Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 54-7 (2021) 577–582

Parametric Identification of a Linear Time Invariant Model for a Subglottal System ⋆

Javier G. Fontanet *,** **Juan I. Yuz** * **Matías Zañartu** *

* *Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123 Chile (e-mail: javier.gonzalezf@sansano.usm.cl, juan.yuz@usm.cl, matias.zanartu@usm.cl).*

** *Department of Automatic Engineering, Universidad de Oriente, Santiago de Cuba, 90500 Cuba.*

Abstract: Models of the human body are key in bio-engineering and medical applications. This study presents the identification, in time and frequency domains, of linear time invariant models of the human subglottal system, for the clinical assessment of vocal function. For time domain identification, the input-output data corresponds to the glottal volume velocity and the acceleration registered by a sensor on the neck skin of the patient. For frequency domain identification, the frequency response of the subglottal tract module is used. Maximum likelihood and prediction error methods are applied. Additionally, the Akaike and Bayes Information Criteria are used to select the models order.

Copyright © 2021 The Authors. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

Keywords: Maximum Likelihood estimator, prediction error methods, identification algorithm, parameter estimation, frequency response.

1. INTRODUCTION

According to medical specialists, common voice disorders, known as vocal hyperfunction (VH) are associated with speech abuse and misuse (Verdolini et al. (2006)). According to Mehta et al. (2012) and Bhattacharyya (2014) approximately 6.6% of the U.S adult population suffers from this disease, showing the relevance of studying new ways of evaluating the health status of the vocal folds. The modeling and simulation of the vocal folds are also important for the design of voice synthesizers and the study of mechanical vibrations, (see Story and Titze (1995)).

Several works have studied the differences between normal vocal function and VH, (Hillman et al. (1989); Espinoza et al. (2017)). Most of these results are based on the use of a circumferentially vented pneumotachograph mask to measure the oral airflow. This airflow is inverse filtered, in order to reduce the influence of vocal tract, estimating parameters of the glottal airflow waveform, such as peak-to-peak AC flow, open quotient and maximum flow declination rate (see Hillman et al. (1989), Perkell et al. (1994), Alku (2011), Drugman et al. (2014)).

Other works show a non-invasive technique to characterize VH by using an accelerometer to measure the vibrations of vocal folds due to speech, (see Cheyne (2006); Mehta et al. (2012); Zañartu et al. (2013); Espinoza Catalán and Zañartu (2014); Cortés et al. (2018)). This research is supported by the fact that glottal aerodynamic measures of subglottal air pressure, and glottal airflow can be used to identify phonatory mechanisms associated with vocal hyperfunction that are different from normal vocal function, (see Espinoza et al. (2017)).

In Zañartu et al. (2013); Mehta et al. (2015) and Cortés et al. (2018), a transfer function model is obtained to use aerodynamic parameters in an ambulatory setting with a neck surface accelerometer, for the evaluation, monitoring and treatment of VH. A particularly difficult problem is to estimate the subject-specific parameters of the subglottal system.

Although several models developed for the vocal folds are non-linear, simpler linear models have been also successfully used, for example, obtaining an impulse response by inverse Fourier transform of frequency domain response, (Cortés et al. (2018)). The linear simplification is reasonable given that the nonlinear terms in the subglottal systems are associated with frequency dependent resistances that have a minor overall effect (Zañartu et al. (2013)). In this paper linear time invariant (LTI) models are proposed to represent vocal folds, that are fitted by using maximum likelihood (ML) and prediction error method PEM. The use of linear models may also be motivated by the need to perform ambulatory studies of speech health, aimed at obtaining the glottal airflow using Kalman filtering. In this paper, the identified parameters correspond to three different model structures: Box-Jenkins (BJ), output−error (OE) and state space (SS). To determine the order of each model structure the AIC and BIC information criteria are applied.

The structure of this paper is as follows: Section 2 presents the description of the subglottal system, the usual way to

⋆ This work was supported by the Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008, ANID; and the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders through Grant P50DC015446. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

2405-8963 Copyright © 2021 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2021.08.422

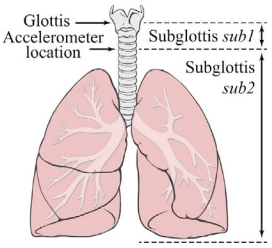

Fig. 1. Representation of subglottal system, Zañartu et al. (2013)

fit a parametric model, and the impedance based frequency response. Section 3 presents the maximum likelihood algorithm and Akaike and Bayes information criteria. Section 4 shows the results of the estimation of model parameters, comparisons between model fit, and the order selection obtained from the information criteria. Finally, in Section 5 we draw some conclusions.

2. SYSTEM DESCRIPTION

In the study of VH, works such as Mehta et al. (2015); Cortés et al. (2018), propose to measure the acceleration of the neck skin generated by the air flow in the glottis (see Fig. (1)). From the acceleration measurements, inverse filtering is carried out. Subglottal impedance based inverse filtering (IBIF) was proposed to obtain an accurate estimation of the aerodynamic source of voice sounds at the glottis in Zañartu et al. (2013). In that paper, a scheme based on mechano−acoustic impedance representations from a physiologically-based transmission line model is used. The IBIF, used as a signal processing tool, provides an estimation of the glottal airflow from an accelerometer placed on the skin over the trachea. From the estimated airflow, specialists are then able to determine certain conditions of the vocal folds, essential for speech.

Considering the individual characteristics of each subject under study, Zañartu et al. (2013); Cortés et al. (2018), highlight the need of IBIF calibration, estimating subject−specific parameters. The parameters to adjust are: the neck resistance (R_m), accelerometer area density on neck skin (M_m), skin stiffness (K_m); the length of the trachea ($L_{trachea}$) and accelerometer placement (L_{sub1}). Calibration of these parameters is performed based on five scaling factors Q_i. According to Cortés et al. (2018), accurate Q_i factors, allow to filter neck skin and subglottal resonances ensuring comparison of the glottal airflow between different patients. To obtain these subject specific parameters, the airflow measured using a pneumotachograph mask is used to find the optimal parameters using Particle Swarm Optimization Algorithm (PSO) for and ad-hoc cost function.

The phonatory system can be represented by the block diagram shown in Fig. (2); *ovv* (oral volume velocity) is obtained from laboratory measurements with a pneumotachograph mask, and *acc* (acceleration) with an accelerometer on the neck skin. $H_1(q)$ represents the oral cavity filter, and $H(q)$ the noise filter, while the subglottal system to be identified and the initial condition are represented by $G(q)$ and x_0, respectively. Once $G(q)$ is parameterized from estimation with synchronized data *ovv* and *acc*, it is possible then to perform ambulatory monitoring of the

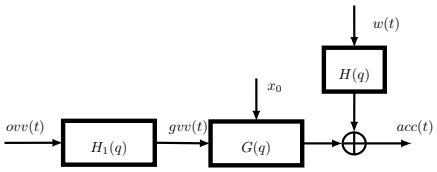

Fig. 2. Block diagram representation of the phonatory system

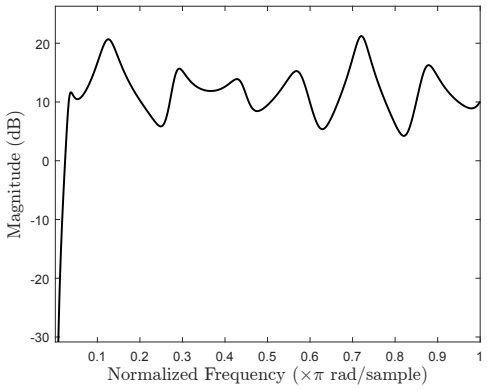

Fig. 3. Frequency response of subglottal system, Zañartu et al. (2013)

subjects under medical study based on the accelerometer signal only.

2.1 Frequency response of the subglottal model

System frequency response is important for system identification, model validation and analysis, (Fragnière and Wartmann (2015)). This has been also used as a tool for robust modeling algorithms and to adjust low complexity models, (see Agüero et al. (2012)).

To obtain the coefficients of dynamic models, the frequency response can be used to adjust an algebraic expression, (see Levy (1959); Deschrijver and Dhaene (2006)). In the current work, the frequency response of the subglottal system is used to perform system identification in the frequency domain. Fig. 3 shows the frequency response of a subglottal model numerically obtained in Zañartu et al. (2013). We are interested in a model that can be represented by:

$$acc(t) = \frac{B(q)}{F(q)} gvv(t) \tag{1}$$

From the frequency response of the subglottal system and applying Levy's algorithm the coefficients that fit the model from data are obtained, (Levy (1959)).

To obtain the coefficients of the numerator and denominator in (1), an equation error or output error model can be used. The equation error model is fit by:

$$\min_{b,a} \sum_{k=1}^{n} wt(k) \left| h(k)A(w(k)) - B(w(k)) \right|^2 \tag{2}$$

where $A(w(k))$ and $B(w(k))$ are the Fourier transform of the polynomials of the numerator and denominator of the model, $h(k)$ and $w(k)$ are the vectors of frequency response and angular frequencies, and $wt(k)$ are weighting factors.

The output error model fit is given by:

$$\min_{b,a} \sum_{k=1}^{n} wt(k) \left| h(k) - \frac{B(w(k))}{A(w(k))} \right|^2 \tag{3}$$

These coefficients will be used to initialize the parameter estimates that are then identified with ML in the time domain.

3. PARAMETER ESTIMATION AND INFORMATION CRITERIA

In this study three different LTI models are estimated for the subglottal system by performing the identification in time and frequency domains.

In order to estimate the parameters of the subglottal system, in time domain, a Box-Jenkins model structure is selected:

$$y(t) = G(q^{-1}, \theta)u(t) + H(q^{-1}, \theta)w(t) \tag{4}$$

For the subglottal system $u(t)$ refers to $gvv(t)$, the glottal volume velocity, $y(t)$ refers to $acc(t)$, the acceleration of the neck skin due to glottal volume velocity, and $w(t)$ is assumed to be a white noise sequence.

For the identification in the frequency domain, we identify two model structures: a discrete-time output error model and a continuous-time state space model. The discrete-time output error model is given by:

$$y(t) = \frac{B(q)}{F(q)}u(t - nk) + e(t) \tag{5}$$

where nk is the input delay, expressed as the number of samples, and $e(t)$ is the error.

The last model structure to be identified is a continuous-time state space model:

$$\begin{aligned} \dot{x}(t) &= A'x(t) + B'u(t) + Ke(t) \\ y(t) &= C'x(t) + D'u(t) + e(t) \end{aligned} \tag{6}$$

3.1 Maximum likelihood

For the model in (4) we assume $G(0, \theta) = 0$ and $H(0, \theta) = 1$, ensuring at least one input-output delay, and we consider that H can be normalized, including the variance of the noise in the parameter vector.

The estimated vector from the input and output data is $\beta = [\theta^T \ \sigma^2]$, where θ^T corresponds to the polynomial coefficients, and σ^2 is the noise variance, (see Åström (1980); Söderström (2002); Agüero et al. (2010)).

The likelihood function is:

$$L_N(\beta) = p(y_{1:N}|u_{1:N}, \beta) \tag{7}$$

where $p(y_{1:N}|u_{1:N}, \beta)$ is the probability of the observed data $y_{1:N}$ given a realization $u_{1:N}$ and the parameter vector β.

Applying the chain rule and Bayes' theorem (taking into account that the input is deterministic and given), and applying the natural logarithm, the log-likelihood function can be expressed as:

$$l(\beta) = \log\left(p(y_1|u_{1:N}, \beta)\right) + \sum_{t=2}^{N} \log\left[p(y_t|y_{1:t-1}, u_{1:N}\beta)\right] \tag{8}$$

The ML estimator is then given by:

$$\hat{\theta}_{ML} = \arg\min_{\theta} \left\{ \frac{1}{H(q^{-1}, \theta)} \left[y_t - G\left(q^{-1}, \theta\right) u_t \right] \right\} \tag{9}$$

3.2 Information Criteria

Model order selection is an important task in the analysis of time series, signal processing, experimental identification of systems, among others, (see Liavas and Regalia (2001)). To determine the order of subglottal model, information criteria such as Akaike and Bayes (AIC and BIC respectively) can be used .

Akaike Information Criteria. The Akaike's criterion, (Akaike (1974)), is a Kullback-Leibler (KL) cross validation approach, in which the selection of the model order consists of minimizing the KL discrepancy between the true probability distribution function and the likelihood of the model,(Stoica and Selen (2004)). AIC is given by the expression:

$$AIC = -2 \ln p_n(y, \hat{\theta}^n) + 2n \tag{10}$$

where y is the vector of available data of size N, and $\hat{\theta}$ is the n-dimensional estimate of the parameter vector.

The value of (10) is obtained for a set of candidate models, among which we select the one that yields the minimum. In this way, the goodness of fit is rewarded and over−fitting is penalized, (see Akaike (1974)).

Bayes Information Criteria. Bayes' information criteria can be obtained, like AIC, from maximizing KL. It can be obtained using the full Bayesian approach, (see Stoica and Selen (2004)). The BIC is given by:

$$BIC = -2 \ln p_n(y, \hat{\theta}^n) + n \ln N \tag{11}$$

As for the case of AIC, BIC is determined for a set of candidates models, to select the one that minimizes the index (11), taking into account model fit and penalizing over−fitting.

4. ESTIMATION RESULTS

In this section, numerical results for the identification of LTI subglottal models, with order selected from AIC and BIC, are shown. The data used for identification correspond to the signals resulting from the sustained pronunciation of a vowel /a/; gvv is obtained filtering the oral volume velocity (ovv) captured with a pneumotachograph mask and synchronized with the accelerometer data. The data length is 5964 data points and the sampling frequency is $11025Hz$.

4.1 Time domain identification

We estimate the parameters of model in (4) using ML for different model orders of $G(q^{-1}, \theta)$, while $H(q^{-1}, \theta)$

is a first order filter. Table 1 shows the results of the information criteria for several candidate models. It can be seen that the best fit is obtained for order 24, coinciding with the minima of AIC and BIC. Simulations of models with order greater than 40 show no further improvements.

For the identification with ML, the initial parameter estimates were first obtained from the frequency response using the equation error criteria in (2).

Table 1. Information criteria and fit of candidates BJ models.

Criteria \ Order	7	8	9	12	24	35	40
AIC (10^4)	5.6	5.5	5.59	5.6	**4.0**	5.8	5.8
BIC (10^5)	2.4	2.1	2.26	2.3	**1.6**	534	493
fit (%)	22	12	17	20	**25**	13	18

Fig. 4 shows the frequency response of the model of order 24 selected taking into account AIC and BIC. Fig. 5 shows the time response of the estimated model simulated with gvv as input signal. It can be observed, that, despite being the best fit obtained under the above conditions, still shows a low fit in the time domain.

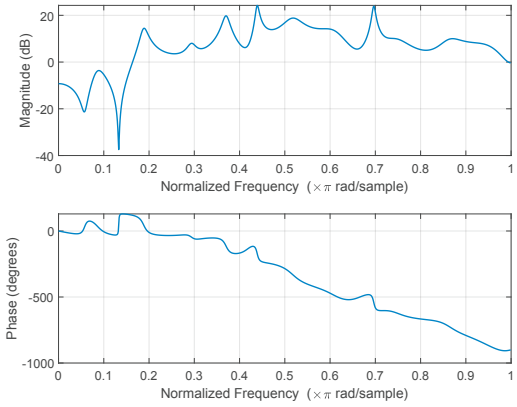

Fig. 4. Frequency response of 24^{th} order estimated model

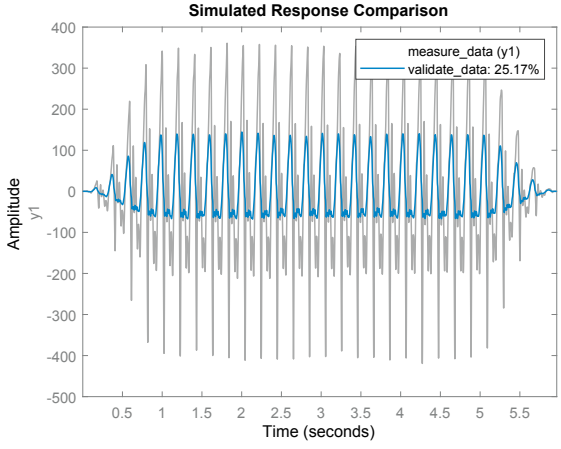

Fig. 5. Time domain response of 24^{th} order estimated model

Table 2 shows AIC and BIC, when ML estimation is initialized with the output error model (3). In this case, the minimum of AIC and BIC does not coincide with the highest fit (NRMSE fitness value). The minimum of both criteria correspond to a 12^{th} order model, meanwhile the highest fit correspond to the order 40. Fig. 6 and Fig. 7 show the frequency and time domain responses of the model with highest fit.

Table 2. Information criteria and fit of candidates models.

Criteria \ Order	7	12	24	35	40	50	60
AIC (10^4)	5.6	**4.0**	4.1	5.5	5.7	5.8	5.8
BIC (10^7)	2.5	**0.01**	2.0	2.1	3.4	5.8	5.0
fit (%)	24	16	30	14	**37**	28	13

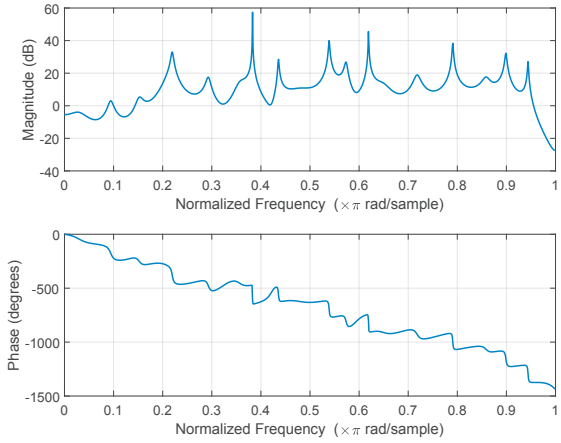

Fig. 6. Frequency response of 40^{th} order estimated model

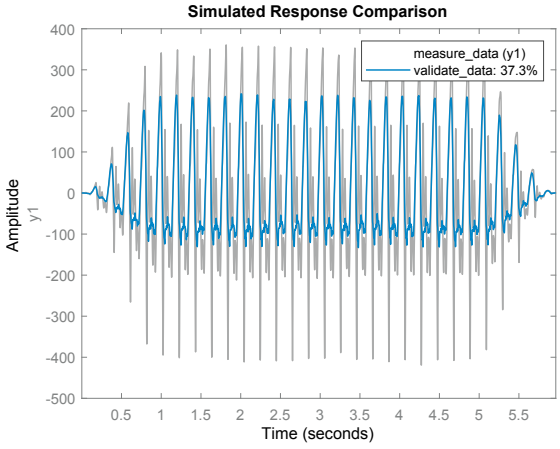

Fig. 7. Time domain response of 40^{th} order estimated model

4.2 Frequency domain identification

This section shows the identification results in the frequency domain. As previously mentioned, the estimated model structures are shown in (5) and (6). The estimation is done using the PEM algorithm in Matlab.

For the identification in the frequency domain, we use as frequency response of the subglottal system the data in Fig. 3 that was obtained numerically simulating a distributed impedance model. Using such data, leads to a frequency fit of model (5) higher than those shown in Tables 1 and 2, but obtaining an unstable model. When stability is introduced as a constraint, the model fit in

the frequency domain decreases (as one would expect). The adjustment percentage results for various candidate models of different orders are shown in Table 3, where the corresponding information criteria are also shown, where U and S refer to unstable and stable models, respectively

Table 3. Information criteria and fit of candidate stable (S) and unstable (U) discrete-time output error models.

Criteria \ Order		8	10	12	45	50	60
AIC (10^4)	U	1.47	1.53	1.42	1.06	1.12	**0.99**
	S	**1.47**	1.55	1.50	1.59	1.52	**1.47**
BIC (10^4)	U	1.48	1.54	1.43	1.11	1.16	**1.04**
	S	**1.48**	1.56	1.52	1.64	1.57	1.53
fit (%)	U	24	17	31	63	59	**67**
	S	24	13	20	9	19	**25**

We now compare the obtained unstable and stable models, with order $F(q) = 60$ and $B(q) = 30$ chosen based on AIC and BIC. Fig. 8 shows the frequency response of the estimated models. If stability is not introduced as a constraint, the obtained model is unstable. It can be seen that the frequency response of this estimated model is similar to the frequency response of the subglottal system obtained in Zañartu et al. (2013).

Fig. 9 shows the comparison between the frequency response of the available frequency domain data for identification (3), and the frequency response of the model obtained when forcing stability. It can be noticed that the model fit is reduced due to the stability constraint.

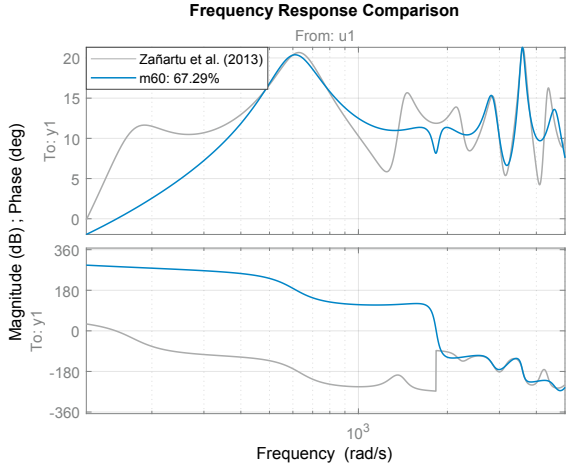

Fig. 8. Frequency response of 60^{th} order unstable estimated output−error model

The results of parameter estimation to fit the state space model (6) are shown in Table 4, which shows the model fit and the associated AIC and BIC.

Fig. 10 shows the frequency response of the estimated unstable state space model of order 50. It can be seen that the model order corresponding to the best information criteria also provides the best model fit.

Fig. 11 shows the frequency response of the estimated stable state space model of order 50. It can be noticed that, even though stability has been enforced in the parameterized model, still a good fit is obtained.

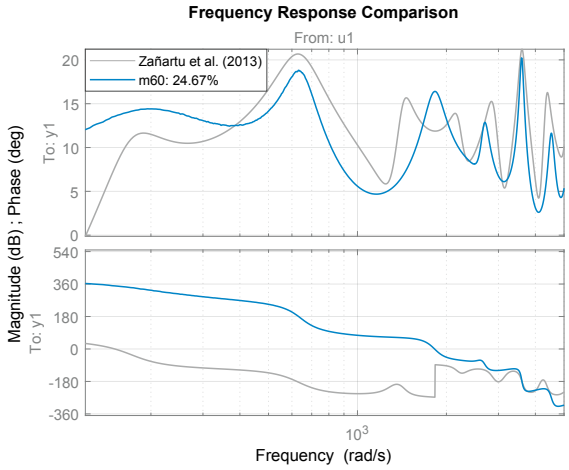

Fig. 9. Frequency response of 60^{th} order stable estimated output−error model

Table 4. Information criteria and fit of candidate stable (S) and unstable (U) space state models.

Criteria \ Order		8	10	12	45	50	60
AIC (10^4)	U	1.41	1.38	1.62	0.35	**-0.02**	1.44
	S	1.41	1.374	1.39	1.13	**0.98**	1.15
BIC (10^4)	U	1.43	1.40	1.63	0.40	**0.04**	1.51
	S	1.43	1.39	1.40	1.18	**1.04**	1.22
fit (%)	U	31	34	4	88	**93**	31
	S	31	36	35	59	**68**	57

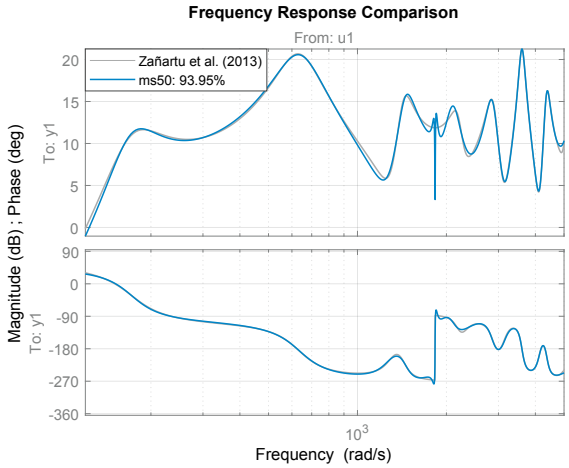

Fig. 10. Frequency response of 50^{th} order unstable estimated space state model

5. CONCLUSION

In this work the parameter estimation of different linear time invariant model structures was performed using ML and PEM algorithms. The identification in the time and frequency domains was carried out, and the goodness of fit and over-fitting were taken into account for the selection of the order, using the Akaike and Bayes information criteria. First, a Box-Jenkins model was obtained in the time domain and the results showed a low fit. The identification was also performed in the frequency domain for two model structures: output-error and space state. In the frequency

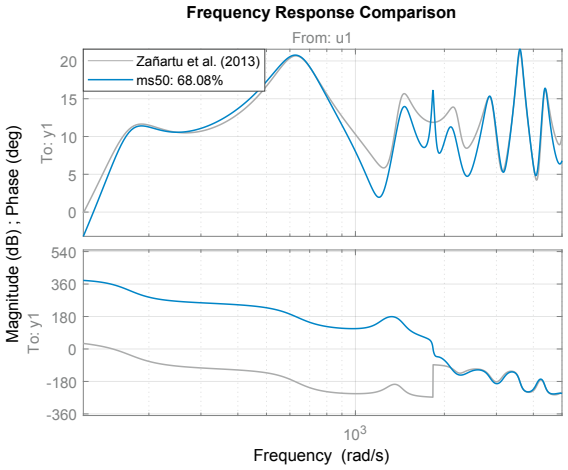

Fig. 11. Frequency response of 50^{th} order stable estimated space state model

domain, the obtained results significantly improved in terms of fitting. However, unstable models were obtained. To avoid this, stability of the model was introduced as a constraint in the estimation. The subglottal system frequency domain identification exhibits better results than time domain identification, especially, when using a state space model structure. The current study shows that the estimation of an LTI model for the subglottal system is a challenging task and model fit accuracy was for the best scenario, only up to 67%.

REFERENCES

Agüero, J.C., Tang, W., Yuz, J.I., Delgado, R., and Goodwin, G.C. (2012). Dual time–frequency domain system identification. *Automatica*, 48(12), 3031 – 3041.

Agüero, J.C., Yuz, J.I., Goodwin, G.C., and Delgado, R.A. (2010). On the equivalence of time and frequency domain maximum likelihood estimation. *Automatica*, 46(2), 260–270.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Alku, P. (2011). 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), 623–650.

Bhattacharyya, N. (2014). The prevalence of voice problems among adults in the United States. *The Laryngoscope*, 124(10), 2359–2362.

Cheyne, H.A. (2006). Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck. In *2006 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*, 118–121.

Cortés, J.P., Espinoza, V.M., Ghassemi, M., Mehta, D.D., Van Stan, J.H., Hillman, R.E., Guttag, J.V., and Zañartu, M. (2018). Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration. *PloS one*, 13(12), e0209017.

Deschrijver, D. and Dhaene, T. (2006). Parametric identification of frequency domain systems using orthonormal rational bases. *IFAC Proceedings Volumes*, 39(1), 837–842.

Drugman, T., Alku, P., Alwan, A., and Yegnanarayana, B. (2014). Glottal source processing: From analysis to applications. *Computer Speech & Language*, 28(5), 1117–1138.

Espinoza, V.M., Zañartu, M., Van Stan, J.H., Mehta, D.D., and Hillman, R.E. (2017). Glottal aerodynamic measures in women with phonotraumatic and non-phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*, 60(8), 2159–2169.

Espinoza Catalán, V. and Zañartu, M. (2014). Estudio dinámico de parámetros de filtrado inverso para el seguimiento ambulatorio de la función vocal. In *IX Congreso Iberoamericano de Acústica*.

Fragnière, B. and Wartmann, J. (2015). Local polynomial method frequency-response calculation for rotorcraft applications. In *AHS 71st Annual Forum*.

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

Levy, E.C. (1959). Complex-curve fitting. *IRE Transactions on Automatic Control*, AC-4(1), 37–43.

Liavas, A.P. and Regalia, P.A. (2001). On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing*, 49(8), 1689–1695.

Mehta, D.D., Van Stan, J.H., Zañartu, M., Ghassemi, M., Guttag, J.V., Espinoza, V.M., Cortés, J.P., Cheyne, H.A., and Hillman, R.E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in bioengineering and biotechnology*, 3, 155.

Mehta, D.D., Zañartu, M., Feng, S.W., Cheyne II, H.A., and Hillman, R.E. (2012). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Biomedical Engineering*, 59(11), 3090–3096.

Perkell, J.S., Hillman, R.E., and Holmberg, E.B. (1994). Group differences in measures of voice production and revised values of maximum airflow declination rate. *The Journal of the Acoustical Society of America*, 96(2), 695–698.

Åström, K. (1980). Maximum likelihood and prediction error methods. *Automatica*, 16(5), 551 – 574.

Söderström, T. (2002). *Discrete-time stochastic systems: estimation and control*. Springer Science & Business Media.

Stoica, P. and Selen, Y. (2004). Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4), 36–47.

Story, B.H. and Titze, I.R. (1995). Voice simulation with a body-cover model of the vocal folds. *The Journal of the Acoustical Society of America*, 97(2), 1249–1260.

Verdolini, K., Rosen, C., Branski, R., et al. (2006). *Classification manual for voice disorders-I*.

Zañartu, M., Ho, J.C., Mehta, D.D., Hillman, R.E., and Wodicka, G.R. (2013). Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(9), 1929–1939.