is conducted using input–output data from the system. In previous studies, the subglottal system has been represented using a linear model, and the impulse response has been obtained by inverse Fourier transformation of the frequency domain response [18,20]. The linearity assumption in the subglottal system is justified since the nonlinearity associated with frequencies dependent on resistances has a minor overall impact on the system behavior [18]. Furthermore in Fant [1], the cavities responsible for transforming the glottal source were presented as linear filters.

In this paper we propose the use of linear time-invariant (LTI) models, to describe the subglottal system in continuous-time and to use the simplified refined instrumental variable method for continuous-time (SRIVC) model estimation method to fit these models using real input–output data.

SRIVC is widely recognized as one of the most successful direct methods for continuous-time model identification from sampled data [33]. This method has been extensively applied in different fields, including chemical processes, electronic circuits, and biological systems modeling [34,35].

Additionally, in this paper we study the model order selection problem. Once the candidate continuous-time models for the subglottal system have been estimated, the Akaike, Bayes, and Young information criteria methods can be employed to determine the optimal order and structure among them. These criteria methods play a crucial role in the trade-off between accuracy of model fitting and complexity. The information criteria enable a systematic and data-driven approach in selecting the most appropriate model order [36].

The remainder of the paper is structured as follows: Section 2 presents the subglottal system description. Then, Section 3 presents SRIVC as the system identification method and the information criteria to be used to select the best model order and structure. In Section 4, the data processing required before performing the estimation for the set of candidate models is explained, and a linearity test is also conducted. Later in Section 5, the results of the model identification are presented together with the discussion of model order selection. Section 6 presents the application of the obtained model for inverse filtering using a Kalman Smoother to obtain estimates of the glottal volume velocity airflow and the associated aerodynamic features. Finally, in Section 7 conclusions are presented.
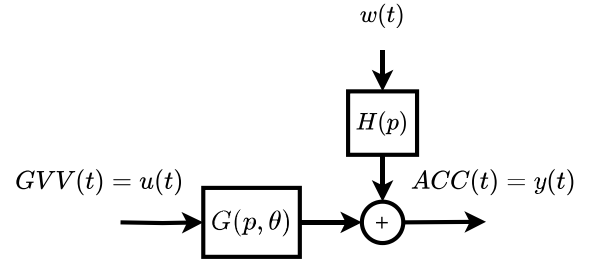

Fig. 1. Block diagram representation of the phonatory system.

2. System description

Several works have proposed to measure the acceleration on the neck skin surface generated by the airflow in the glottis to study VH. The acceleration data has been used to estimate certain parameters of an impedance-base (IB) model, which is a mechano-acoustic representation of a physiologically-based transmission line [18]. An inverse filtering has been then applied to the IB model to obtain an accurate estimation of the aerodynamic source of voice sounds at the glottis [18–20].

The skin acceleration is measured by attaching an accelerometer to the neck surface between the thyroid prominence and the suprasternal notch [17]. The main advantage of measuring acceleration is that it is a non-invasive method of studying the health of the speech system. Additionally, it is also immune to noise, making it particularly suitable for ambulatory studies [21,37].

We consider the subglottal system as represented in the block diagram shown in Fig. 1, where $u(t)$ is the input signal GVV, $y(t)$ is the measured signal ACC, and $w(t)$ is assumed to be white measurement noise.

As mentioned in the introduction, the input GVV is not directly measured, however, it is obtained from OVV using a linear prediction (LP) filter [38,39].

The subglottal system in Fig. 1 can be expressed as the following continuous-time model

$$y(t) = G(p, \theta)u(t) + H(p)w(t) \tag{1}$$

where p represents the differential operator $p = \frac{d}{dt}$, $H(p) = 1$ is assumed as the noise filter, and the subglottal system is then given by

$$G(p, \theta) = \frac{B(p, \theta)}{A(p, \theta)} = \frac{b_m p^m + b_{m-1} p^{m-1} + \cdots + b_0}{p^n + a_{n-1} p^{n-1} + \cdots + a_0} \tag{2}$$

where $G(p, \theta)$ is assumed to be proper ($n \geq m$) and the parameter vector is $\theta = \begin{bmatrix} a_{n-1} \cdots a_0 & b_m \cdots b_0 \end{bmatrix}^T \in \mathbb{R}^{n+m+1}$.

Our interest in this paper is to estimate the parameter vector θ in (2), for a given model structure (i.e., n and m), based on input and output sampled data: $u(t_k)$ corresponding to GVV and $y(t_k)$ corresponding to ACC, for a regular sampling interval T_s. The sampling instants are $t_k = k T_s$, for $k = 1, \ldots, N$, where N is the number of available synchronized data points. Thus, for identification purposes we consider the following hybrid model

$$y_u(t) = G(p, \theta)u(t)$$
$$y(t_k) = y_u(t_k) + w(t_k) \tag{3}$$

where $y_u(t)$ is the noise free output of the subglottal system, and $w(t_k)$ is assumed to be a discrete-time Gaussian distributed white measurement noise $w(t_k) \sim \mathcal{N}(0, \sigma_w^2)$.

3. Parameter estimation and information criteria

This section presents the parameter estimation method used to perform identification of the subglottal system from synchronized input–output data. The input signal is the glottal flow GVV and the output is the accelerometer signal ACC.

Furthermore, this section also presents the information criteria employed for model order selection from a set of candidate models. The criteria used are the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Young Information Criterion (YIC), which are commonly used to select the best model order from a set of candidate models.

3.1. Continuous-time transfer function estimation using time-domain data

When the theory of system identification first appeared, it was mainly aimed at the estimation of continuous-time LTI models. However, with the advance of digital computers and data acquisition systems, the estimation of discrete-time models gained prominence. In the last decades, the increase both in computation power and sampling frequency has led to a resurgence of the estimation of continuous-time models from sampled data [32,40].

Some of the difficulties that appear in discrete-time identification are related to the sampling time, which may lead to loss of information when the sampling period T_s is small. Numerical issues may appear in discrete-time since all the poles of the transfer function approach the point $z = 1$ (in the z-plane) for fast sampling rates [40,41].

The estimation of continuous-time models from a sampled dataset has been shown to have advantages over discrete-time methods when using real data [34]. Moreover, the obtained continuous-time models can be discretized for different sampling rates if needed [32].

Simplified refined instrumental variable method for continuous-time (SRIVC) modeling is one of the most applied continuous-time identification methods in the literature. It is an iterative instrumental variable method that generates filters (that depend on the system parameters) that are applied to the available input–output data at each iteration. It has been shown that SRIVC provides a consistent and optimal estimator for additive white noise [42,43].

One might consider applying the Refined Instrumental Variables for Continuous-Time Model (RIVC) method for identification, given the likelihood of colored noise in practical scenarios. However, the results that we obtained with such approach closely resemble those obtained with SRIVC. This is common in applications where the noise level is relatively small. [32].

For SRIVC the error function is given by the output error

$$\varepsilon(t_k, \theta) = y(t_k) - \frac{B(p, \theta)}{A(p, \theta)} u(t_k) \tag{4}$$

The input–output signals in (4) are filtered by the following parameter-dependent filter

$$L(p, \theta) = \frac{1}{A(p, \theta)} \tag{5}$$

Then, from (4), we obtain

$$\varepsilon(t_k, \theta) = A(p, \theta) \tilde{y}(t_k) - B(p, \theta) \tilde{u}(t_k) \tag{6}$$

where $\tilde{y}(t_k)$ and $\tilde{u}(t_k)$ are the variables pre-filtered by $L(p, \theta)$

A drawback of the previous formulation is that the polynomial in the denominator of the transfer function $G(p, \theta)$ and its parameters are unknown, and, as a consequence, an initial parameter estimate is required [33]. Then, the error in (6) can be iteratively optimized by solving in every iteration the following recursion

$$\tilde{y}^{(n)}(t_k, \hat{\theta}^j) = \tilde{\phi}^T(t_k, \hat{\theta})\theta^{j+1} + \varepsilon(t_k, \hat{\theta}^j) \tag{7}$$

where the regressor vector is given by

$$\tilde{\phi}^T(t_k, \hat{\theta}) = \left[\tilde{u}^{(m)}(t_k, \hat{\theta}^j), \ldots, \tilde{u}(t_k, \hat{\theta}^j), \\ -\tilde{y}^{(n-1)}(t_k, \hat{\theta}^j), \ldots, -\tilde{y}(t_k, \hat{\theta}^j) \right] \tag{8}$$

where $\hat{\theta}^j$ and $\hat{\theta}^{j+1}$ are the estimated parameters at the jth and $(j+1)^{th}$ iterations, respectively, and the superscripts (n), (m), and $(n-1)$

correspond to (approximate) time derivatives. The SRIVC estimator is then given by

$$\hat{\theta}_{SRIVC}^{j+1} = \left[\frac{1}{N} \sum_{k=1}^{N} \tilde{Z}(t_k, \hat{\theta}^j) \tilde{\phi}^T(t_k, \hat{\theta}^j) \right]^{-1} \\ \left[\frac{1}{N} \sum_{k=1}^{N} \tilde{Z}^T(t_k, \hat{\theta}^j) \tilde{y}^{(n)}(t_k, \hat{\theta}^j) \right] \tag{9}$$

where $\tilde{Z}(t_k)$ denotes the filtered instrument and is given by

$$\tilde{Z}^{(i)}(t_k, \hat{\theta}^j) = \frac{p^{(i)}}{A(p, \hat{\theta}^j)} \tilde{\phi}(t_k, \hat{\theta}^j) \tag{10}$$

More details about the SRIVC algorithm and its applications can be found, for example, in Garnier [35],Young [42],Young and Jakeman [44],González et al. [45].

3.2. Information criteria

Model order selection is an important task in the analysis of time series, signal processing, and system identification [46]. To determine the best order and structure of the subglottal LTI model, information criteria such as Akaike, Bayes and Young's Information Criterion (AIC, BIC and YIC respectively) can be used.

Akaike information criterion

AIC is a Kullback–Leibler (KL) cross validation approach, in which the model order is chosen such that it minimizes the KL discrepancy between the true probability distribution function and the likelihood of the model [47]. The AIC is given by

$$AIC = -2\ln p_{n_\theta}(y, \hat{\theta}^\eta) + 2n_\theta \tag{11}$$

where y is the vector of available output data of size N, n_θ is the number of parameters, $\hat{\theta}^\eta$ is the estimated parameter vector, and $p_n(y, \hat{\theta}^\eta)$ is the probability density function of the data vector. The value of the above expression is obtained for a set of candidate models, among which the one that yields the minimum value is chosen. In this way, the goodness of fit is rewarded and over-fitting is penalized [48].

Bayes information criterion

BIC is a second possible information criterion used to select model order that is given by

$$BIC = -2\ln p_{n_\theta}(y, \hat{\theta}^\eta) + n_\theta \ln N \tag{12}$$

Eq. (12) has a structure similar to the AIC. However, it differs in last penalty component that is associated with the assumption that $p(\theta)$ is independent of N, resulting in an increasing ratio of estimation to validation samples as the number of data points N grows [47].

Similar to the case of AIC, BIC is determined for a set of candidates model, taking into account model fit and penalizing over-fitting, and the model order that minimizes the BIC in (12) is then selected.

Young information criterion

YIC is a third model order selection criterion that, as a difference compared to AIC and BIC, uses the covariance matrix estimate associated with the estimated parameters [32,49,50].

In fact, YIC involves the coefficient of determination R_T^2 (13) and the variance-error norm to define the model that best fits the data and best estimates the parameters [49].

The coefficient of determination R_T^2 is a metric that assesses the accuracy of the identified model to represent the actual model [33,51], by considering the simulation error. The R_T^2 values range between 0 and 1, indicating the relation between the identified model and the real-world observations. The R_T^2 is defined by Garnier [35]

$$R_T^2 = 1 - \frac{\sigma_{\hat{\varepsilon}}^2}{\sigma_y^2} \tag{13}$$

where $\sigma_{\hat{\varepsilon}}^2$ is the variance of the estimated error, and σ_y^2 is the variance of the measured output which are respectively given by

$$\sigma_{\hat{\varepsilon}}^2 = \frac{1}{N} \sum_{k=1}^{N} (\hat{\varepsilon}(t_k) - \bar{\varepsilon}(t_k))^2 \tag{14}$$

$$\sigma_y^2 = \frac{1}{N} \sum_{k=1}^{N} (y(t_k) - \bar{y}(t_k))^2 \tag{15}$$

$$\hat{\varepsilon}(t_k) = y(t_k) - \hat{y}(t_k, \hat{\theta}^\eta) \tag{16}$$

where $\bar{\varepsilon}(t_k)$ and $\bar{y}(t_k)$ are the mean values of the estimated noise and the measured output signal of the system, $\hat{y}(t_k, \hat{\theta}^\eta)$ is the estimated output signal, and $\hat{\theta}^\eta$ is the estimated parameter vector given by the SRIVC algorithm.

The YIC is then given by

$$YIC = \ln \left(\frac{\sigma_{\hat{\varepsilon}}^2}{\sigma_y^2} \right) + \ln \left(\frac{1}{n_\theta} \sum_{i=1}^{n_\theta} \frac{\hat{\pi}_{ii}}{\hat{\theta}_i^2} \right) \tag{17}$$

where n_θ is the number of estimated parameters, $\hat{\pi}_{ii}$ is the ith diagonal element of the covariance matrix of the estimated parameter vector $\hat{\theta}^\eta$ [32].

4. Data pre-processing

To estimate the parameters of different possible models, we have an input–output dataset comprising 262165 samples, sampled at $20\,kHz$, corresponding to the pronunciation of the vowel a five times with temporal separation. It is worth mentioning that high-frequency signal processing can be computationally expensive and, in some cases, unnecessary for accurate system identification. Therefore, down-sampling the high-frequency signals to a lower sampling rate can be beneficial [52].

Considering the impact of high-frequency data on parametric estimation, we resample the ACC and GVV signals to a frequency of $F_{sub} = 8192$ Hz. A similar down-sampling strategy has been applied, for example, in Espinoza et al. [15]. The power spectrum of the resampled data is shown in Fig. 2. The analysis of the power spectrum is crucial for understanding signal characteristics such as dominant frequencies, energy distribution, and gain. Moreover, it may enhance experimental identification results.

Furthermore, we opted for the second vocalization from the resampled dataset, comprising 2000 time data points, with the assumption that the first vocalization may exhibit a greater transient effect. This choice was made to enhance the accuracy and fit of our experimental identification results.

The resampled GVV and ACC signals are shown in Figs. 3 and 4, respectively. In these figures, the data points in red belong to down-sampled signal, and the blue lines represent the original signal sampled at high-frequency (20 kHz).

Previous studies in the literature assume a linear model for the subglottal system, neglecting the impact of frequency-dependent resistances in the physically based model since they do not significantly impact the overall behavior of the system [18]. In this section, we test this assumption using the coherence method. The analysis shows that the assumption of linearity between the measurements of ACC and GVV could be, in fact valid in the frequency bandwidth of interest.

Spectral coherence is a powerful signal processing tool, that has been applied to analyze the relationship between two random signals or processes [53,54], and has extensively used in diverse fields, including communication systems, acoustics, and biomedical engineering [55]. Spectral coherence measures the degree of linear correlation between two given signals in the frequency domain, providing a coherence value spanning from 0 (indicating no correlation) to 1 (indicating perfect correlation).

The coherence between two signals x and y is defined by the following quotient

$$C_{(x,y)}(f) = \frac{\left| P_{(x,y)}(f) \right|^2}{P_{(x,x)}(f) P_{(y,y)}(f)} \quad \in [0,1] \tag{18}$$
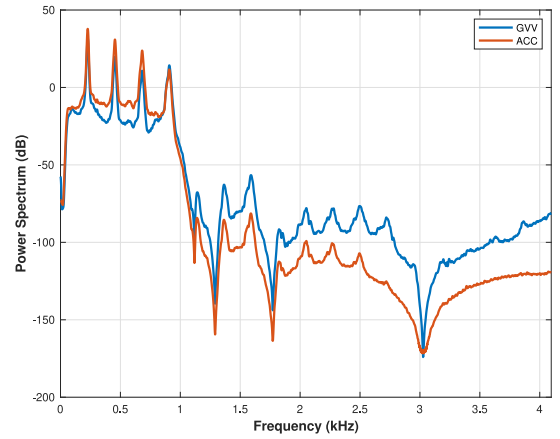

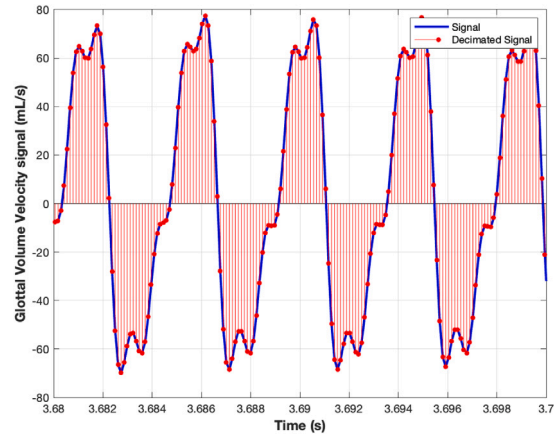**Fig. 2.** Input–output resampled dataset power spectrum.

Fig. 3. Glottal volume velocity (GVV) original and resampled signal.
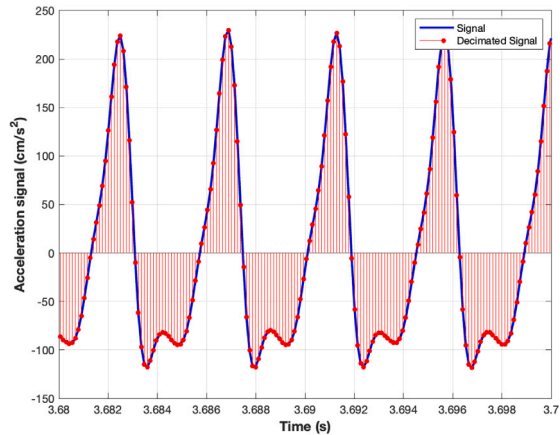

Fig. 4. Acceleration (ACC) signal original and resampled signal.

where $P_{(x,x)}(f)$ and $P_{(y,y)}(f)$ are the power spectral densities of each signal, and $P_{(x,y)}(f)$ is the cross power spectral density between two signals.

The coherence function provides a direct and independent measure of system excitation, data quality, and system response linearity [56]. Poor coherence value can be attributed to either a poor signal to noise ratio (which is not the case in our application) or to nonlinear effects in the dynamics. Therefore, good coherence data is important for determination the linearity assumption of the system. The input–output
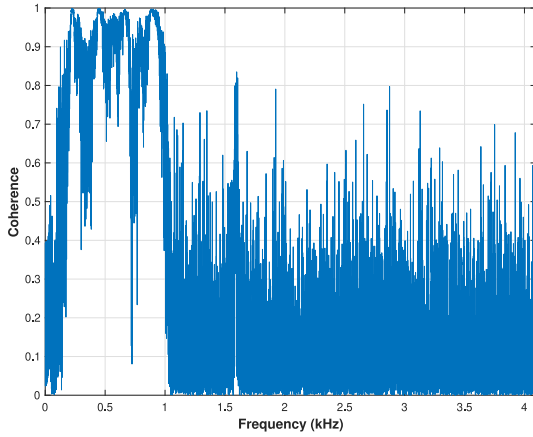
Fig. 5. Coherence between ACC and GVV.

Fig. 6. Heatmap of AIC coefficients.

Fig. 7. Heatmap of BIC coefficients.

Fig. 8. Heatmap of YIC coefficients.

data can then be used to estimate the parameters of the selected linear model.

In our case, the signals utilized for coherence analysis are the accelerometer measurements ACC and the glottal airflow GVV, which are the same signals used to perform the model parameter estimation. The results are presented in Fig. 5, which shows the coherence between the two signals as a function of frequency. It can be noticed that, in the frequency range from approximately 220 Hz to 1kHz, the average coherence value is higher than 0.7. This suggests that the signals are highly correlated in that frequency range, and a linear model may be appropriate. This result is closely related to the graph of the power spectrum of the signals shown in Fig. 2, where the peaks appear within the same frequency bandwidth.

5. Continuous time model identification

In this section we present the identification results obtained for the subglottal system using the SRIVC estimator.

For the system identification procedure, the input–output data are of the glottal airflow signal GVV and the accelerometer measurement signal ACC, which have already been resampled, as explained in Section 4.

In this study, system identification is conducted for a range of candidate models, for different model orders and relative degrees. Then we have to consider the trade-off between the flexibility of the estimated model that may overfit the training data and having a parsimonious model that may offer a more concise representation [57]. Parameter estimation is performed for all possible combinations of candidate models, using coefficients ranging from 1 to 10 in both the denominator and numerator. Then the SRIVC algorithm is applied for each of them to obtain parameter estimates.

The AIC, BIC, and YIC information criteria are utilized to compare the performance of each candidate model. These criteria provide a systematic approach to assess the trade-off between model complexity and goodness of fit, allowing for the selection of the best model among the candidates. Figs. 6, 7, 8 show the AIC, BIC, and YIC matrices, respectively, obtained when continuous-time system identification is performed using SRIVC.

From the information criteria shown in Figs. 6, 7 and 8 it can be noticed that the three information criteria suggest that the best model structure is the one having 5 poles and 4 zeros.

In order to complement the analysis provided by the information criteria, we use the root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y(t_k) - \hat{y}(t_k, \hat{\theta}^\eta))^2} \qquad (19)$$
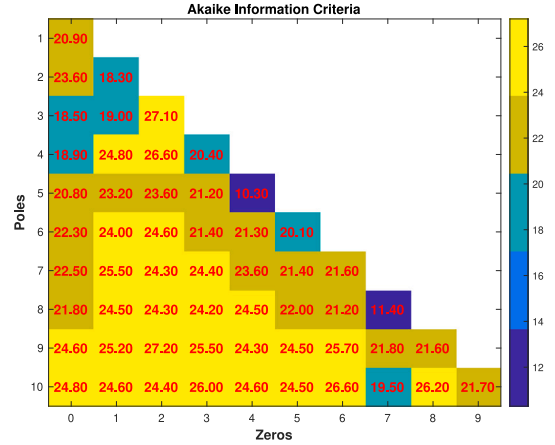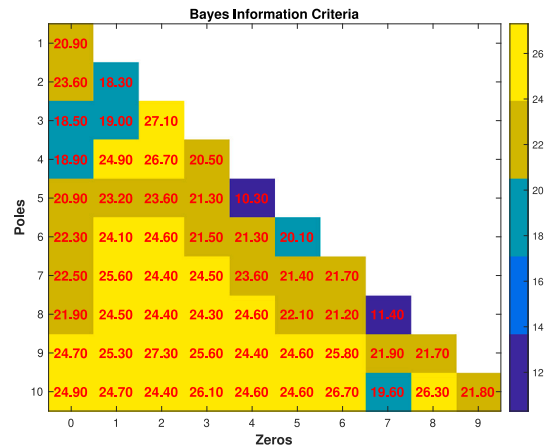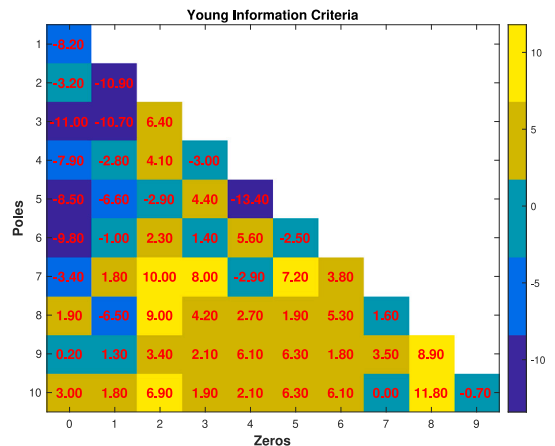
where N is the number of samples.

The RMSE between the measured ACC signal and the simulated ACC signal is computed for the group of candidate models (see Fig. 9). In this figure it can be seen that the structure with 5 poles and 4 zeros has the lowest RMSE value, which corresponds to the same structure selected by the three information criteria.

Moreover, Fig. 10 shows the coefficient of determination obtained for the estimated models. It can be noticed that the highest value of the
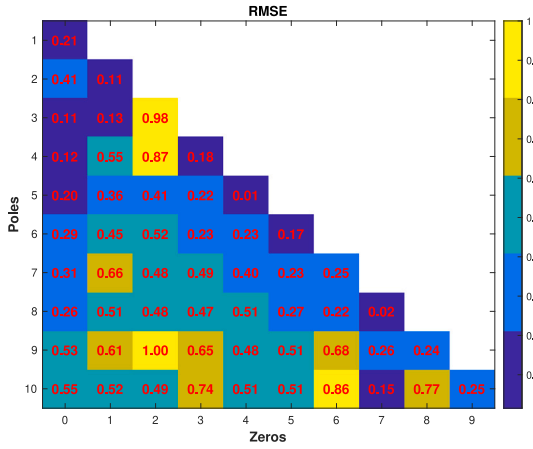
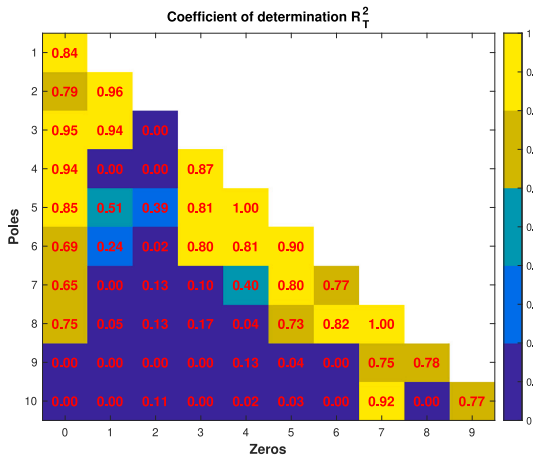**Fig. 9.** Root mean squared error of the candidates models.

Fig. 10. Heatmap of R_T^2 coefficients.

Fig. 11. Poles and zeros map of the selected model estimated using SRIVC.

Fig. 12. Bode diagram of the selected model estimated using SRIVC.

Fig. 13. Comparison of the estimated model with SRIVC and resampled data simulation against validation data.

coefficient $R_T^2 = 1$ is obtained for the same model structure selected by the information criteria, i.e., the model having 5 poles and 4 zeros.

From the analysis above, the choice of the best model structure is clear since information criteria, RMSE, and coefficient of determination provide the same result, namely, an LTI model for the subglottal system having 5 poles and 4 zeros. Moreover, the analysis above also shows that the model structure with 8 poles and 7 zeros, could also be considered, however, at the expense of more complexity and possible overfitting.

Additionally, Figs. 11 and 12 show the pole/zero map and Bode magnitude plot of the selected model (with 5 poles and 4 zeros), and Fig. 13 shows the comparison between the simulated response corresponding to the selected model and the measured output (i.e. the accelerometer signal ACC). In that figure, the comparison between the measured acceleration validation data and the simulated acceleration data results in a model fit of 93.9% computed by:

$$FIT = \left(1 - \frac{\|y(t_k) - \hat{y}(t_k, \hat{\theta}^\eta)\|}{\|y(t_k) - \bar{y}(t_k)\|}\right)100\% \qquad (20)$$

6. Inverse filtering

In the previous section, system identification was performed applying SRIVC and information criteria were used to select a model order and structure that provides a good fit between simulated and observed acceleration. However, a key objective of obtaining a model
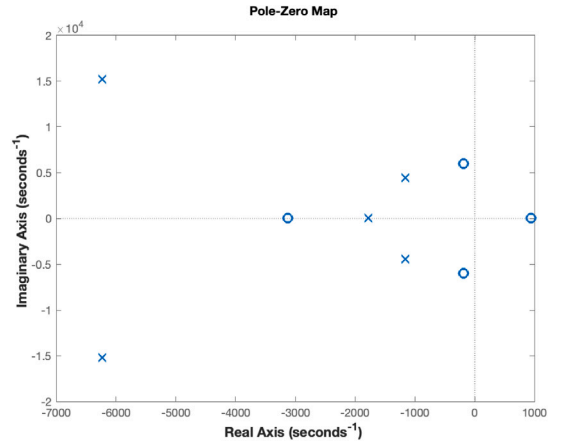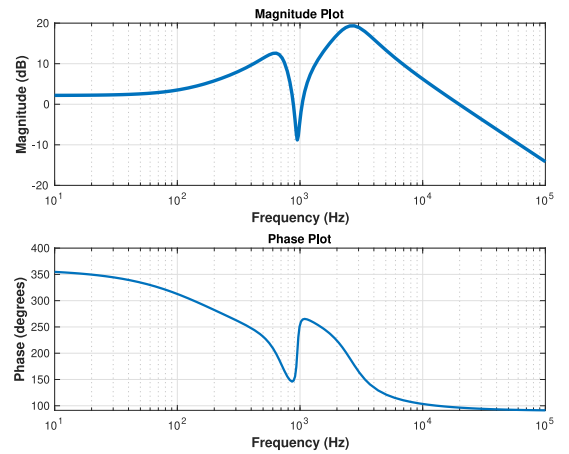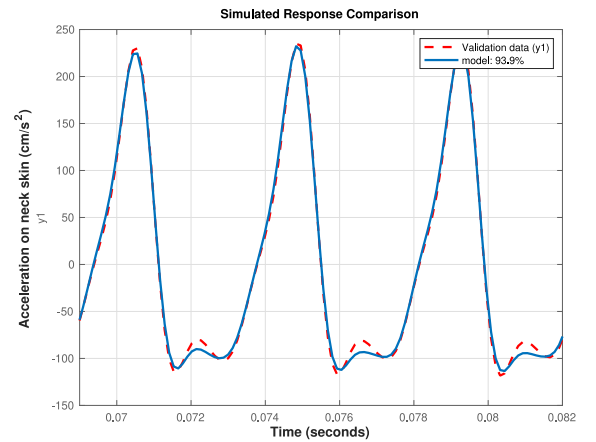
for the subglottal system is to apply it for inverse filtering, i.e., to estimate the glottal airflow (GVV, the input signal) from accelerometer measurements (ACC, the output signal). In fact, in ambulatory studies such as [18], the estimated glottal airflow is used to obtain indicators of the speech health of the patients.

In this work, inverse filtering is performed applying a Kalman smoother strategy as presented in Morales et al. [58]. A key difference
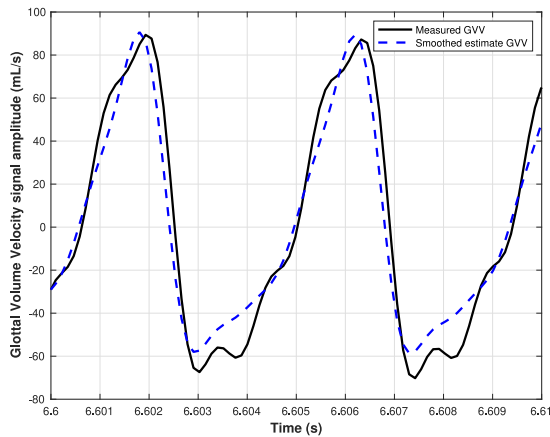
Fig. 14. GVV estimation with Kalman Smoother and SRIVC estimated model.

Table 1
Aerodynamic metrics computed from GVV.

Measures \ Model	GVV_ACC	GVV_OVV
H1-H2 (dB)	8.43 ± 2.5	9.82 ± 3.7
HRF (dB)	7.19 ± 3.3	7.98 ± 4.4
MFDR (L/s^2)	142.1 ± 83.8	178.37 ± 107.53
ACFL (mL/s)	97.99 ± 56.2	112.07 ± 65.7
NAQ	0.15 ± 0.03	0.15 ± 0.04

with that work is that the linear models obtained in that paper for the subglottal system were based in a frequency response of an impedance-based model which parameters were previously fitted by particle swarm optimization.

The results of the Kalman smoother estimation of the glottal airflow (using the optimal model obtained by SRIVC) are shown in Fig. 14. In this figure, the GVV signal derived from OVV and Kalman filtered GVV signal from the SRIVC estimated model can be compared. The associated RMSE obtained is 10.62×10^{-3} mL/s.

To be able to assess the validity of the estimated glottal airflow, acoustic and aerodynamic measures obtained from GVV are used. These measures consider the first two harmonics (H1-H2), the harmonic richness factor (HRF), the maximum flow declination rate (MFDR), the AC flow (ACFL), and the normalized amplitude quotient (NAQ) [59].

Table 1 shows the aerodynamic features computed using the *measured* GVV (that is actually derived from OVV) and the aerodynamic features computed using the estimated GVV obtained from ACC measurements by Kalman filtering. In the table, these two sets of features are labeled GVV_OVV and GVV_ACC, respectively.

Moreover, aerodynamic features shown in Table 1 obtained from ACC are similar to previous results in the literature when applying Kalman filtering, such as, for example Table 2 in Cortés et al. [60] and Table 4 in Morales et al. [58]. However, compared to those previous results, in this study we have used a lower-order model (5 poles and 4 zeros) estimated directly from time-domain data. This implies a reduction in model order of 77% (5/22) against recent efforts [58] and 98.6% (5/350) against previous ones [60]. The clear reduction in the model order for the subglottal system certainly leads to an important reduction in the computational time and complexity, which are key issues for real-time monitoring applications.

7. Conclusions

In this paper a continuous-time, linear time-invariant models were obtained for the subglottal system. The identification process employed sampled data, encompassing both neck skin acceleration and glottal airflow. Utilizing the SRIVC estimator, system identification was performed for a range of candidate models. Selection of the optimal parsimonious model was based on information criteria, and further verified by RMSE and coefficient of determination analysis. Our results indicate that the SRIVC algorithm effectively estimates a low-order model, achieving a high level of fit. Moreover, this model was applied in a Kalman smoother for inverse filtering, enabling the extraction of glottal airflow estimates based on neck skin acceleration measurements. The approach allows us to estimate glottal aerodynamic features using a lower order model compared to previous studies in the literature, offering a novel alternative for real-time ambulatory assessment of vocal function.

CRediT authorship contribution statement

Javier G. Fontanet: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Juan I. Yuz:** Formal analysis, Investigation, Supervision, Writing – review & editing, Methodology, Funding acquisition. **Hugues Garnier:** Formal analysis, Software, Supervision, Writing – review & editing. **Arturo Morales:** Software, Writing – review & editing. **Juan Pablo Cortés:** Data curation, Writing – review & editing. **Matías Zañartu:** Data curation, Formal analysis, Supervision, Writing – review & editing, Funding acquisition, Project administration, Resources.

Declaration of competing interest

• Javier G. Fontanet reports a relationship with ANID that includes: funding grants. Javier G. Fontanet reports a relationship with Federico Santa Maria Technical University that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
• Juan I. Yuz reports a relationship with ANID that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
• Juan Pablo Cortes reports financial support was provided by Lanek SPA. Matias Zanartu reports financial support was provided by Lanek SPA. Juan Pablo Cortes reports a relationship with Lanek SPA that includes: employment. Matias Zanartu reports a relationship with Lanek SPA that includes: equity or stocks. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
The co-authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] G. Fant, Acoustic theory of speech production, in: Acoustic Theory of Speech Production, 1971, http://dx.doi.org/10.1515/9783110873429.

[2] K.N. Stevens, Acoustic phonetics, in: Acoustic Phonetics, 2000, http://dx.doi.org/10.7551/mitpress/1072.001.0001.

[3] S.M. Lulich, A. Bachrach, N. Malyska, A role for the second subglottal resonance in lexical access, J. Acoust. Soc. Am. 122 (4) (2007) http://dx.doi.org/10.1121/1.2772227.

[4] I.R. Titze, Nonlinear source–filter coupling in phonation: Theory, J. Acoust. Soc. Am. 123 (5) (2008) http://dx.doi.org/10.1121/1.2832337.

[5] I. Titze, T. Riede, P. Popolo, Nonlinear source–filter coupling in phonation: Vocal exercises, J. Acoust. Soc. Am. 123 (4) (2008) http://dx.doi.org/10.1121/1.2832339.

[6] J.C. Ho, M. Zañartu, G.R. Wodicka, An anatomically based, time-domain acoustic model of the subglottal system for speech production, J. Acoust. Soc. Am. 129 (3) (2011) 1531–1547, http://dx.doi.org/10.1121/1.3543971.

[7] M. Zañartu, L. Mongeau, G.R. Wodicka, Influence of acoustic loading on an effective single mass model of the vocal folds, J. Acoust. Soc. Am. 121 (2) (2007) http://dx.doi.org/10.1121/1.2409491.

[8] Z. Zhang, J. Neubauer, D.A. Berry, The influence of subglottal acoustics on laboratory models of phonation, J. Acoust. Soc. Am. 120 (3) (2006) http://dx.doi.org/10.1121/1.2225682.

[9] X. Chi, M. Sonderegger, Subglottal coupling and its influence on vowel formants, J. Acoust. Soc. Am. 122 (3) (2007) http://dx.doi.org/10.1121/1.2756793.

[10] S.M. Lulich, Subglottal resonances and distinctive features, J. Phonetics 38 (1) (2010) 20–32, http://dx.doi.org/10.1016/j.wocn.2008.10.006.

[11] K. Verdolini, C. Rosen, R.C. Branski, et al., Classification manual for voice disorders-I, American Speech-Language Hearing Division. Mahwah, NJ: Erlbaum, Psychology Press, 2006.

[12] N. Roy, R.M. Merrill, S.D. Gray, E.M. Smith, Voice disorders in the general population: Prevalence, risk factors, and occupational impact, The Laryngoscope 115 (11) (2005) 1988–1995.

[13] N. Bhattacharyya, The prevalence of voice problems among adults in the United States, Laryngoscope 124 (10) (2014) 2359–2362, http://dx.doi.org/10.1002/lary.24740.

[14] R.E. Hillman, E.B. Holmberg, J.S. Perkell, M. Walsh, C. Vaughan, Objective assessment of vocal hyperfunction: An experimental framework and initial results, J. Speech Lang. Hearing Res. 32 (2) (1989) 373–392.

[15] V.M. Espinoza, M. Zañartu, J.H. Van Stan, D.D. Mehta, R.E. Hillman, Glottal aerodynamic measures in women with phonotraumatic and nonphonotraumatic vocal hyperfunction, J. Speech Lang. Hearing Res. 60 (8) (2017) 2159–2169, http://dx.doi.org/10.1044/2017-JSLHR-S-16-0337.

[16] M. Rothenberg, New inverse-filtering technique for deriving the glottal air flow waveform during voicing, J. Acoust. Soc. Am. 48 (1A_Supplement) (1970) http://dx.doi.org/10.1121/1.1975066.

[17] H.A. Cheyne, Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck, in: Proceedings of the 3rd IEEE-EMBS International Summer School and Symposium on Medical Devices and Biosensors, ISSS-MDBS 2006, 2006, pp. 118–121, http://dx.doi.org/10.1109/ISSMDBS.2006.360113.

[18] M. Zañartu, J.C. Ho, D.D. Mehta, R.E. Hillman, G.R. Wodicka, Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration, IEEE Trans. Audio Speech Lang. Process. 21 (9) (2013) 1929–1939.

[19] D.D. Mehta, J.H. Van Stan, M. Zañartu, M. Ghassemi, J.V. Guttag, V.M. Espinoza, J.P. Cortés, H.A. Cheyne, R.E. Hillman, Using ambulatory voice monitoring to investigate common voice disorders: Research update, Front. Bioeng. Biotechnol. 3 (OCT) (2015) 1–14, http://dx.doi.org/10.3389/fbioe.2015.00155.

[20] J.P. Cortés, V.M. Espinoza, M. Ghassemi, D.D. Mehta, J.H. Van Stan, R.E. Hillman, J.V. Guttag, M. Zañartu, Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration, PLoS ONE 13 (12) (2018) 1–22, http://dx.doi.org/10.1371/journal.pone.0209017.

[21] D.D. Mehta, M. Zañartu, S.W. Feng, H.A. Cheyne, R.E. Hillman, Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform, IEEE Trans. Biomed. Eng. 59 (12 PART2) (2012) 3090–3096, http://dx.doi.org/10.1109/TBME.2012.2207896.

[22] G.R. Wodicka, K.N. Stevens, H.L. Golub, E.G. Cravalho, D.C. Shannon, A model of acoustic trnsmission in the respirtor system, IEEE Trans. Biomed. Eng. 36 (9) (1989) http://dx.doi.org/10.1109/10.35301.

[23] P. Harper, S.S. Kraman, H. Pasterkamp, G.R. Wodicka, An acoustic model of the respiratory tract, IEEE Trans. Biomed. Eng. 48 (5) (2001) http://dx.doi.org/10.1109/10.918593.

[24] V.P. Harper, H. Pasterkamp, H. Kiyokawa, G.R. Wodicka, Modeling and measurement of flow effects on tracheal sounds, IEEE Trans. Biomed. Eng. 50 (1) (2003) http://dx.doi.org/10.1109/TBME.2002.807327.

[25] J.S. Perkell, R.E. Hillman, E.B. Holmberg, Group differences in measures of voice production and revised values of maximum airflow declination rate, J. Acoust. Soc. Am. 96 (2) (1994) 695–698.

[26] 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 36 (5) (2011) 623–650.

[27] T. Drugman, P. Alku, A. Alwan, B. Yegnanarayana, Glottal source processing: From analysis to applications, Comput. Speech Lang. 28 (5) (2014) 1117–1138.

[28] B. Henry, T.J. Royston, A multiscale analytical model of bronchial airway acoustics, J. Acoust. Soc. Am. 143 (3) (2018) http://dx.doi.org/10.1121/1.5027239.

[29] N. Hanna, J. Smith, J. Wolfe, How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips, J. Acoust. Soc. Am. 143 (5) (2018) http://dx.doi.org/10.1121/1.5033330.

[30] T. Söderström, P. Stoica, System Identification, Prentice Hall International, 1989.

[31] L. Ljung, System Identification: Theory for the User, Second ed., Pearson, 1998.

[32] H. Garnier, L. Wang, Identification of Continuous-time Models from Sampled Data, Springer, 2008, p. 429.

[33] H. Garnier, M. Gilson, P.C. Young, E. Huselstein, An optimal IV technique for identifying continuous-time transfer function model of multiple input systems, Control Eng. Pract. 15 (4) (2007) 471–486, http://dx.doi.org/10.1016/j.conengprac.2006.09.004.

[34] H. Garnier, P.C. Young, The advantages of directly identifying continuous-time transfer function models in practical applications, Internat. J. Control 87 (7) (2014) 1319–1338, http://dx.doi.org/10.1080/00207179.2013.840053.

[35] H. Garnier, Direct continuous-time approaches to system identification. Overview and benefits for practical applications, Eur. J. Control 24 (2015) 50–62, http://dx.doi.org/10.1016/j.ejcon.2015.04.003.

[36] K.P. Burnham, D.R. Anderson, Multimodel inference: A practical information-theoretic approach, in: Sociological Methods and Research, 2004.

[37] P.S. Popolo, J.G. Švec, I.R. Titze, Adaptation of a pocket PC for use as a wearable voice dosimeter, J. Speech Lang. Hearing Res. 48 (4) (2005) 780–791, http://dx.doi.org/10.1044/1092-4388(2005/054).

[38] P. Alku, C. Magi, S. Yrttiaho, T. Bäckström, B. Story, Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering, J. Acoust. Soc. Am. 125 (5) (2009) http://dx.doi.org/10.1121/1.3095801.

[39] G.P. Kafentzis, Y. Stylianou, P. Alku, Glottal inverse filtering using stabilised weighted linear prediction, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2011, pp. 5408–5411, http://dx.doi.org/10.1109/ICASSP.2011.5947581.

[40] H. Garnier, M. Mensler, A. Richard, Continuous-time model identification from sampled data: Implementation issues and performance evaluation, Internat. J. Control 76 (13) (2003) 1337–1357, http://dx.doi.org/10.1080/0020717031000149636.

[41] V. Pascu, H. Garnier, L. Ljung, A. Janot, Benchmark problems for continuous-time model identification: Design aspects, results and perspectives, Automatica 107 (2019) http://dx.doi.org/10.1016/j.automatica.2019.06.011.

[42] P.C. Young, Recursive estimation and time-series analysis, in: Paper Knowledge . Toward a Media History of Documents. no. 2, second ed., Springer Science & Business Media, London, 2012, http://dx.doi.org/10.1007/978-3-642-21981-8.

[43] S. Pan, R.A. González, J.S. Welsh, C.R. Rojas, Consistency analysis of the simplified refined instrumental variable method for continuous-time systems, in: Automatica, vol. 113, 2020, http://dx.doi.org/10.1016/j.automatica.2019.108767.

[44] P. Young, A. Jakeman, Refined instrumental variable methods of recursive time-series analysis: Multivariable systems, Internat. J. Control 31 (4) (1980) 741–764, http://dx.doi.org/10.1080/00207177908922724.

[45] R.A. González, C.R. Rojas, S. Pan, J.S. Welsh, On the relation between discrete and continuous-time refined instrumental variable methods, IEEE Control Syst. Lett. 7 (2023) 2233–2238, http://dx.doi.org/10.1109/LCSYS.2023.3282445, URL https://ieeexplore.ieee.org/document/10143357/.

[46] A.P. Liavas, P.A. Regalia, On the behavior of information theoretic criteria for model order selection, IEEE Trans. Signal Process. 49 (8) (2001) 1689–1695.

[47] P. Stoica, Y. Selén, Model-order selection: A review of information criterion rules, IEEE Signal Process. Mag. 21 (4) (2004) 36–47.

[48] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control 19 (6) (1974) 716–723, http://dx.doi.org/10.1109/TAC.1974.1100705.

[49] P. Young, Recursive estimation, forecasting, and adaptive control, Control Dyn. Syst. 30 (1989) 119–165, http://dx.doi.org/10.1016/B978-0-12-012730-6.50011-0.

[50] V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire, H. Garnier, A new data-based modelling method for identifying parsimonious nonlinear rainfall/flow models, in: Modelling for Environment's Sake: Proceedings of the 5th Biennial Conference of the International Environmental Modelling and Software Society, iEMSs 2010, vol. 3, 2010, pp. 2044–2052.

[51] P. Young, H. Garnier, Identification and estimation of continuous-time rainfall-flow models, IFAC Proc. Vol. 39 (1) (2006) http://dx.doi.org/10.3182/20060329-3-au-2901.00206.

[52] N.K. Rout, D.P. Das, G. Panda, Computationally efficient algorithm for high sampling-frequency operation of active noise control, Mech. Syst. Signal Process. 56 (2015) http://dx.doi.org/10.1016/j.ymssp.2014.10.009.

[53] A. Gómez González, J. Rodríguez, X. Sagartzazu, A. Schuhmacher, I. Isasa, Multiple coherence method in time domain for the analysis of the transmission paths of noise and vibrations with non stationary signals, in: Proceedings of ISMA 2010 - International Conference on Noise and Vibration Engineering, Including USD 2010, 2010, pp. 3927–3942.

[54] A. Klein, T. Sauer, A. Jedynak, W. Skrandies, Conventional and wavelet coherence applied to sensory-evoked electrical brain activity, IEEE Trans. Biomed. Eng. 53 (2) (2006) http://dx.doi.org/10.1109/TBME.2005.862535.

[55] P. Stoica, R.L. Moses, Spectral analysis of signals, Pearson/Prentice Hall, 2005, p. 452.

[56] R. Johansson, System Modeling and Identification, Prentice-Hall, 1993.

[57] R.A. González, C. Rojas, S. Pan, J.S. Welsh, Parsimonious identification of continuous-time systems: A block-coordinate descent approach, 2023, arXiv preprint arXiv:2304.03259.

[58] A. Morales, J.I. Yuz, J.P. Cortés, J.G. Fontanet, M. Zañartu, Glottal airflow estimation using neck surface acceleration and low-order Kalman smoothing, IEEE/ACM Trans. Audio Speech Lang. Process. 31 (2023) 2055–2066, http://dx.doi.org/10.1109/TASLP.2023.3277269.

[59] E.B. Holmberg, P. Doyle, J.S. Perkell, B. Hammarberg, R.E. Hillman, Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy, J. Voice 17 (3) (2003) http://dx.doi.org/10.1067/S0892-1997(03)00076-6.

[60] J.P. Cortés, G.A. Alzamendi, A.J. Weinstein, J.I. Yuz, V.M. Espinoza, D.D. Mehta, R.E. Hillman, M. Zañartu, Kalman filter implementation of subglottal impedance-based inverse filtering to estimate glottal airflow during phonation, Appl. Sci. (Switzerland) 12 (1) (2022) 1–20, http://dx.doi.org/10.3390/app12010401.