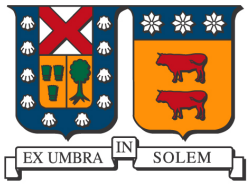
Parametric Identification Methods for the Subglottal System



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Javier Gonzalo González Fontanet

Supervisor: Dr. Juan I. Yuz Eissmann

Department of Electronic Engineering
Universidad Técnica Federico Santa María

This thesis is submitted for the degree of
Ph.D. in Electronic Engineering

August 2024

Thesis Title:

Parametric Identification Methods for the Subglottal System

Author:

Javier Gonzalo González Fontanet

Thesis Work, submitted in partial fulfillment of the requirements for the degree of Ph.D. in Electronic Engineering at Universidad Técnica Federico Santa María.

Supervisor

Dr. Juan I. Yuz Eissmann _____

Internal Reviewer and President of the Committee

Dr. Juan C. Agüero _____

Internal Reviewer

Dr. Matías Zañartu _____

External Reviewer

Dr. Jorge F. Silva _____

To my wife Isdanny Morales, my son Javier Abel and my nephew Aarón, my parents Miriam and Gonzalo, my siblings Yisel and Gonzalo Abel and my grandparents Nidia and Idalberto.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Javier Gonzalo González Fontanet
August 2024

Acknowledgements

This work has received financial support from multiple sources, including ANID, through grants Advanced Center for Electrical and Electronic Engineering FB0008, ECOS 210008, FONDECYT 1230623, and Doctorado Nacional 21202402 scholarship. Additionally, funding was provided by Universidad Técnica Federico Santa María through PIIC grant 015/2021. We also acknowledge support from the National Institute of Health and the National Institute on Deafness and Other Communication Disorders through grant NIH P50DC015446. It is important to note that the views expressed in this content are solely the responsibility of the authors and do not necessarily reflect the official perspectives of the NIH.

I want to thank my colleagues and friends from the Fourier Lab: Claudia Sanches, María Coronel, Rafael Orellana, Ángel Cedeño, and Marco Gordón. They welcomed me at USM, and I have received their unconditional support all these years.

I also want to thank the Cuban team I met in the Department of Electronic Engineering, who became family in Chile: Patricia Morales, Patricia Franco, Reinier López, and Jorge Bermúdez. Other friends who cannot be left out of these acknowledgments are Eusebio Hernández, Yeiniel Alfonso, Marileyxis López, Mariana Rivero, Marais del Río, Alexander Blanco, Orlando Vargas, and Yanelis Ruíz, among many others.

I dedicate this work to my former colleagues from the Department of Automatic Engineering at the Faculty of Electrical Engineering of the Universidad of Oriente, Cuba. Special thanks to Ania Lussón, Irina Bausa, Daily Milanés, Ernesto Estremera, Ksenia Arias, Michel Sanz, Mercedes Ramírez, and José Pullés.

I dedicate this work to my entire family, especially: my parents, my siblings, my grandparents, my nephew Aaronci, and my uncle, who I didn't get the chance to say goodbye to and now lives in my heart.

Perhaps the person who missed me the most during my absence, and whom I missed dearly, was my beloved son. Javi, all my efforts past and future are dedicated to you.

Finally, Chile allowed me to meet my wife, Isdanny Morales. Thanks to you I have achieved so much. This thesis is dedicated to you.

Special thanks to Professor Hugues Garnier for his advice and supervision. Furthermore, I would like to extend my gratitude to my supervisor, Juan I. Yuz Eissmann, for his unwavering commitment to my project and his genuine care for my well-being throughout this remarkable journey.

Abstract

This thesis addresses the problem of mathematical modeling of the subglottal system, motivated by the high rate of affection of voice pathologies in the adult population, supported by the elevated prevalence of documented affections in the literature. Mathematical models serve as cornerstone instruments for advancing medical science and supporting innovative clinical interventions. Utilizing mathematical modeling presents an alternative non-invasive approach that significantly influences the diagnosis and treatment of voice disorders.

This work initially presents the results obtained from the bond-graph modeling of the body cover model (BCM). Once the bond-graph modeling is obtained, it is appropriate to derive the Port-Hamiltonian representation of the BCM, and then proceed with the estimation of the model parameters and states. In this section, for the state estimation is used and extended Kalman filter, and for the parameters estimation maximum likelihood (ML).

Later is presented 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 utilizes glottal volume velocity as the input parameter and yields neck skin acceleration as the corresponding output.

In contrast to preceding efforts that employed frequency-domain methods, the present thesis leverages system identification techniques to derive a parsimonious continuous-time model of the subglottal tract using time-domain data samples. To perform estimation from these time-domain data samples, our emphasis lies on the application of Simplified Refined Instrumental Variables for the Continuous-Time Model (SRIVC). Obtaining results that indicate the SRIVC algorithm effectively estimates a low-order model (with an order model reduction over 70% compared with previous efforts), achieving a high level of fit.

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.

Resumen

Esta tesis aborda el problema del modelado matemático del sistema subglótico, motivada por la alta tasa de afectación de las patologías vocales en la población adulta. Lo anterior se respaldada por la elevada prevalencia de afectaciones documentadas en la literatura. Los modelos matemáticos sirven como instrumentos fundamentales para el avance de la ciencia médica y el apoyo a intervenciones clínicas innovadoras. La utilización del modelado matemático presenta un enfoque alternativo no invasivo que influye significativamente en el diagnóstico y tratamiento de los trastornos de la voz.

Este trabajo presenta inicialmente los resultados obtenidos del modelado usando bond graph del Body Cover Model (BCM). Una vez obtenido el modelado dicho modelo, es apropiado derivar la representación Port-Hamiltoniana del BCM, y luego proceder con la estimación de los parámetros y estados del modelo. En esta sección, para la estimación de estados se utiliza un filtro de Kalman extendido, y para la estimación de parámetros máxima verosimilitud.

Posteriormente, se presenta el modelado del tracto subglótico y las propiedades de la piel del cuello para su uso en la evaluación ambulatoria de la función vocal, permitiendo la monitorización no invasiva del flujo de aire glotal a través de un acelerómetro en la superficie del cuello. Para que la técnica sea efectiva, se requiere el desarrollo de un modelo preciso del tracto subglótico. Dicho modelo utiliza la velocidad de volumen glotal como parámetro de entrada y produce la aceleración de la piel del cuello como salida correspondiente.

En contraste con esfuerzos previos que emplearon métodos en el dominio de la frecuencia, la presente tesis aprovecha técnicas de identificación de sistemas para derivar un modelo continuo en tiempo continuo del tracto subglótico utilizando muestras de datos en el dominio del tiempo. Para realizar la estimación a partir de estas muestras de datos en el dominio del tiempo, nuestro énfasis recae en la aplicación de Variables Instrumentales Refinadas Simplificadas para el Modelo en Tiempo Continuo. Se obtienen resultados que indican que el algoritmo SRIVC estima eficazmente un modelo

de bajo orden (con una reducción del orden del modelo superior al 70% en comparación con esfuerzos anteriores), logrando un alto nivel de ajuste.

Adicionalmente, se realiza un examen del orden de los modelos a través de la aplicación de varios criterios de información. Una vez que un modelo de bajo orden se ajusta exitosamente, se utiliza un filtro inverso basado en un suavizador de Kalman para la estimación de la velocidad de volumen glotal y métricas aerodinámicas relacionadas, constituyendo la ejecución más eficiente de estas estimaciones hasta la fecha.

Contents

List of Figures **xvii**

List of Tables **xix**

List of abbreviations **xxi**

1 Introduction **1**

1.1 Introduction 1

1.1.1 What does system identification have to offer? 5

1.2 Related work 6

1.2.1 Continuous time model parameters estimation method 7

1.3 Thesis outline and contribution 8

1.3.1 Thesis contribution 10

1.4 Associated publications 10

2 Mathematical representation of the subglottal system **13**

2.1 Body Cover Model representation 13

2.1.1 Body-cover model system description 14

2.1.2 Bond-graph representation of the body-cover model 15

2.1.3 Port-Hamiltonian representation. 17

2.2 Impedance-Based Model representation of the subglottal system 23

2.3 Continuous Linear Time Invariant model of the subglottal system . . . 27

2.4 Chapter summary 29

3 Model identification methods 31

3.1 Maximum Likelihood 31

3.1.1 Maximum likelihood for linear algebraic systems 34

3.1.2 General properties of the ML estimator 36

3.2 Simple Refined Instrumental Variable for Continuous time OE model . 37

3.2.1 Continuous-time modeling using frequency-domain data 39

3.3 Model Order selection 42

3.3.1 Akaike information criterion 42

3.3.2 Bayesian information criterion 47

3.3.3 Young information criterion 49

3.3.4 The coherence method 50

3.4 Chapter summary 50

4 Continuous-Time Model Identification of the Subglottal System 53

4.1 Body Cover Model parameter estimation 53

4.1.1 Model simulation 54

4.1.2 State Estimation 56

4.1.3 Parameter Estimation 57

4.1.4 Numerical results 58

4.2 Impedance-base model parameter estimation 60

4.2.1 Data pre-processing 61

4.2.2 Continuous-time model from frequency-domain data 64

4.2.3 Continuous-time model from sampled time-domain data 66

4.3 Linear Time-Invariant model parameter estimation 69

4.3.1 Inverse filtering 73

4.4 Chapter summary 76

5 Conclusions **77**

5.1 Future work 78

Bibliography **81**

List of Figures

2.1 Speech production 14

2.2 Body-cover model for a single vocal fold [1]. 15

2.3 Schematic and bond representation of a single vocal fold, and flow into fluid capacitance and through a narrow pipe [2]. 16

2.4 Schematic and bond representation of a single vocal fold [3]. 17

2.5 Schematic diagram of vocal-fold vibration [4]. 18

2.6 The tract section above the accelerometer is labeled as $sub1$, while the tract section below this point, extending to the alveoli, is labeled as $sub2$. Taken from [5] 24

2.7 Block diagram representation of the phonatory system 24

2.8 Block diagram representation of the phonatory system 28

3.1 Block diagram representation of maximum likelihood algorithm 32

4.1 Lower mass displacement system 60

4.2 Upper mass displacement system 60

4.3 Body mass displacement system 61

4.4 Input force 61

4.5 Lower mass collision displacement 61

4.6 Upper mass collision displacement 61

4.7 Lower mass collision displacement 62

4.8 Upper mass collision displacement 62

4.9 Body mass collision displacement 62
4.10 Input-output resampled dataset power spectrum 63
4.11 Glottal volume velocity (GVV) original and resampled signal 63
4.12 Acceleration (ACC) signal original and resampled signal 63
4.13 Coherence between ACC and GVV 64
4.14 MSE between estimated $H_{sub1}(p)$ model and the data used for estimation 65
4.15 Coherence between simulated data from estimated $H_{sub1}(p)$ model and the data used for estimation 65
4.16 $H_{sub1}(p)$ model stability up to order 14 65
4.17 MSE between estimated $Z_{sub2}(p)$ model and the data used for estimation 66
4.18 Coherence between simulated data from estimated $Z_{sub2}(p)$ model and the data used for estimation 66
4.19 $Z_{sub2}(p)$ model stability up to order 14 66
4.20 Comparison of $G(p) = F(p)T(p)$ model against validation data 67
4.21 Frequency response comparison 68
4.22 Frequency response comparison 70
4.23 Heatmap of AIC coefficients 71
4.24 Heatmap of BIC coefficients 71
4.25 Heatmap of YIC coefficients 71
4.26 Root mean squared error (RMSE) of the candidates models 71
4.27 Heatmap of R_T^2 coefficients 71
4.28 Poles and zeros map of the selected model estimated using SRIVC . . . 72
4.29 Bode diagram of the selected model estimated using SRIVC 72
4.30 Comparison of the estimated model with SRIVC and resampled data simulation against validation data 73
4.31 GVV estimation with Kalman Smoother and SRIVC estimated model . 75

List of Tables

2.1 Main variables and relationships for the mechanical model 18

2.2 Values of the mechanical impedance model parameters of the neck skin surface 26

4.1 True ($\bar{\theta}$) and Estimated Parameters using EKF ($\hat{\theta}_{EKF}$) and Maximum-Likelihood ($\hat{\theta}_{ML}$) estimation 59

4.2 Mean-square error of simulated state $x(k\Delta + \Delta)$ and estimated state $\hat{x}(k\Delta + \Delta)$ 60

4.3 Estimates of the parameters Q_i for five vocalizations of three different subjects 69

4.4 Aerodynamic metrics computed from GVV 75

4.5 Aerodynamic Features computed from GVV_OVV for three subjects under study 76

List of abbreviations

Acronyms / Abbreviations

ACC Acceleration on neck skin

AIC Akaike Information Criterion

BCM Body cover model

BIC Bayes Information Criterion

EKF Extended Kalman Filter

GVV Glottal volume velocity

i.i.d. Independently and identically distributed

IBIF Impedance-based inverse filtering

KL Kullback-Leibler discrepancy function

LTI Linear time invariant

ML Maximum likelihood

OVV Oral volume velocity or oral airflow

pdf Probability density function

PEM Prediction Error Method

SRIVC Simple Refined Instrumental Variables for the Continuous-Time Model

VH Vocal hyperfunction

YIC Young Information Criterion

Chapter 1

Introduction

1.1 Introduction

The study of voice pathologies is of great interest in the scientific community. From a clinical point of view, a model to describe the behavior of the vocal folds is needed in order to simulate different laryngeal adjustments or pathological variations [6] by studying specific aspects of the mechanics of the vibration. Models of the vocal fold have also been used for speech synthesis systems [7].

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 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 [8, 9], 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 [10–13], trigger bifurcations and voice breaks [14, 15], and yield natural quantal differences for the vowel space [9, 16, 17].

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 [18], and affects

approximately 6.6% of the adult population with a lifetime prevalence of 30% [19, 20]. These statistics are even higher for people who rely on their voices for work, particularly in the female education sector. According to the Chilean Institute of Public Health, speech-related conditions are among the first occupational problems in the country, especially in the female educational sector [21]. This highlights the critical need for new non-invasive approaches to assess the overall health of the vocal apparatus.

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) [22, 23] from recordings of oral airflow (or oral volume velocity, OVV) with a pneumotachograph mask [24]. Notably, the GVV signal can also be obtained using subglottal inverse filtering methods using models of the subglottal system [25, 5] 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 [23, 26, 27] through ambulatory voice monitoring devices, e.g [28].

Mathematical representations of the subglottal system [29–31] have played an important role in the development of the signal processing components needed for subglottal inverse filtering. These models have shown advances in the simulation of different laryngeal adjustments and pathological variations. In other cases, numerical models of the vocal folds and vocal tract have been used to design voice synthesizers and to study vibration modes of the vocal folds [7].

Measuring the vibration of the neck skin due to the modulation of the glottal airflow by the vocal folds is another non-invasive alternative approach to studying vocal hyperfunction. This technique has been extensively studied in the literature and provides valuable insights into the features associated with vocal hyperfunction [25, 32, 5, 33, 34]. This method involves deriving GVV from the oral airflow (or oral volume velocity, OVV) with a circumferentially vented pneumotachograph mask, as described in [22, 23]. Then, a transmission line model that represents the subglottal system was proposed by [5] to physiologically relate the ACC and GVV signals, in an approach referred to as impedance-based inverse filtering (IBIF).

As a result, aerodynamic features can be obtained from GVV such as peak-to-peak AC flow, open quotient, and maximum flow declination rate [35–37]. 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 [23]. 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 [5, 26, 27]. 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.

To remove the influence of the vocal tract, the OVV is filtered, and an impedance-based model that represents the subglottal system connected to the neck-surface model is fitted using the neck-skin accelerometer (ACC) signal. Once the model is fitted, estimations of the GVV waveform can be obtained by applying the inverse of the impedance-based model (i.e., filter) to the neck-skin acceleration signal. As a result, aerodynamic features can be obtained from GVV such as peak-to-peak AC flow, open quotient, and maximum flow declination rate [35–37]. 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 [23]. 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 [5, 38, 34]. 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 [30, 39, 40]. 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 [41–43]. In this work, 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 [27, 5]. 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 [5]. Furthermore in [8], the cavities responsible for transforming the glottal source were presented as linear filters.

The **main goal** of this work is to propose a methodology to estimate the parameters of the body cover model, the subglottal IB model, and black-box model, by using statistical parametric estimation techniques based on system identification. The **specific objectives** of this thesis are: to implement identification methods to obtain the parameters of the body cover model; to implement identification methods to obtain the parameters of the IB model; to propose and identify the parameters of Linear Time Invariant models; and to study the trade-off between model order and complexity using information criteria. The assumed **hypothesis** is that the estimation of the subglottal model parameters by using statistical methods will provide quantitative information beyond point estimates based on the optimization of the ad-hoc deterministic fit function. We believe that these results will contribute to improving the ambulatory studies of voice pathologies.

In this thesis we present the use of linear time-invariant models, to describe the subglottal system in continuous-time. The fit of the obtained model will be performed by using system identification techniques, specifically simplified refined instrumental variables for the continuous-time model. SRIVC is widely recognized as one of the most successful direct methods for continuous-time model identification from sampled data [44]. This method has been extensively applied in different fields, including chemical processes, electronic circuits, and biological systems modeling [45, 46].

The purpose of this work is not to develop new models for the subglottal system because the existing ones are not sufficiently good. However, it is important to note that the impedance-based model (IB), when accounting for all the branching of the lungs, yields a model of order 350. Working with such high-order models presents several challenges that must be considered, such as increased computational complexity, numerical instability, interpretability issues, overfitting, and implementation problems [42].

The application of this thesis is focused on ambulatory studies, making it advisable to avoid working with overly simplistic models that merely fit the system. Moreover, previous studies have focused on point estimation of parameters. However, it is of great interest to not only obtain point estimates but also statistical distributions of these parameters. In this regard, system identification techniques offer an advantage,

allowing for better characterization of the system's behavior and providing a more robust framework for parameter estimation.

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

1.1.1 What does system identification have to offer?

System identification [47] is a diverse field of science that deals with the challenge of constructing mathematical models for dynamic systems based on measured data from them[48, 41]. According to Ljung (2010), it serves as the interface between real-world applications and mathematical abstractions [49]. Among the sciences that are interconnected to build the theoretical background of system identification are systems theory, signal processing, optimization, and statistics. Some fundamental tasks in system identification include experimental design, model structure selection, parameter estimation, and model validation [50].

System identification has a wide range of applications, extending from process industries and mechatronics to the aerospace industry, with notable growth in the field of bioengineering in recent years. In this diverse context, one of the primary objectives of employing system identification is to obtain a more accurate and realistic representation of the physical system to be used for fitting controllers based on internal models [51], enabling better adaptation and optimization of systems across various industries and disciplines.

Some relevant applications of system identification in the field of bioengineering are related to the identification and control of organs, such as the case of an artificial pancreas [52–54]; the heart or cardiovascular system [55–57]; and human-like walking, postural dynamics, and motion control [58–61].

The need for well fit mathematical models, rather than relying solely on practical experiments is an important application of system identification, justified by several factors such as: expensive experiments can be financially prohibitive, especially in advanced research areas or situations requiring substantial resources. Additionally,

when experiments pose a safety risk to researchers or the environment, simulation based on mathematical models offers a safe and ethical alternative. Moreover, in instances where system time constants are extremely large, practical experiments may become impractical due to their slow pace, making simulation an invaluable tool for efficiently studying system behavior [62].

1.2 Related work

Several methods have been employed for studying speech production, including inverse filtering, digital high-speed imaging of the vocal folds [63–65], and electroglottography [66–70]. In studies relying on inverse filtering, it is common to utilize a circumferentially vented pneumotachograph mask, commonly known as the Rothenberg mask [24], to capture oral airflow. This approach highlights the distinction between normal vocal function and vocal hyperfunction (VH). However, data of this nature can only be obtained within laboratory settings. A significant breakthrough in VH research emerged from the work of [71], proposing the use of neck skin accelerometers to measure the neck skin acceleration. The skin acceleration is measured by attaching an accelerometer to the neck surface between the thyroid prominence and the suprasternal notch [25]. 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 [72, 28].

Measuring the acceleration on the neck skin due to the airflow through the glottis has emerged as a popular method for assessing vocal function [38, 34]. By deriving estimates of the mechano-acoustic transfer function $T_{skin}(\omega)$ from OVV and ACC, these values can be applied in the impedance-based inverse filtering (IBIF) strategies. This enables an accurate estimation of the aerodynamic source of voice sounds at the glottis [5]. Specialists can then use this estimated airflow to compute aerodynamic features. 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) [73].

The solutions based on IBIF requires the optimization of the model, for example using Particle Swarm Optimization (PSO). It has been reported that for multiple iterations of the PSO algorithm, it continuously gives different estimation results for the same parameter under the same conditions [23]; in addition, statistics of the estimated parameters are based on multiple realizations and not on any statistical analysis.

Additional endeavors has yielded state-space representations of an impedance-based model for the subglottal system, derived from the system's frequency response[74], leading to a precise model. Additionally, this study proposes a Bayesian approach utilizing the Kalman smoother to estimate GVV from neck-skin signal acceleration.

Another model studied for vocal assessment is the Body Cover Model (BCM), which is based on the mechanical properties of the vocal folds. In its simplest expression, BCM is represented by a set of lumped masses. For this model, a Port-Hamiltonian Systems (PHS) approach has also been proposed based on the energy function of the system [1, 75].

The mathematical modeling of the vocal folds is supported by extensive previous work on models for human and animal body organs. Among the most studied cases are the pancreas and type 1 diabetes [76–78], the cardiovascular and pulmonary system, and the hemodynamics of the heart [79, 80].

1.2.1 Continuous time model parameters estimation method

When the theory of system identification first appeared, it was mainly aimed at the estimation of continuous-time models [43]. 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 [81].

The parameters of a physical system often have a direct interpretation in continuous time, while the parameters of a discrete-time model may lack physical meaning. This lack of significance can complicate the expert validation of the identified model, as the parameters may not align with the prior knowledge of the expert about the physical system [82]. Some other difficulties that arise in discrete-time identification are related to the sampling time, which may result in the 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 [81, 83].

The estimation of continuous-time models from a sampled dataset has been shown to have advantages over discrete-time methods when using real data [45]. Moreover, the obtained continuous-time models can be discretized for different sampling rates if

needed. A more comprehensive list of advantages and disadvantages of continuous-time model identification can be found in [43, 45, 46].

Simplified refined instrumental variables for continuous-time model is one of the most applied continuous-time model 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 [84, 85].

1.3 Thesis outline and contribution

In this section is provided the outline of the thesis and the contribution of each chapter.

Chapter 2 describes three mathematical models: one for the vocal folds and two for the complete subglottal system. In the first case, the body cover model (BCM) is described, providing an overview of its characteristics. Subsequently, the interactions of the mass-spring-damper system that constitutes the BCM are depicted using bond-graph theory. This approach facilitates the derivation of the Port-Hamiltonian representation of the BCM, which is also addressed in this section.

Next, the state of the art of impedance-based mechano-acoustic model is presented, along with a description of the equations that constitute this model. Starting from a block diagram representation of the subglottal system, the equations to be estimated are introduced. This involves presenting separately both the acoustic model and the mechanical model.

Finally, building upon studies demonstrating the feasibility of attaining a linear representation of the subglottal system, the output error structure is introduced. This structure will be subsequently employed to consider Linear Time Invariant (LTI) models having different orders.

Chapter 3 presents two estimation methods: Maximum Likelihood (ML) and Simple Refined Instrumental Variables for the Continuous-Time Model (SRIVC). It outlines the general properties of ML and its development for general systems, as well as the derivation of SRIVC from sampled data and the system's frequency response. Additionally, information criteria are introduced —Akaike Information Criterion , Bayes Information Criterion, and Young Information Criterion— that are utilized in the subsequent chapters to determine the order of Linear Time Invariant (LTI)

models. The chapter also introduces the coherence method, employed to demonstrate that, within the frequency range of interest, the subglottal system can be effectively represented by a linear model.

Chapter 4 presents the estimation results for the models introduced in Chapter 2. Firstly, simulations of the Body Cover Model are conducted, highlighting the challenges in estimating states and parameters for this model. The estimations from Maximum Likelihood are then compared with those from a Kalman filter.

Subsequently, estimations of the IB model are presented. The acoustic model parameters are estimated from frequency response, while the mechanical model parameters are derived from sampled input-output data. The pre-processing of the input-output dataset is also detailed. This section compares estimation results across multiple subjects and various vocalizations.

Finally, the subglottal system is presented as a Linear Time Invariant model. To select the best model, a set of candidate models is estimated, and the optimal is selected based on information criteria. The chapter also introduces inverse filtering to obtain glottal airflow from the estimated model, employing a Kalman smoother.

Chapter 4 concludes with general findings from this thesis and outlines future research directions.

1.3.1 Thesis contribution

Contributions

(i) The representation of the BCM through bond-graph modeling is proposed, facilitating the derivation of the Port-Hamiltonian model of the vocal folds.

(ii) The use of parametric estimation techniques is proposed for obtaining continuous-time models from sampled data to represent the subglottal system based on LTI models.

(iii) The estimation of 3 out of the 5 $Q_{1,2,\ldots,5}$ parameters of interest in the continuous-time impedance-based model was performed using parametric estimation techniques and sampled data-time.

(iv) LTI models were obtained with a reduction in order of over 70% compared to previous efforts, resulting in lower computational cost and time, which represents a significant advantage for ambulatory assessment.

1.4 Associated publications

This thesis is complemented by a collection of publications co-authored by the candidate, along with other relevant publications produced during the research period.

Journal

J.1 Arturo Morales, Juan I. Yuz, Juan Pablo Cortés, Matías Zañartu, **Javier G. Fontanet**. 2023. Glottal Airflow Estimation using Neck Surface Acceleration and Low-Order Kalman Smoothing. IEEE/ACM Transactions on Audio, Speech, and Language Processing 31, 2055–2066. doi:10.1109/TASLP.2023.3277269.

J.2 **Javier G. Fontanet**, Juan I. Yuz, Hugues Garnier, Arturo Morales, Juan Pablo Cortés, Matías Zañartu. (2024). "Continuous-Time Model Identification of the Subglottal System". Biomedical Signal Processing and Control journal. Volume 95, Part A, No. 106394, doi: https://doi.org/10.1016/j.bspc.2024.106394, ISSN: 1746-8094.

Conference

C.1 **Javier G. Fontanet**, Juan I. Yuz, Javier A. Torres, Marcos A. Gordon, Héctor Ramírez. "Port-Hamiltonian modeling of the vocal folds using bond-graph representation". 2021 IEEE International Conference on Automation/XXIV Congress of the Chilean Association of Automatic Control (ICA-ACCA). Doi: 10.1109/ICAACCA51523.2021.9465330

C.2 **Javier G. Fontanet**, Juan I. Yuz, Matías Zañartu. "Parametric Identification of a Linear Time Invariant Model for a Subglottal System". 19th IFAC Symposium System Identification: learning models for decision and control. 2021. IFAC-PapersOnLine, 54(7), 577–582. doi:10.1016/j.ifacol.2021.08.422.

C.3 **Javier G. Fontanet**, Juan I. Yuz, Hugues Garnier, Víctor Espinoza, Matías Zañartu. (2023). "Subglottal Impedance-Based Model Parameter Estimation via System Identification". Presented in the: 20th IFAC symposium on system identification SYSID 2024.

Article J1 proposes a low-order Kalman smoother to estimate the glottal airflow from the acceleration measured on the skin of the neck. Low-order state-space representations of an impedance-based model of the subglottal system are obtained from the system's frequency response. At the time of publication, the proposed model achieved a 94% order reduction compared to previous works, reducing computational time to 1.5%, which is attractive for ambulatory monitoring applications. Although there was collaboration with the primary author on this work, the results are part of Arturo Morales' master's thesis requirements.

Chapter 2

Mathematical representation of the subglottal system

The purpose of this chapter is to introduce background regarding the mathematical representations of the speech system. In particular we present two physically based models that appear in the literature: the body cover model (BCM) and the impedance based model (IB). Also, an LTI model is presented that is used in the rest chapters for system identification.

2.1 Body Cover Model representation

From a clinical perspective, a model is required to describe the behavior of the vocal folds in order to simulate various laryngeal adjustments or pathological variations, thereby studying specific aspects of the mechanics of vibration.

Human beings, along with other animal species, shape the produced sound to convert it into understandable sound waves. In voice production, the airflow originating from the lungs at a certain pressure causes the vocal folds to vibrate, modulating this flow through the glottis [86]. Subsequently, this airflow, having passed through the pharyngeal filter, undergoes further filtration in the oral cavity to produce vowels and audible consonants. The primary organ involved in vocal production is the larynx (see Figure 2.1). The vocal mechanism is also integral to the respiratory system, coordinated with several cortical areas.

Figure 2.1 Speech production

2.1.1 Body-cover model system description

The morphological structure of the vocal cords, biomechanically divided into the body and the cover layer, was proposed in 1974 by Hirano in [6]. In this theory, the mechanical properties of the thyroarytenoid muscle are described, which, along with vocalis muscle fibers, work as a single vibration unit during phonation, and this is called "body" [6, 87]. The "cover' is formed by the vocal fold epithelium and the superficial layer of the lamina propria, which is responsible for providing support and nutrition to the epithelium.

The development of BCM traces its origins to works such as [88], where the vocal fold is conceptualized as a mass-spring oscillator driven by the airflow from the lungs. However, this model had a drawback as it could not generate the necessary vertical phase difference for flow-induced oscillation. This limitation was addressed in [86], which successfully introduced the required degrees of freedom by incorporating a second mass-spring system. Other significant contributions to the development of this theory are found in the works of Titze (1973 - 1974) [89, 90], wherein the vocal folds are modeled using 16 masses to gain a better understanding of how the voice source operates [7].

In the work of Story and Titze in 1995, other influential studies are presented before proposing a 3-mass model. In this model, the third mass is added to simulate the effect of the body component.

The mechanical properties of the vocal folds, defined by the elastic constant, position, shape, and mass, are determined by the intrinsic and extrinsic laryngeal muscles [6]. Based on these mechanical properties and previous models, in [7] a model of three lumped masses for the vocal folds is proposed as an approximation of the body-cover model, as shown in Figure 2.2.

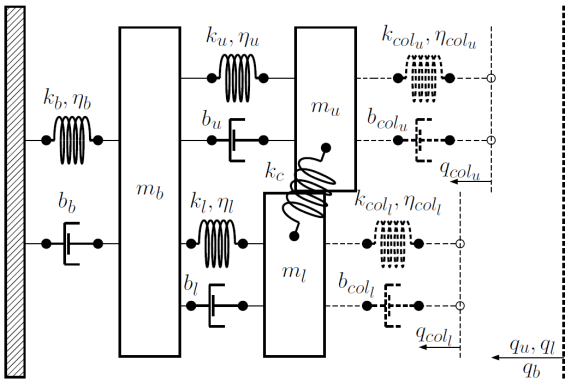


Figure 2.2 Body-cover model for a single vocal fold [1].

In Figure 2.2, an elongated mass (m_b) is used to simulate the body layer, attached through a spring and a damper to the thyroid cartilage rigid boundary. To this mass m_b, two other masses (m_u, m_l) are joined through springs and dampers, coupled in turn by a spring, which represents the shear forces on the cover [7, 1]. A common assumption is that the vocal folds are symmetrical, that is, the movement that occurs in one vocal fold is replicated exactly in the other. As a consequence, the vocal folds only collide in the glottal mid-line.

The spring k_b represents the stiffness of the body that varies with the contraction and relaxation of the thyroarytenoid and cricothyroid muscles. The springs that join the mass m_b with m_u and m_l represent the stiffness of the cover tissue and between the body and the cover [7].

The vocal folds are essential for speech. They vibrate to modulate the airflow through the glottis. This oscillation chops up the air when the vocal folds collide. Collisions can be modeled by switches that activate or deactivate certain mechanical elements, represented by springs and dampers (k_{col_u}, η_{col_u}, b_{col_u}, k_{col_l}, η_{col_l}, b_{col_l}), due to their similarity to the plastic and elastic nature of the collision [1]. In Figure 2.2, the elements representing the collisions are represented by broken lines: they are considered only when the vocal folds collide, and when the folds are open they are not considered in the model.

2.1.2 Bond-graph representation of the body-cover model

A bond-graph is an intuitive approach for graphic representation of a system, independent of the physical domain under study. The latter is convenient for vocal folds

modelling since a number of elements, both from the mechanical and pneumatic domain, can be used in this way. In [2], bond-graph model of anura vocal system is presented, serving as a starting point for the development of the current paper.

Figure 2.3 shows two commonly used sub-models of grouped elements for the representation of vocal production. The first one (A) can be seen as the simplest diagram of a vocal fold, formed by a mass, a spring and a damper, and its representation in a bond-graph. From this idea, the bond-graph of the complete vocal folds system proposed by Encina et al (2015) in [1] will be obtained. The **Se** element in the bond graph represents the addition of force to the system from an external source.

In Figure 2.3-B, the airflow from the lungs and its circulation through the vocal tract is represented as a capacity and an airflow coming from it, compressed to circulate through a pipe, expressed in the pneumatic domain. The **Sf** element in the bond-graph represents the addition of force to the system from an external source.

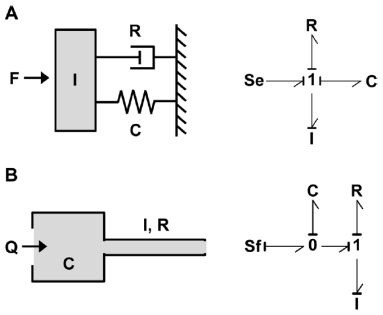


Figure 2.3 Schematic and bond representation of a single vocal fold, and flow into fluid capacitance and through a narrow pipe [2].

From the model of a vocal fold, the elements are assembled in a bond-graph model following the energy flow through the system. After expressing the model, represented by physical elements such as dampers and springs, in a bond-graph model, a number of methods are available to derive the graphical model, and convert the bond-graph into a set of equations that can then used to simulate the system [91].

From the system diagram in Figure 2.2, five dissipative elements are identified, corresponding to the mechanical damping elements. On the other hand, there are three inertial elements, corresponding to the body layer mass and the cover layer masses. In addition, there are six energy storing elements due to the springs. The resulting translation of the model into graphic language, is shown in Figure 2.4.

Taking into account a more general scheme, as shown in [2], the energy source of the system is the airflow produced by compression of the lungs. Part of this flow

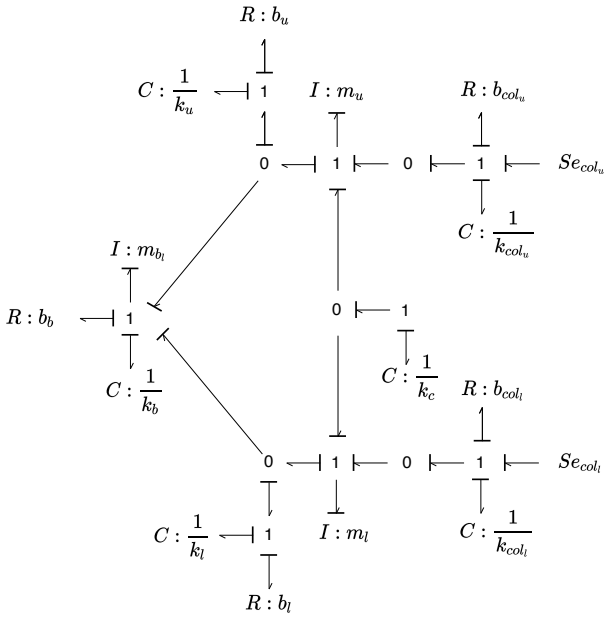


Figure 2.4 Schematic and bond representation of a single vocal fold [3].

remains in the lungs, and the rest is stored in the posterior chamber of the larynx determining a pressure within the chamber, which is transduced into a force acting on each vocal fold. Some of the forces applied to these folds are stored in the momenta of the mass of the skin tissue. In fact, the speed of vibration of the fold depends on its momenta.

2.1.3 Port-Hamiltonian representation.

As shown in Figure 2.2, the body-cover model proposed by Story and Titze in [7] has eight energy storage elements: three masses and five springs. The basic relationships between velocities and forces are shown in Table 2.1.

We refer to $q_u(t)$, $q_l(t)$ and $q_b(t)$ as the displacement of upper, lower and body masses respectively, and $q_{col_u}(t)$ and $q_{col_l}(t)$ are the collision displacements, and $p_u(t)$, $p_l(t)$ and $p_b(t)$ are the momenta of the corresponding masses.

The different approaches to obtain a linear or a non-linear model depend on the choice of the Hamiltonian, (since a non-linear term is added to the energy stored in the springs), both models the interconnection and dissipation matrices will be the same. Taking into account the above, only the non-linear case is presented.

Table 2.1 Main variables and relationships for the mechanical model

	Power conjugate variables	
Elements	**Flow**	**Effort**
Spring	$v_k = \dfrac{dq}{dt} = \dot{q}$	$F_{k,\eta} = F(q)$
Mass	$v_m = \dfrac{p_m}{m}$	$F_m = \dfrac{dp_m}{dt} = \dot{p}_m$
Damper	$Fb = b\ v_b$	

The upper and lower collisions are modeled as switches s_u, s_l, where a 0 value indicates that the vocal fold are separated and 1 indicates the collision mode. For instance, Figure 2.5 shows a schematic diagram of the vocal-fold vibration.

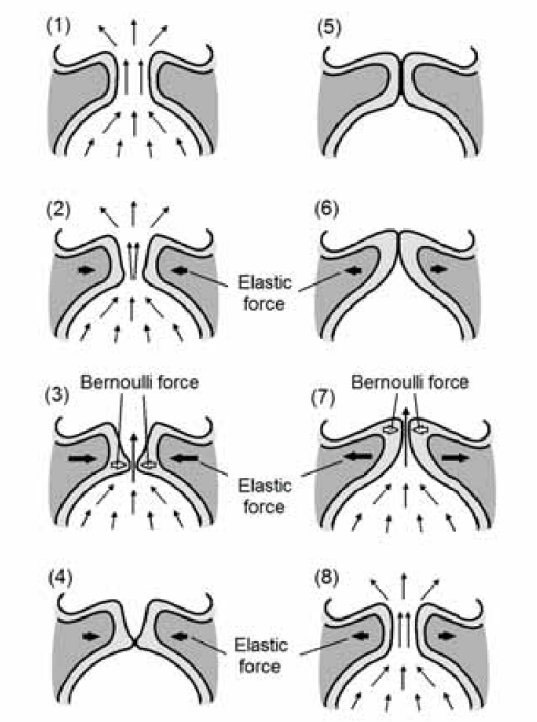


Figure 2.5 Schematic diagram of vocal-fold vibration [4].

- Displacement of the masses and springs:

$$\dot{q}_b = v_b = \frac{p_b}{m_b} \tag{2.1}$$

$$\dot{q}_l = v_l = \frac{p_l}{m_l} \tag{2.2}$$

$$\dot{q}_u = v_u = \frac{p_u}{m_u} \tag{2.3}$$

$$\dot{q}_{col_u} = s_u \; v_{col_u} = s_u \left(\frac{p_u}{m_u}\right) \tag{2.4}$$

$$\dot{q}_{col_l} = s_l \; v_{col_l} = s_l \left(\frac{p_l}{m_l}\right) \tag{2.5}$$

- Momentum of the upper mass ($\dot{p}_u$):

$$\begin{aligned}
\dot{p}_u &= F_{m_u} \\
&= F_{in_u} - [F_{k_u} + F_{b_u} + F_{c-u} + s_u(F_{col_{ku}} + F_{col_{bu}})] \\
&= F_{in_u} - [k_u((q_u - q_{u_0}) - (q_b - q_{b_0})) + b_u(v_u - v_b) \\
&+k_c((q_u - q_{u_0}) - (q_l - q_{l_0})) + s_u(k_{col_u} q_{col_u} + + b_{col_u}(v_u - v_b))] \\
&= F_{in_u} - [k_u((q_u - q_{u_0}) - (q_b - q_{b_0})) + k_c((q_u - q_{u_0}) \\
&-(q_l - q_{l_0})) + s_u(k_{col_u} q_{col_u}) + (b_u + s_u b_{col_u}) \left(\frac{p_u}{m_u}\right) - (b_u + s_u b_{col_u}) \left(\frac{p_b}{m_b}\right)\Big]
\end{aligned} \tag{2.6}$$

and let $\alpha = (b_u + s_u b_{col_u})$, then

$$\begin{aligned}
\dot{p}_u = F_{in_u} - [&k_u((q_u - q_{u_0}) - (q_b - q_{b_0})) + k_c((q_u - q_{u_0}) - \\
&-(q_l - q_{l_0})) + s_u(k_{col_u} q_{col_u}) + \alpha \left(\frac{p_u}{m_u}\right) - \alpha \left(\frac{p_b}{m_b}\right)\Big]
\end{aligned} \tag{2.7}$$

- Momentum of the lower mass ($\dot{p}_l$):

$$\begin{aligned}
\dot{p}_l &= F_{m_l} \\
&= F_{in_l} - [F_{k_l} + F_{b_l} + F_{c-l} + s_l(F_{col_{kl}} + F_{col_{bl}})] \\
&= F_{in_l} - [k_l((q_l - q_{l_0}) - (q_b - q_{b_0})) + b_l(v_l - v_b) + \\
&+ k_c((q_l - q_{l_0}) - (q_u - q_{u_0})) + s_l(k_{col_l} q_{col_l} + b_{col_l}(v_l - v_b))] \\
&= F_{in_l} - [k_l((q_l - q_{l_0}) - (q_b - q_{b_0})) + k_c((q_l - q_{l_0}) - \\
&- (q_u - q_{u_0})) + s_l(k_{col_l} q_{col_l}) + (b_l + s_l b_{col_l}) \left(\frac{p_l}{m_l}\right) - (b_l + s_l b_{col_l}) \left(\frac{p_b}{m_b}\right)\Big]
\end{aligned} \tag{2.8}$$

and let $\beta = (b_l + s_l b_{col_l})$, then

$$\begin{aligned}
\dot{p}_l = F_{in_l} - [&k_l((q_l - q_{l_0}) - (q_b - q_{b_0})) + k_c((q_l - q_{l_0}) - \\
&- (q_u - q_{u_0})) + s_l(k_{col_l} q_{col_l}) + \beta \left(\frac{p_l}{m_l}\right) - \beta \left(\frac{p_b}{m_b}\right)\Big]
\end{aligned} \tag{2.9}$$

- Momentum of the body mass ($\dot{p}_b$):

$$\begin{aligned}
\dot{p}_b &= F_{m_b} \\
&= F_{k_u} + F_{b_u} + F_{k_l} + F_{b_l} + s_u F_{col_{bu}} + s_l F_{col_{bl}} - F_{b_b} - F_{k_b} \\
&= k_u((q_u - q_{u0}) - (q_b - q_{b0})) + b_u(v_u - v_b) + k_l((q_l - q_{l_0}) - \\
&- (q_b - q_{b_0})) + b_l(v_l - v_b) + s_u b_{col_u}(v_u - v_b) + \\
&+ s_l b_{col_l}(v_l - v_b) - b_b v_b - k_b(q_b - q_{b_0}) \\
&= (b_u + s_u b_{col_u}) \left(\frac{p_u}{m_u}\right) + (b_l + s_l b_{col_l}) \left(\frac{p_l}{m_l}\right) - \\
&- (b_u + b_l + s_u b_{col_u} + s_l b_{col_l}) \left(\frac{p_b}{m_b}\right) + k_u((q_u - q_{u0}) - (q_b - q_{b0})) \\
&+ k_l((q_l - q_{l_0}) - (q_b - q_{b_0})) - k_b(q_b - q_{b_0})
\end{aligned} \tag{2.10}$$

and let $\gamma = \alpha + \beta + b_b$, then

$$\begin{aligned}
\dot{p}_b = \alpha \left(\frac{p_u}{m_u}\right) + \beta \left(\frac{p_l}{m_l}\right) - \gamma \left(\frac{p_b}{m_b}\right) + k_u((q_u - q_{u_0}) - (q_b - q_{b_0})) + \\
+ k_l((q_l - q_{l_0}) - (q_b - q_{b_0})) - k_b(q_b - q_{b_0})
\end{aligned} \tag{2.11}$$

Equations (2.1)-(2.11) describe the model of a vocal fold and they were obtained from the bond-graph shown in Figure 2.4. Now, we obtain the energy function of the system to express the vocal folds model as a port-Hamiltonian system.

- Energy of the system: The energy of the system is the sum of potential energy in the springs ($U's$) and the kinetic energy in the masses ($K's$), i.e, the Hamiltonian is given by:

$$\begin{aligned} H(x(t)) = U_u(q_u, q_b) + U_l(q_u, q_b) + U_b(q_b) + U_c(q_l, q_u) + U_{col_u}(q_{col_u}) \\ + U_{col_l}(q_{col_l}) + K_u(p_u) + K_l(p_l) + K_b(p_b) \end{aligned} \tag{2.12}$$

- Energy stored in the springs:

$$U_u(q_u, q_b) = \frac{1}{2}\frac{\left[(q_u(t) - q_{u_0}) - (q_b(t) - q_{b_0})\right]^2}{(k_u)^{-1}} + \frac{1}{4}\frac{\left[(q_u(t) - q_{u_0}) - (q_b(t) - q_{b_0})\right]^4}{(k_u\eta_u)^{-1}} \tag{2.13}$$

$$U_l(q_l, q_l) = \frac{1}{2}\frac{\left[(q_l(t) - q_{l_0}) - (q_b(t) - q_{b_0})\right]^2}{(k_l)^{-1}} + \frac{1}{4}\frac{\left[(q_l(t) - q_{l_0}) - (q_b(t) - q_{b_0})\right]^4}{(k_l\eta_l)^{-1}} \tag{2.14}$$

$$U_b(q_b) = \frac{1}{2}\frac{\left[q_b(t) - q_{b_0}\right]^2}{(k_b)^{-1}} + \frac{1}{4}\frac{\left[q_b(t) - q_{b_0}\right]^4}{(k_b\eta_b)^{-1}} \tag{2.15}$$

$$U_{col_u}(q_{col_u}) = \frac{1}{2}\frac{\left[q_{col_u}(t)\right]^2}{(k_{col_u})^{-1}} + \frac{1}{4}\frac{\left[q_{col_u}(t)\right]^4}{(k_{col_u}\eta_{col_u})^{-1}} \tag{2.16}$$

$$U_{col_l}(q_{col_l}) = \frac{1}{2}\frac{\left[q_{col_l}(t)\right]^2}{(k_{col_l})^{-1}} + \frac{1}{4}\frac{\left[q_{col_l}(t)\right]^4}{(k_{col_l}\eta_{col_l})^{-1}} \tag{2.17}$$

  where the η's are the nonlinear coefficients.

- Energy stored in the masses:

$$K_u(p_u) = \frac{1}{2}\frac{(p_u(t))^2}{m_u} \tag{2.18}$$

$$K_l(p_l) = \frac{1}{2}\frac{(p_l(t))^2}{m_l} \tag{2.19}$$

$$K_b(p_b) = \frac{1}{2}\frac{(p_b(t))^2}{m_b} \tag{2.20}$$

- Port-Hamiltonian representation of the system: An n-dimensional input state output port-Hamiltonian system is given by [92]:

$$\begin{aligned}\dot{x}(t) &= [J(x(t)) - R(x(t))]\frac{\partial H(x(t))}{\partial x(t)} + g(x(t))u(t)\\ y &= g^T(x(t))\frac{\partial H(x(t))}{\partial x(t)}\end{aligned} \tag{2.21}$$

where the state $x(t)$, for the proposed model, is defined by:

$$\begin{aligned}x(t) = &[q_u(t), q_{col_u}(t), p_u(t), q_l(t), q_{col_l}(t), p_l(t), q_b(t)\\ &p_b(t)]^T\end{aligned}$$

and the matrices $J(x) = -J^T(x)$ and $R(x) = R^T(x) \geq 0$. These matrices and the input map $g(x)$ are given by:

$$J(x) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & s_u & 0 & 0 & 0 & 0 & 0\\ -1 & -s_u & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & s_l & 0 & 0\\ 0 & 0 & 0 & -1 & -s_l & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \quad R(x) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \alpha & 0 & 0 & 0 & 0 & -\alpha\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \beta & 0 & -\beta\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & -\alpha & 0 & 0 & -\beta & 0 & \gamma \end{bmatrix}$$

$$g(x) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}^T$$

and $\alpha = (b_u + s_u b_{col_u})$, $\beta = (b_l + s_l b_{col_l})$ and $\gamma = \alpha + \beta + b_b$. Note that the output is the velocity of lower mass. Then, the energy gradient is given by:

$$\frac{\partial H(x(t))}{\partial x(t)} = \begin{bmatrix}\frac{\partial H}{\partial q_u}, & \frac{\partial H}{\partial q_{col_u}}, & \frac{\partial H}{\partial p_u}, & \frac{\partial H}{\partial q_l}, & \frac{\partial H}{\partial q_{col_l}}, & \frac{\partial H}{\partial p_l}, & \frac{\partial H}{\partial q_b}, & \frac{\partial H}{\partial p_b}\end{bmatrix}^T \tag{2.22}$$

2.2 Impedance-Based Model representation of the subglottal system

A well-known and extensively studied model for the glottal system is the so-called impedance-based model, obtained from a physiologically-based transmission line model. This model integrates the acoustic model for the lungs combined with a mechanical model based on specific skin characteristics. The IB model holds significant value in the ambulatory study of the subglottal system. Once an IB model is fitted to a patient, it allows for the estimation of GVV and the associated aerodynamic features from ACC through inverse filtering. The estimation of GVV is a key process that allows for evaluating, monitoring, and discerning differences between VH and healthy controls [5, 38, 34, 93]. Nonetheless, a noteworthy hurdle in these investigations lies in the precise estimation of individual-specific parameters associated with the subglottal system and the mechanical properties of the neck skin.

In the literature, a common method to obtain patient-specific parameters of the IB model is based on running a Particle Swarm Optimization algorithm (PSO) [5]. In this process, a normalized weighted absolute error function between oral-based and neck-based GVV is optimized. The obtained optimal parameters are those that best represent the anatomical characteristics of the patients.

Ambulatory assessment of vocal fold health has been performed measuring the neck skin acceleration generated by airflow through the glottis. For this purpose, an accelerometer is placed between the thyroid prominence and the suprasternal notch [25], see Figure 2.6. This non-invasive method is advantageous for evaluating the speech system condition, and it is particularly well-suited for ambulatory research due to its noise immunity. The data obtained from skin acceleration, combined with a physiologically based impedance-based model [5], allows the customization of specific parameters for individual patient assessment of the vocal folds. Achieving a patient-specific and accurately fitted model is crucial for obtaining a precise estimation of the aerodynamic source of voice sound at the glottis [38, 34, 74].

We consider to represent the subglottal system through the block diagram in Figure 2.7, where $u(t)$ is the input signal GVV, and $y(t)$ is the measured signal ACC obtained with a neck-mounted accelerometer, and $w(t)$ is assumed to be white measurement noise. The input GVV is not directly measured, however, it is obtained from OVV using a linear prediction (LP) filter [94, 95]. Simultaneously, the subglottal system, represented by the transfer function $G(p)$ is decomposed into two functions, $F(p)$ and

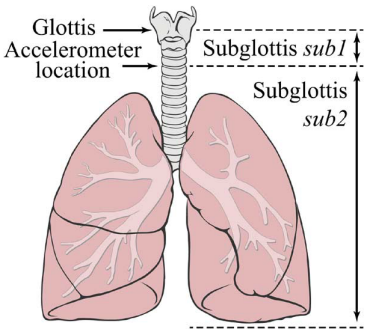


Figure 2.6 The tract section above the accelerometer is labeled as $sub1$, while the tract section below this point, extending to the alveoli, is labeled as $sub2$. Taken from [5]

$T(p)$. The input to filter $F(p)$ is the signal GVV, producing the output signal GVV^*, serving as the input to filter $T(p)$. The parameters of $T(p)$ are determined through system identification using synchronized GVV^* and ACC data. The resulting model $G(p) = T(p)F(p)$ can be utilized for ambulatory monitoring, allowing the estimation of GVV input signal from ACC output signal measurements.

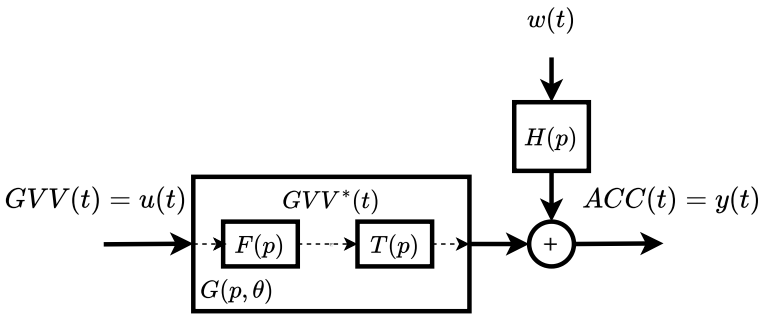


Figure 2.7 Block diagram representation of the phonatory system

On the other hand, according to [5], the relationship between the glottal volume velocity U_{sub} (i.e. GVV) to the neck surface acceleration $\dot{U}_{skin}$ (i.e. ACC), can be represented as a mechano-acoustic transmission line:

$$T_{skin}(\omega) = -\frac{\dot{U}_{skin}}{U_{sub}} = \frac{H_{sub1}(\omega)Z_{sub2}(\omega)j\omega}{Z_{sub2}(\omega) + Z_{skin}(\omega)} \tag{2.23}$$

where

$\dot{U}_{skin}$ is obtained from differentiating the skin volume velocity U_{skin},

$H_{sub1} = \dfrac{U_{sub1}}{U_{sub}}$ is the frequency response of the transfer function of subglottal section $sub1$,

$j\omega$ represents the ideal derivative filter,

Z_{skin} refers to the neck skin impedance given in (2.24), and

Z_{sub2} is the frequency-dependent driving-point impedance of the corresponding tract [5].

As described in [96], $|Z_{skin}(\omega)| \gg |Z_{sub_2}(\omega)|$, which allows us to assume that $Z_{sub_2}(\omega) + Z_{skin}(\omega) \approx Z_{skin}(\omega)$.

Moreover,

$$Z_{skin}(\omega) = \frac{1}{A_{acc}} \left(Z_m(\omega) + Z_{rad}(\omega)\right) \tag{2.24}$$

where

$$Z_m(\omega) = R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) \tag{2.25}$$

and where R_m is the neck skin resistance, M_m corresponds to accelerometer area density on the neck skin, K_m is the skin stiffness, and Z_{rad} is the radiation impedance due to accelerometer, given by

$$Z_{rad}(\omega) = j\omega M_{acc} \tag{2.26}$$

Then, replacing in (2.24), we have that

$$Z_{skin}(\omega) = \frac{1}{A_{acc}} \left[R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) + j\omega M_{acc}\right] \tag{2.27}$$

where: M_{acc} and A_{acc} are the accelerometer mass and contact area.

Taking into account the unique characteristics of each patient, as highlighted in prior studies [5], [34], it is essential to obtain subject-specific parameters for the IB model. Typical values for these parameters are given in Table 2.2, which includes R_m, M_m, and K_m, defined previously and also the trachea length $L_{trachea}$ and the accelerometer position L_{sub1}. Scaling factors Q_i, $\{i = 1, 2, 3\}$, have been typically used for the parameters R_m, M_m, and K_m, respectively, and they will be used in the sequel.

From (2.23)-(2.27), and using the scaling factors Q_i, we can combine the impedance and transfer functions models to obtain:

$$Z_{skin}(\omega) = \frac{1}{A_{acc}} \left[Q_1 R_m + Q_2 M_m j\omega - Q_3 j\frac{K_m}{\omega} + j\omega M_{acc}\right] \tag{2.28}$$

Table 2.2 Values of the mechanical impedance model parameters of the neck skin surface

subject specific parameter		unit
$R_m =$	2320	$[g \cdot s^{-1} cm^{-2}]$
$M_m =$	2.4	$[g \cdot cm^{-2}\ ]$
$K_m =$	491000	$[dyn \cdot cm^{-3}]$
$L_{trachea} =$	10	[cm]
$L_{sub1} =$	5	[cm]

Assuming equality conditions between frequency and Laplace domains and letting p to be the differential operator $p = \frac{d}{dt}$:

$$Z_{skin}(p) = \frac{1}{A_{acc}} \left[Q_1 R_m + Q_2 M_m p + \frac{Q_3 K_m}{p} + p M_{acc} \right] \tag{2.29}$$

the subglottal system transfer function is given by:

$$T_{skin}(p) = \frac{H_{sub1}(p) Z_{sub2}(p) p}{\frac{1}{A_{acc}} \left(Q_1 R_m + p Q_2 M_m + Q_3 \frac{K_m}{p} + p M_{acc} \right)} \tag{2.30}$$

From Figure 2.7, an approximation of the IB model is presented as in [93], i.e,

$$\begin{aligned} T_{skin}(p) &= G(p) = F(p) T(p) \\ &= \frac{H_{sub1}(p) Z_{sub2}(p) p^2}{\frac{1}{A_{acc}} \left[p^2 (Q_2 M_m + M_{acc}) + p Q_1 R_m + Q_3 K_m \right]} \end{aligned} \tag{2.31}$$

where we define

$$F(p) = H_{sub1}(p) Z_{sub2}(p) \tag{2.32}$$

and

$$T(p) = \frac{p^2 b}{p^2 + p a_1 + a_2} \tag{2.33}$$

The parameters of transfer function $T(p)$ in (2.33), are then given by

$$\begin{aligned} b &= \frac{A_{acc}}{Q_2 M_m + M_{acc}} \\ a_1 &= \frac{Q_1 R_m}{Q_2 M_m + M_{acc}} \\ a_2 &= \frac{Q_3 K_m}{Q_2 M_m + M_{acc}} \end{aligned} \tag{2.34}$$

From the developments above, we will proceed with estimating the two transfer functions $F(p)$ and $T(p)$. First, a model for $F(p)$ will be estimated based on the frequency responses of $H_{sub_1}(\omega)$ and $Z_{sub_2}(\omega)$. These frequency responses are obtained by simulating the high-order model that considers the branching structure of the tracheobronchial airways with 24 levels of depth [97, 13, 5]. The hypothesis for estimating these models relies on the model order reduction.

On the other hand, once $F(p)$ has been estimated it will be used to obtain the signal GVV^* from GVV. Then $T(p)$ will be estimated as a second-order system (with the same number of poles and zeros), where $Q_{\{1,2,3\}}$ are the parameters to be estimated.

2.3 Continuous Linear Time Invariant model 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 [34]. 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 [5].

In the preceding section, the subglottal model was divided into two series-connected models based on Figure 2.7. Both models were linear time-invariant (LTI), yet featured a partially constrained structure, particularly in the mechanical part of the mechano-acoustic model. In this section, we refrain from imposing any constraints on the model structure or order. Instead, we will perform estimations across a set of candidate models, each having different orders and relative degrees. To select the appropriate model we will apply the information criteria theory.

In this section linear time-invariant (LTI) models are proposed to represent the complete subglottal system, fitted by using SRIVC. 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.

As mentioned before, 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 [5]. 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 [5, 26, 27].

The skin acceleration is measured by attaching an accelerometer to the neck surface between the thyroid prominence and the suprasternal notch ([25]). 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 [72, 28].

We consider the phonatory system as represented by the block diagram shown in Figure 2.8, 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 (we have assumed no measurement error) using a linear prediction (LP) filter [94, 95].

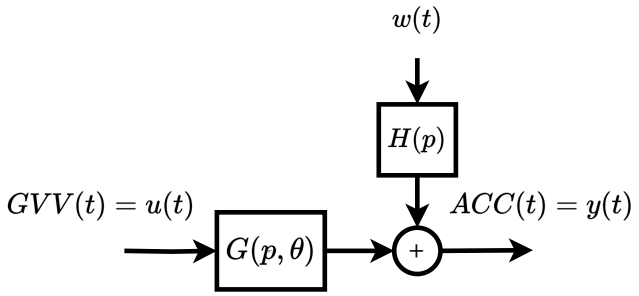


Figure 2.8 Block diagram representation of the phonatory system

The subglottal system in Figure 2.8 can be expressed as the following continuous-time model

$$y(t) = G(p, \theta)u(t) + H(p)w(t) \tag{2.35}$$

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

$$G(p,\theta) = \frac{B(p,\theta)}{A(p,\theta)} = \frac{b_m p^m + b_{m-1}p^{m-1} + ... + b_0}{p^n + a_{n-1}p^{n-1} + ... + a_0} \tag{2.36}$$

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

Our interest in this paper is to estimate the parameter vector θ in (2.36), 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

$$\begin{aligned} y_u(t) &= G(p,\theta)u(t) \\ y(t_k) &= y_u(t_k) + w(t_k) \end{aligned} \tag{2.37}$$

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 $w(t_k) \sim \mathcal{N}(0, \sigma_w^2)$.

2.4 Chapter summary

This chapter introduced two models physically based utilized for the study of the speech system, along with the proposal of a more general Linear Time Invariant model for the same purpose. Firstly, the Body Cover Model was described, and its modeling in bond-graph was presented, leading to the derivation of the Port-Hamiltonian model. Subsequently, the impedance-based model was introduced, presented as the product of two models—one to be estimated from frequency response and a second-order model for which sampled input-output data will be used for estimation. Finally, the structure of the proposed LTI system is described to represent the subglottal model.

Chapter 3

Model identification methods

The different models proposed for the subglottal system are defined by parameters that are subject dependents. In this chapter, the theoretical fundamentals of the methods employed to estimate the parameters of the proposed models using data from the subglottal system are introduced.

Specifically, the theory of Maximum Likelihood and Simple Refined Instrumental Variable for Continuous-time model will be introduced. Additionally, three information criteria for selecting the order that best fits the dynamic system will be presented, along with other statistical tools. The goal is to gather comprehensive statistical information about the estimated models.

While a specific case of estimation is presented based on the frequency response of a system, this is done for convenience since the data for estimation in this particular system precisely comprises its frequency response. For all other cases, the estimation of model coefficients is carried out using sampled input-output data from the system.

3.1 Maximum Likelihood

Maximum Likelihood (ML) is a general parameter estimation method [98] that has been extensively studied and applied in statistics and system identification literature. The likelihood function is the conditional probability density function (pdf) of the observations given the parameters [99, 100].

Given a model defined by a parameter vector β_0; let y_t be a stochastic variable whose distribution depends on an unknown vector β_0, and ω_t (see Figure 3.1) a stochastic disturbance and u_t a deterministic input.

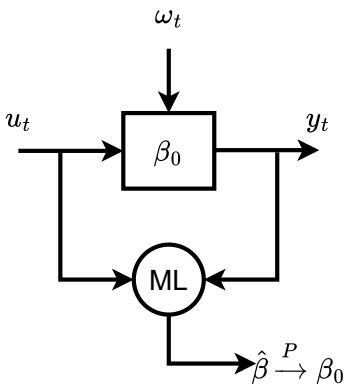


Figure 3.1 Block diagram representation of maximum likelihood algorithm

From measured data $y_{1,2,\ldots,N}$, the likelihood function is given by:

$$L_N(\beta) = p(y_{1:N}|\beta) \tag{3.1}$$

and maximizing (3.1) we obtain the ML estimate of β_0:

$$\hat{\beta}_N = \arg\max_{\beta} L_N(\beta) \tag{3.2}$$

This estimator satisfies general properties that are discussed in section 3.1.2.

Consider the next example, where the data $y_t \sim \mathcal{N}(\mu, \sigma^2)$ is independently and identically distributed (i.i.d.) with Gaussian distribution. The goal is to find the ML estimate for $\beta_0 = [\theta^T \quad \sigma^2]^T$ using the data $y_{1:N}$.

In order to maximize (3.2) we need to know the pdf $p(y_{1:N}|\beta)$. Considering that the data $y_{1:N}$ is i.i.d. and normal distributed, the pdf can be expressed as follows.

$$\begin{aligned} p(y_{1:N}|\beta) &= \prod_{t=1}^{N} p(y_t|\beta) \\ &= \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2}\right\} \end{aligned} \tag{3.3}$$

Since optimizing a function with a product term could be challenging, it is common to leverage the properties of logarithms to simplify the operation. With this approach,

the natural logarithm is applied to the likelihood function for ease of computation.

$$\begin{aligned} \log(L_N(\beta)) &= \log\left[\prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2}\right\}\right] \\ l_N(\beta) &= \sum_{t=1}^{N}\left[\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \left(-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2}\right)\right] \\ l_N(\beta) &= -\frac{N}{2}\log(\sigma^2) - \frac{N}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(y_t-\mu)^2 \end{aligned} \tag{3.4}$$

Taking into account that in the previous equation, the term $\frac{N}{2}\log(2\pi)$ does not contain the parameters of interest, from the optimization standpoint, we can disregard it.

$$l_N(\beta) = -\frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(y_t-\mu)^2 \tag{3.5}$$

The bijective transformation $\alpha = \frac{1}{\sigma^2}$ is then carried out, followed by the optimization in the new domain. The likelihood in the new domain is given by:

$$\begin{aligned} \bar{l}_N(\alpha,\mu) &= -\frac{N}{2}\log\left(\frac{1}{\alpha}\right) - \frac{\alpha}{2}\sum_{t=1}^{N}(y_t-\mu)^2 \\ &= \frac{N}{2}\log(\alpha) - \frac{\alpha}{2}\sum_{t=1}^{N}(y_t-\mu)^2 \end{aligned} \tag{3.6}$$

Taking the derivative with respect to alpha and setting it equal to zero

$$\begin{aligned} \frac{\partial \bar{l}_N(\alpha,\mu)}{\partial\alpha} &= 0 \\ \frac{\partial \bar{l}_N(\alpha,\mu)}{\partial\alpha} &= \frac{N}{2\alpha} - \frac{1}{2}\sum_{t=1}^{N}(y_t-\mu)^2 \\ 0 &= \frac{N}{2\alpha} - \frac{1}{2}\sum_{t=1}^{N}(y_t-\mu)^2 \\ \frac{N}{2\alpha} &= \frac{1}{2}\sum_{t=1}^{N}(y_t-\mu)^2 \end{aligned} \tag{3.7}$$

By replacing $\alpha = \frac{1}{\sigma^2}$

$$\sigma^2 = \frac{1}{N}\sum_{t=1}^{N}(y_t-\mu)^2 \tag{3.8}$$

Then, by replacing σ^2 in (3.5) is obtained the concentrated cost function $V(\mu)$.

$$V(\mu) = -\frac{N}{2}\log\left(\frac{1}{N}\sum_{t=1}^{N}(y_t-\mu)^2\right) - \frac{N}{2} \tag{3.9}$$

Notice that the second term of the concentrated cost functional does not depend on μ, so it is only necessary to optimize the argument of the logarithm

$$V(\mu) = \frac{1}{N}\sum_{t=1}^{N}(y_t-\mu)^2 \tag{3.10}$$

Applying the derivative with respect to μ to the concentrated cost functional and setting it equal to zero we obtain the fallowing almost surely (a.s.) convergences:

$$\begin{aligned} \hat{\mu} &= \frac{1}{N}\sum_{t=1}^{N} y_t \quad \xrightarrow{a.s.} \quad \mathbb{E}\{y_t\} = \mu \\ \hat{\sigma}^2 &= \frac{1}{N}\sum_{t=1}^{N}(y_t-\hat{\mu})^2 \quad \xrightarrow{a.s.} \quad \sigma^2 \end{aligned} \tag{3.11}$$

3.1.1 Maximum likelihood for linear algebraic systems

Giving the following linear algebraic systems:

$$y_t = G(q^{-1},\theta_0)u_t + H(q^{-1},\theta_0)\omega_t \tag{3.12}$$

where q^{-1} is the backward shift operator and:

$$\begin{aligned} G(q^{-1},\theta) &= \frac{B(q^{-1},\theta)}{A(q^{-1},\theta)} \\ H(q^{-1},\theta) &= \frac{C(q^{-1},\theta)}{D(q^{-1},\theta)} \end{aligned} \tag{3.13}$$

and $\omega_t \sim \mathcal{N}(0,\sigma^2)$, u_t is a deterministic known signal. $y_{1:N}$, and $u_{1:N}$.

From (3.1):

$$L_N(\beta) = p(y_{1:N}|u_{1:N},\beta) \tag{3.14}$$

Taking into account Bayes theorem:

$$L_N(\beta) = p(y_1|u_{1:N},\beta)\prod_{t=2}^{N} p(y_t|y_{1:t-1},u_{1:N},\beta) \tag{3.15}$$

Re-writing (3.12):

$$y_t = G(q^{-1},\theta)u_t + (H(q^{-1},\theta)-1)\omega_t + \omega_t \tag{3.16}$$

and solving for ω_t:

$$w_t = \frac{1}{H(q^{-1},\theta)}\left\{y_t - G(q^{-1},\theta)u_t\right\} \tag{3.17}$$

then:

$$y_t = G(q^{-1},\theta)u_t + \left[H(q^{-1},\theta)-1\right]\frac{1}{H(q^{-1},\theta)}\left\{y_t - G(q^{-1},\theta)u_t\right\} + \omega_t \tag{3.18}$$

next assuming $H(0,\theta)=1$ and $G(0,\theta)=0$, i.e. there is a delay from the input to the output signal, and the gain of the noise transfer function is considered into the noise variance. This means that:

$$\begin{aligned} G(q^{-1},\theta) &= g_1 q^{-1} + g_2 q^{-2} + ... \\ H(q^{-1},\theta) &= 1 + h_1 q^{-1} + h_2 q^{-2} + ... \end{aligned} \tag{3.19}$$

where:

$$p(y_t|y_{1:t-1}) = \mathcal{N}(\hat{y}_{t|t-1},\sigma^2) \tag{3.20}$$

$$\hat{y}_{t|t-1} = G(q^{-1},\theta)u_t + \frac{H(q^{-1},\theta)-1}{H(q^{-1},\theta)}\left\{y_t - G(q^{-1},\theta)u_t\right\} \tag{3.21}$$

After some basic mathematical operations:

$$\hat{y}_{t|t-1} = y_t - \frac{1}{H(q^{-1},\theta)}\left(y_t - G(q^{-1},\theta)u_t\right) \tag{3.22}$$

thus

$$\varepsilon_t(\theta) = y_t - \hat{y}_{t|t-1} = \frac{1}{H(q^{-1},\theta)}\left(y_t - G(q^{-1},\theta)u_t\right) \tag{3.23}$$

where $\varepsilon_t(\theta)$ is known as the prediction error. Then, from (3.20) we have:

$$\begin{aligned} p(y_t|y_{1:t-1}) &= \mathcal{N}(\hat{y}_{t|t-1},\sigma^2) \\ &= \prod_{t=1}^{N}\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2}\frac{(y_t-\hat{y}_{t|t-1})^2}{\sigma^2}\right\} \\ &= \prod_{t=1}^{N}\frac{1}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\varepsilon_t(\theta)^2\right\} \end{aligned} \tag{3.24}$$

Finally assuming $y_1 \sim \mathcal{N}(0, \sigma^2) \quad \Rightarrow \varepsilon_1 = y_1$,

$$\begin{aligned} l_N(\beta) &= \log\left[p(y_1|\beta)\right] - \frac{N-1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{N}\varepsilon_t(\theta)^2 \\ &= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{N}\varepsilon_t(\theta)^2 \end{aligned} \tag{3.25}$$

After proceedings from (3.5) to (3.10) is obtained:

$$\begin{aligned} \hat{\theta}_{ML} &= \arg\min_{\theta} V_N(\theta) \\ \hat{\sigma}^2 &= \frac{1}{N}\sum_{t=2}^{N}\varepsilon_t(\hat{\theta}_{ML})^2 \\ V_N(\theta) &= \frac{1}{N}\sum_{t=2}^{N}\varepsilon_t(\theta)^2 \end{aligned} \tag{3.26}$$

The significance of the ML approach to parameter estimation in statistical models becomes apparent. ML aims to maximize the likelihood function to optimize the agreement between the model predictions and observed data. It can be shown that $\hat{y}_{t|t-1} = \mathbb{E}\{y_t|y_{1:t-1}\}$, where $\mathbb{E}\{\cdot\}$ denotes the expectation operator, and directly minimizing $V_N(\theta)$ constitutes the Prediction Error Method (PEM).

3.1.2 General properties of the ML estimator

In the previous sections we presented maximum likelihood estimator. In this section we present the properties of this estimator. These fundamental properties underscore the utility of the maximum likelihood approach in parameter estimation for statistical models. Some properties of maximum likelihood estimator are listed below [99]:

1. **Principle of invariance**. The ML estimator of a parametric function, can be computed by taking the parametric function of the ML estimates, i.e., if $\hat{\theta}$ is a ML estimator (MLE) of $\theta \in \Theta \subset \mathbb{R}^p$, then $g(\hat{\theta})$ is a MLE of $g(\theta)$, function $g : \Theta \to \mathbb{Z} \subset \mathbb{R}^n$, and $n \leqslant p$.

2. **Cramer-Rao lower bound**. An unbiased estimator that satisfies the Cramer-Rao lower bound, is then also a MLE.

3. **Consistency**. Let $\hat{\theta}_N$ be a MLE of θ of a random variables $y = (x_1, ...x_N)^T$, independent and identically distributed (i.i.d), then $\hat{\theta}_N \xrightarrow{a.s} \theta$.

4. **Asymptotic normality**. Let $\hat{\theta}_N$ be a MLE of θ based on N i.i.d variables $y = (x_1, ..., x_N)^T$, then $\hat{\theta}_N$ converges in distribution to a normal random variable.

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{law} \beta \tag{3.27}$$

   where

$$\beta \sim \mathcal{N}(0, \mathcal{I}_\theta^{-1})$$

   and $\mathcal{I}_\theta$ is the average value of Fisher's information matrix per sample.

5. **Efficiency**. $\hat{\theta}$ is efficient in the sense that it asymptotically attains the Cramer-Rao lower bound.

With the properties of ML now established, it is crucial to acknowledge that, despite being an optimal estimator, it may exhibit sensitivity to deviations from the true distribution [101]. Furthermore, uncertainty stemming from errors in measurements, inputs, implementation, model, or numerical errors may impact the quality of estimates [102].

3.2 Simple Refined Instrumental Variable for Continuous time OE model

Natural processes evolve in continuous time and, as a consequence, in the early years of system identification theory, the focus was primarily on estimating continuous-time models. This approach was later overshadowed by discrete-time system identification, due to the technological development of digital computers and data acquisition systems. Nevertheless, in recent years there has been a renewed interest in the estimation of continuous-time models [81, 103].

The choice between continuous- or discrete-time modeling depends both on the application and the available data. The literature highlights a range of advantages and benefits associated with each option. However, one notable drawback in discrete-time identification arises for small sampling time, resulting in transfer functions having poles very close to the point $z = 1$, leading to numerical problems [81, 83]. On the other hand, continuous-time models offer advantages such as the interpretation of estimated parameters, reflecting real system properties, and allowing a better understanding of the physical behavior. Moreover, continuous-time models are independent of the sampling

period, making them suitable for applications with fast sampling rates [104, 103]. Finally, a continuous-time model can be discretized for different sampling rates if needed [43].

Simplified Refined Instrumental Variable method for Continuous-time systems (SRIVC) [105] is a widely and reliable algorithm for parameter estimation of continuous-time models from sampled data. This iterative method is based on instrumental variables (IV) that generate parameter-dependant filters that are applied to input-output data at each iteration. It has been shown that SRIVC provides a consistent and optimal estimator for additive white noise [106].

From the model defined in (2.37)

$$\begin{aligned} y_u(t) &= G(p,\theta)u(t) \\ y(t_k) &= y_u(t_k) + w(t_k) \end{aligned} \tag{3.28}$$

The error function used in SRIVC is the so-called generalized equation error (GEE) [107] given by

$$\varepsilon(t_k,\theta) = y(t_k) - \frac{B(p,\theta)}{A(p,\theta)}u(t_k) \tag{3.29}$$

In the method, the input and output signals are filtered by

$$L(p,\theta) = \frac{1}{A(p,\theta)} \tag{3.30}$$

Then, (3.29) can be written as

$$\varepsilon(t_k,\theta) = A(p,\theta)\tilde{y}(t_k) - B(p,\theta)\tilde{u}(t_k) \tag{3.31}$$

where $\tilde{\chi}(t_k)$ are the input and output filtered signals.

An important issue here is that the filter $L(p,\theta)$ is not known a priori, and, thus, an initial parameter estimate is required. Notice that, using the filtered signals, leads to (3.31) which is a linear regression [44].

Additionally, the instrumental variable approach allows to solve bias problem when the measurement noise is colored, i.e., when $H(p) \neq 1$ in (2.35).

The cost function associated with (3.31) can be (iteratively) optimized as follows:

$$\tilde{y}^{(n)}(t_k,\hat{\theta}^j) = \tilde{\phi}^T(t_k,\hat{\theta})\theta^{j+1} + \varepsilon(t_k,\hat{\theta}^j) \tag{3.32}$$

where the regressor vector is given by

$$\begin{aligned}\tilde{\phi}^T(t_k,\hat{\theta}) = \Big[&\tilde{u}^{(m)}(t_k,\hat{\theta}^j),\cdots,\tilde{u}(t_k,\hat{\theta}^j),\\ &-\tilde{y}^{(n-1)}(t_k,\hat{\theta}^j),\cdots,-\tilde{y}(t_k,\hat{\theta}^j)\Big]\end{aligned} \tag{3.33}$$

where $\hat{\theta}^j$ and $\hat{\theta}^{j+1}$ are the estimated parameters at the j^{th} and $(j+1)^{th}$ iterations, respectively, and the superscripts (n), (m), and $(n-1)$ correspond to (approximate) time derivatives. The optimal SRIVC estimator is then given by

$$\begin{aligned}\hat{\theta}^{j+1}_{SRIVC} = &\left[\frac{1}{N}\sum_{k=1}^{N}\tilde{Z}(t_k,\hat{\theta}^j)\tilde{\phi}^T(t_k,\hat{\theta}^j)\right]^{-1}\\ &\left[\frac{1}{N}\sum_{k=1}^{N}\tilde{Z}^T(t_k,\hat{\theta}^j)\tilde{y}^{(n)}(t_k,\hat{\theta}^j)\right]\end{aligned} \tag{3.34}$$

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

$$\tilde{Z}^{(i)}(t_k,\hat{\theta}^j) = \frac{p^{(i)}}{A(p,\hat{\theta}^j)}\tilde{\phi}(t_k,\hat{\theta}^j) \tag{3.35}$$

A summary of the method is shown in Algorithm 1 and more details can be found, for example, in [106, 108, 82].

The iterations of SRIVC stop under two conditions: the first is when the number of iterations is reached, or when the error in parameter estimation becomes sufficiently small, resulting in a rapidly converging algorithm. In this work, the CONTSID Toolbox for Matlab is employed, which has proven to be highly robust in practical applications [109, 110].

3.2.1 Continuous-time modeling using frequency-domain data

The problem of identification in the frequency domain lies in obtaining an estimate of the system based on its frequency response. The frequency domain approach allows, for example, for the combination of data from various experiments, especially when state-space models are employed, providing a concise and flexible representation of the system. Additionally, data can be pre-filtered, and large datasets can be condensed [111]. Alternatively, emphasis can be placed on a specific bandwidth to enhance robustness [112].

Algorithm 1 SRIVC Algorithm

procedure SRIVC(data, model order (n, m), initialization method, tolerance value (ε), number of iterations $(MaxIter)$)
  Initialization: generate an initial estimate vector parameter, and use it to define the initial continuous-time prefilter $(L(p, \theta_0))$
  Iterative estimation: j $\leftarrow 1$, k $\leftarrow 0$
  while convergence $\leftarrow 1$ && j $\leq$ MaxIter **do**
    Prefilter sampled input-output data to generate the filtered derivatives of these signals

$$\tilde{\phi}^T(t_k, \hat{\theta}) \leftarrow \Big[\tilde{u}^{(m)}(t_k, \hat{\theta}^j), \cdots, \tilde{u}(t_k, \hat{\theta}^j), \\ -\tilde{y}^{(n-1)}(t_k, \hat{\theta}^j), \cdots, -\tilde{y}(t_k, \hat{\theta}^j)\Big]$$

    Compute estimation

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

    if $\dfrac{\|\hat{\theta}^{j+1}_{SRIVC} - \hat{\theta}^j_{SRIVC}\|_2}{\|\hat{\theta}^j_{SRIVC}\|_2} < \varepsilon$ **then**
      k $\leftarrow 1$
    end if
    $j \leftarrow j + 1$
  end while
  Output: $\hat{\theta}^j_{SRIVC}$
end procedure

A model as the one presented in (2.35) can be considered in the frequency domain as a relationship between input and output spectra

$$Y(j\omega) = F_0(j\omega, \theta)U(j\omega) + W(j\omega) \tag{3.36}$$

where $F_0(j\omega)$ denotes the measurement of the system frequency response, and the noise filter is considered to be unitary.

Assuming $F(j\omega, \theta)$ is included in a similar model as the the one in (2.36), the output form (3.36) can be written as [113, 114]:

$$(j\omega)^n Y(j\omega) = \phi(j\omega)\theta_0 + A(j\omega)W(j\omega) \tag{3.37}$$

with the regressor

$$\begin{aligned}\phi(j\omega_k) = [&U(j\omega_k) \dots (j\omega_k)^m U(j\omega_k) \\ &-Y(j\omega_k) \cdots - (j\omega_k)^{n-1} Y(j\omega_k)\big]\end{aligned} \tag{3.38}$$

where ω_k denotes the frequency grid of the measured frequency response.

The estimation problem is then to estimate the parameter vector θ using the measured frequency domain data and minimizing a cost function [115], i.e.,

$$\hat{\theta} = \arg\min_{\theta} V(\theta) \tag{3.39}$$

where

$$V(\theta) = \sum_{k=1}^{N} \left| F_0(j\omega_k) - \frac{B(j\omega_k, \theta)}{A(j\omega_k, \theta)} \right|^2 \tag{3.40}$$

Then we have that the estimate of the parameter vector is given by

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} \left| Z(j\omega_k)L(j\omega_k)Y(j\omega_k) - \phi^*(j\omega_k)\theta \right|^2 \tag{3.41}$$

where $(\cdot)^*$ denotes the complex conjugate transpose, and the instrument $Z(j\omega_k)$ and the filter $L(j\omega_k)$ are given by

$$Z(j\omega_k) = L(j\omega_k)\mathring{\phi}(j\omega_k) \tag{3.42}$$

$$L(j\omega_k) = \frac{1}{A_0(j\omega_k)} \tag{3.43}$$

where $\mathring{\phi}(j\omega_k)$ is the noise free part of $\phi(j\omega_k)$.

The algorithm for SRIVC in the frequency domain closely resembles the one presented in 1. The complete algorithm for the method in the frequency domain can be found in [113].

3.3 Model Order selection

Model order selection is an important task in the analysis of time series, signal processing, and system identification [116]. To select the "best" order for the subglottal LTI model, information criteria such as Akaike, Bayesian and Young Information Criteria (AIC, BIC and YIC respectively) can be used. Furthermore, in this section, we will introduce the determination coefficient and the coherence method, both of which will be used for characterizing the estimated models.

Information criteria play a crucial role in selecting mathematical models that strike a balance between accuracy and complexity. By applying these criteria, more complex models (i.e., higher-order models) are penalized, helping to avoid overfitting—where the model fits the training data well but performs poorly on new data. A key advantage of using information criteria is that they facilitate an objective comparison of candidate models, ensuring an optimal trade-off between model fit and simplicity [117–119].

3.3.1 Akaike information criterion

The AIC is a fundamental tool in model selection within statistics and computer science. Developed by Hirotugu Akaike in the 1970s, AIC addresses the challenge of choosing between different statistical models by balancing model fit with complexity. Essentially, AIC aims to choose, from a set of candidate models, the one that offers a good fit to measured data while also being simple enough to avoid overfitting. Since its justification relies solely on the conventional large-sample properties of maximum likelihood estimators [120], this criterion has evolved into an indispensable tool in scientific practice, extensively applied across diverse fields.

The development of AIC (3.44) is based on 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

[121].

$$AIC \quad = -2\ln(\text{maximum likelihood}) + 2n \tag{3.44}$$

The Kullback-Liebler is an information criterion, also known as a discrepancy function. It involves the true probability density function $p_0(y)$ of observed data y and the *pdf* $\hat{p}(y)$ of an estimate y [122, 121]. The KL is given by

$$D(p_0, \hat{p}) = \mathbb{E}_0 \left\{ \ln \left(\frac{p_0(y)}{\hat{p}(y)} \right) \right\} = \mathbb{E}_0 \left\{ \ln(p_0(y)) \right\} - \mathbb{E}_0 \left\{ \ln(\hat{p}(y)) \right\} \tag{3.45}$$

where $\hat{p}(y)$ is the likelihood of the model (see (3.1)), being $-\mathbb{E}\left\{\ln \hat{p}(y)\right\}$ the model dependent part.

Lemma 3.3.1. *The discrepancy function is almost positive definite $D(p_0, \hat{p} \geq 0)$ with equality if and only if $\hat{p}(y) \equiv p_0(y)$, i.e.*

$$D(p_0, \hat{p}) \geq 0 = 0 \Longleftrightarrow \hat{p}(y) \equiv p_0(y) \tag{3.46}$$

The previous (3.46) can be verify [121] considering that

$$\begin{aligned} -\ln(\lambda) &\geq 1 - \lambda \quad \forall \lambda > 0 \\ -\ln(\lambda) &\geq 1 - \lambda \Longleftrightarrow \lambda = 1 \end{aligned} \tag{3.47}$$

Then, assuming

$$\lambda(y) = \frac{\hat{p}(y)}{p_0(y)} \tag{3.48}$$

Proof.

$$\begin{aligned} D(p_0, \hat{p}) &= \int p_0(y) \ln \left(\frac{p_0(y)}{\hat{p}(y)} \right) dy \\ &= \int p_0(y) \ln \left(\frac{1}{\lambda(y)} \right) dy \\ &= \int p_0(y) \left[-\ln \left(\lambda(y) \right) \right] dy \\ &\geq \int p_0(y) \left(1 - \lambda(y) \right) dy \\ &= \int p_0(y) \left(1 - \frac{\hat{p}(y)}{p_0(y)} \right) dy \end{aligned} \tag{3.49}$$

This shows $D(p_0, \hat{p}) = 0$ only holds if $\hat{p}(y) \equiv p_0(y)$, for other cases it will give an information of the loss[1]. □

The maximization of the model depend part in (3.45) is equivalent to minimize

$$I(p_0, \hat{p}(y)) = \mathbb{E}_0 \left\{ \ln \left(\hat{p}(y) \right) \right\} \tag{3.50}$$

As the model likelihood function is not available, we can use the following

$$\hat{p}(y) = p(y, \hat{\theta}) \tag{3.51}$$

Then:

$$I(p_0, p(y, \hat{\theta})) = \mathbb{E} \left\{ \ln \left(p(y, \hat{\theta}) \right) \right\} \tag{3.52}$$

To evaluate the previous $\mathbb{E}\{\cdot\}$ is not possible because as was commented earlier, the true *pdf* is not available.Therefore, the unbiased estimator of $I(p_0, p(y, \hat{\theta}))$ may be used.

$$\hat{I} = \ln \left(p(y, \hat{\theta}) \right) \tag{3.53}$$

Alternatively, a second-order Taylor expansion around $\hat{\theta}^n$ can be used, because the order selection rule that maximizes (3.53) does not have satisfactory properties [121].

$$\begin{aligned} \ln \left(p_n(y, \theta^n) \right) \approx \ln \left(p_n(y, \hat{\theta}^n) \right) + \left(\theta^n - \hat{\theta}^n \right)^T \left[\left. \frac{\partial \ln \left(p_n(y, \theta^n) \right)}{\partial \theta^n} \right|_{\theta^n = \hat{\theta}^n} \right] + \frac{1}{2} \left(\theta^n - \hat{\theta}^n \right)^T \\ \times \left[\left. \frac{\partial^2 \ln \left(p_n(y, \theta^n) \right)}{\partial \theta^n \partial (\theta^n)^T} \right|_{\theta^n = \hat{\theta}^n} \right] \left(\theta^n - \hat{\theta}^n \right) \end{aligned} \tag{3.54}$$

where n stands for the model order.

Considering that $\hat{\theta}^n$ is the parameter estimate that maximizes $p_n(y, \theta^n)$, the second term in (3.54) is zero. Therefore, it can be rewritten as:

$$\ln \left(p_n(y, \theta^n) \right) \approx \ln \left(p_n(y, \hat{\theta}^n) \right) + \frac{1}{2} \left(\theta^n - \hat{\theta}^n \right)^T J \left(\theta^n - \hat{\theta}^n \right) \tag{3.55}$$

Under regularity conditions, the *pdf* of ML estimate $\hat{\theta}$, as the dataset size $N \to \infty$, converges to a Gaussian *pdf* with mean θ and covariance equal to the Cramér-Rao

[1]This is essentially the proof on [122] on page 82.

bound. Therefore:

$$p_n(\hat{\theta}^n) = \frac{2}{(2\pi)^{\frac{n}{2}}\,|-J|^{\frac{1}{2}}} e^{\left\{-\frac{1}{2}(\hat{\theta}-\theta)^T J(\hat{\theta}-\theta)\right\}} \tag{3.56}$$

$$J = -\mathbb{E}\left\{\frac{\partial^2 \ln\left(p_n(y,\theta^n)\right)}{\partial\theta^n \partial(\theta^n)^T}\right\} \tag{3.57}$$

The expectation in (3.52) is then obtained as:

$$\begin{aligned} I(p_0, p_n(y,\hat{\theta}_y^n)) &= \mathbb{E}\left\{\ln\left(\hat{p}(y)\right)\right\} \\ &= \mathbb{E}\left\{\ln\left(p_n(y,\hat{\theta}^n)\right) - \frac{1}{2}(\theta^n - \hat{\theta}^n)^T J(\theta^n - \hat{\theta}^n)\right\} \\ &= \mathbb{E}\left\{\ln\left(p_n(y,\hat{\theta}^n)\right)\right\} - \mathbb{E}\left\{\frac{1}{2}(\theta^n - \hat{\theta}^n)^T J(\theta^n - \hat{\theta}^n)\right\} \end{aligned} \tag{3.58}$$

and

$$\begin{aligned} \mathbb{E}\left\{\frac{1}{2}(\theta^n - \hat{\theta}^n)^T J(\theta^n - \hat{\theta}^n)\right\} &= \frac{1}{2} tr\left[J\mathbb{E}\left\{(\theta^n - \hat{\theta}^n)(\theta^n - \hat{\theta}^n)^T\right\}\right] \\ &= \frac{1}{2} tr\left[I_n\right] \\ &= \frac{n}{2} \end{aligned} \tag{3.59}$$

where tr is the trace operator.

An unbiased estimator of the KL divergence is then provided by:

$$\ln\left(p_n(y,\hat{\theta}^n)\right) - \frac{n}{2} \tag{3.60}$$

Now, if a vector x of the same length as y is denoted, with the same *pdf*, and is also independent of y, with $\hat{\theta}_x^n$ the ML estimator of x, then it can be written as:

$$\ln(\hat{p}(y)) = \mathbb{E}_x\left\{\ln\left(p_n(y,\hat{\theta}_x^n)\right)\right\} \tag{3.61}$$

substituting in (3.50):

$$\begin{aligned} I(p_0, \hat{p}(y)) &= \mathbb{E}_0\left\{\ln\left(\hat{p}(y)\right)\right\} \\ &= \mathbb{E}_y\left\{\mathbb{E}_x\left\{\ln\left(p_n(y,\hat{\theta}_x^n)\right)\right\}\right\} \end{aligned} \tag{3.62}$$

The selection of $\hat{p}_n(y)$ has the advantage of using samples from x for estimation and those from y for validation [117, 123]. If a second-order Taylor series of $\ln\left(p_n(y,\hat{\theta}_x^n)\right)$

around $\hat{\theta}_y^n$ is applied, as in (3.54) [117, 121], it yields to:

$$\begin{aligned}\ln\left(p_n(y,\hat{\theta}_x^n)\right) &\approx \ln\left(p_n(y,\hat{\theta}_y^n)\right) + (\hat{\theta}_x^n - \hat{\theta}_y^n)^T \left[\left.\frac{\partial \ln\left(p_n(y,\theta)\right)}{\partial\theta}\right|_{\theta=\hat{\theta}_y^n}\right] \\ &+ \frac{1}{2}(\hat{\theta}_x^n - \hat{\theta}_y^n)^T \left[\left.\frac{\partial^2 \ln\left(p_n(y,\theta)\right)}{\partial\theta\partial\theta^T}\right|_{\theta=\hat{\theta}_y^n}\right](\hat{\theta}_x^n - \hat{\theta}_y^n) \\ &\approx \ln\left(p_n(y,\hat{\theta}_y^n)\right) - \frac{1}{2}(\hat{\theta}_x^n - \hat{\theta}_y^n)^T J_y (\hat{\theta}_x^n - \hat{\theta}_y^n)\end{aligned} \tag{3.63}$$

where J_y is the same defined in (3.57) associated with the data y.

As it is assumed that x and y have the same probability density function and are independent, it can be shown that:

$$\begin{aligned}\mathbb{E}_y\left\{\mathbb{E}_x\left\{(\hat{\theta}_x^n - \hat{\theta}_y^n)^T J_y (\hat{\theta}_x^n - \hat{\theta}_y^n)\right\}\right\} &= \mathbb{E}_y\left\{\mathbb{E}_x\left\{tr\left[\left(\left(\hat{\theta}_x^n - \theta\right) - \left(\hat{\theta}_y^n - \theta\right)\right)\left(\left(\hat{\theta}_x^n - \theta\right) - \left(\hat{\theta}_y^n - \theta\right)\right)^T\right]\right\}\right\} \\ &= tr\left[J_y\left(J_x^{-1} + J_y^{-1}\right)\right] \\ &= 2n\end{aligned} \tag{3.64}$$

Substituting the previous result (3.64) into (3.63) yields the asymptotic approximation:

$$I(p_0, \hat{p}(y)) = \mathbb{E}_y\left\{\ln\left(p_n(y,\hat{\theta}_y^n)\right)\right\} - n \tag{3.65}$$

And minimizing (3.65) respect to n is the same that maximizing the AIC:

$$AIC \quad = -2\ln p_n(y,\hat{\theta}^n) + 2n \tag{3.66}$$

The reason for multiplying by -2 the equation (3.65) is given by the convenience of replacing $-2\ln p_n(y,\hat{\theta}^n)$ term with a logarithmic scale of the residual variance [121].

$$-2\ln p_n(y,\hat{\theta}^n) = N\ln\hat{\sigma}_n^2 + costant \tag{3.67}$$

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 [117].

3.3.2 Bayesian information criterion

BIC is a second possible information criterion used to select model order similar to the case of the Akaike criterion, one of the variants for obtaining BIC is related to maximizing the KL in (3.50)

$$I(p_0, \hat{p}(y)) = \mathbb{E}_0 \left\{ \ln \left(\hat{p}(y) \right) \right\} \tag{3.68}$$

As the likelihood of the model is not available, as in the previous case then:

$$\hat{p}(y) \neq p_n(y, \theta^n) \tag{3.69}$$

Thus, it is once again possible to assume a sample, denoted as x, to perform inference on the parameter θ^n by using the *pdf* $p(\theta^n)$ of $\hat{\theta}^n$.

It is assumed that $p(\theta^n)$ is approximately constant in the vicinity of the estimate, i.e.

$$p(\theta^n) \text{ is constant around } \hat{\theta^n} \tag{3.70}$$

and

$$p(\theta^n) \text{ is independent of } N \tag{3.71}$$

The fact that the fictitious sample x is independent of N implies that the relationship between the sizes of the validation sample and x to grow without bounds as N increases. This leads to an order selection rule with a penalty term greater than in the case of AIC, thereby reducing the risk of overfitting [121].

The surrogate likelihood is given by:

$$\hat{p}(y) = \mathbb{E}_\theta \left\{ p_n(y, \theta^n) \right\} = \int p_n(y, \theta^n) p(\theta^n) d\theta \tag{3.72}$$

The asymptotic approximation of (3.72) is obtained by using the approximation (3.54) - (3.55), assuming regularity conditions and $N \to \infty$ (the *pdf* converges to a Gaussian *pdf*).

$$p_n(y, \theta^n) \approx p(y, \hat{\theta}^n) e^{\left\{ -\frac{1}{2} (\hat{\theta}^n - \theta^n)^T \hat{J} (\hat{\theta}^n - \theta^n) \right\}} \tag{3.73}$$

Then by substituting in (3.72):

$$\begin{aligned}\hat{p}(y) &\approx p(y,\hat{\theta}^n)p(\hat{\theta}^n)\int e^{\left\{-\frac{1}{2}(\hat{\theta}^n-\theta^n)^T\hat{J}(\hat{\theta}^n-\theta^n)\right\}} \\ &\approx \frac{p(y,\hat{\theta}^n)p(\hat{\theta}^n)(2\pi)^{\frac{n}{2}}}{|\hat{J}|^{\frac{1}{2}}}\int \frac{1}{(2\pi)^{\frac{n}{2}}|\hat{J}|^{-\frac{1}{2}}}e^{\left\{-\frac{1}{2}(\hat{\theta}^n-\theta^n)^T\hat{J}(\hat{\theta}^n-\theta^n)\right\}}d\theta \\ &\approx \frac{p(y,\hat{\theta}^n)p(\hat{\theta}^n)(2\pi)^{\frac{n}{2}}}{|\hat{J}|^{\frac{1}{2}}}\end{aligned} \tag{3.74}$$

For the above approximation, (3.70) has been taken into account along with the fact that $p_n(y,\theta^n)$ is asymptotically larger at $\theta^n=\hat{\theta}^n$, i.e. when the estimation $\hat{\theta}^n$ converges towards the true value of the parameter θ the probability distribution $p_n(y,\theta^n)$ becomes more spread out around the measured value of y.

Then from (3.50) and (3.74):

$$\hat{I} = \ln\left(p_n(y,\hat{\theta}^n)\right) + \ln\left(p(\hat{\theta}^n)\right) + \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\hat{J}| \tag{3.75}$$

To eliminate the dependence on the prior *pdf* of θ, the fact that J grows without bounds as N increases is utilized:

$$\begin{aligned}\ln|\hat{J}| &= \ln\left|N\frac{1}{N}\hat{J}\right| \\ &= n\ln|N| + \ln\left|\frac{1}{N}\hat{J}\right| \\ &= n\ln|N| + \mathcal{O}(1)\end{aligned} \tag{3.76}$$

In (3.76), the fact that $|NJ| = N^n|J|$ is used, where N is a scalar and J is an $n\times n$ matrix. $\mathcal{O}(1)$ is a term that tends to a constant as $N\to\infty$.

Taking into account that $p_n(\theta^n)$ is independent of N:

$$\hat{I} = \ln\left(p_n(y,\hat{\theta}^n)\right) - \frac{n}{2}\ln N \tag{3.77}$$

Finally, BIC selects the order that maximizes (3.77)

$$BIC = -2\ln p_n(y,\hat{\theta}^n) + n\ln N \tag{3.78}$$

Equation (3.78) has a structure similar to the AIC. However, it differs in the 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 [121].

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 (3.78) is then selected.

3.3.3 Young information criterion

YIC is a model order selection criteria that, as a difference compared to AIC and BIC, uses the covariance matrix estimate associated with the estimated parameters [124, 43, 125]. In fact, YIC involves the coefficient of determination and the variance-error norm to define the model that best fits the data and best estimates the parameters [124].

The coefficient of determination R_T^2 is a metric that assesses the accuracy of the identified model to represent the actual model [126, 44], by considering the simulation error. The values of R_T^2 range between 0 and 1, indicating the relation between the identified model and the real-world observations, and is defined by [46]:

$$R_T^2 = 1 - \frac{\sigma_{\hat{\varepsilon}}^2}{\sigma_y^2} \tag{3.79}$$

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

$$\sigma_{\hat{\varepsilon}}^2 = \frac{1}{N}\sum_{k=1}^{N}(\hat{\varepsilon}(t_k) - \bar{\varepsilon}(t_k))^2 \tag{3.80}$$

$$\sigma_y^2 = \frac{1}{N}\sum_{k=1}^{N}(y(t_k) - \bar{y}(t_k))^2 \tag{3.81}$$

$$\hat{\varepsilon}(t_k) = y(t_k) - \hat{y}(t_k, \hat{\theta}^n) \tag{3.82}$$

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, and $\hat{\theta}^\eta$ is the estimated parameter vector given by the SRIVC algorithm.

The YIC is then given by

$$YIC = \ln\left(\frac{\sigma_{\hat{\varepsilon}}^2}{\sigma_y^2}\right) + \ln\left(\frac{1}{n_\theta}\sum_{i=1}^{n_\theta}\frac{\hat{\pi}_{ii}}{\hat{\theta}_i^2}\right) \tag{3.83}$$

where n_θ is the number of estimated parameters, $\hat{\pi}_{ii}$ is the i^{th} diagonal element of the covariance matrix of the estimated parameter vector $\hat{\theta}^n$ [43].

3.3.4 The coherence method

Spectral coherence is a powerful signal processing tool, that has been applied to analyze the relationship between two random signals or processes [127, 128]. It has been extensively used in different fields, including communication systems, acoustics, and biomedical engineering [129]. 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

$$C_{(x,y)}(f) = \frac{\left|P_{(x,y)}(f)\right|^2}{P_{(x,x)}(f)P_{(y,y)}(f)} \tag{3.84}$$

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 [130]. 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 determining the linearity assumption of the system. The input-output data can then be used to estimate the parameters of the selected linear model.

3.4 Chapter summary

In this chapter, the theoretical foundations of maximum likelihood parametric estimation methods for discrete-time models from sampled data were presented. Similarly, simple refined instrumental variables for continuous-time OE model methods were

introduced, considering both sampled data and the frequency response of a system. The advantages and disadvantages of each method were discussed, laying the groundwork for model order selection for subglottal representation in this thesis.

A key focus of this work is to obtain low-order models and information criteria, specifically, AIC and BIC have been proposed for this purpose. These criteria, based on Kullback-Leibler divergence optimization, were discussed since they have been widely used for model selection to prevent overfitting.

Additionally, YIC was introduced, which leverages the advantage of using the covariance matrix estimate associated with the estimated parameters. These methods collectively strive to select the best-fitting model, aligning with the data used in the estimation process.

Finally, the coherence method was introduced, offering a measure of linear relationship between two signals. This method sets the stage for justifying the use of LTI models in representing the subglottal model in the subsequent chapter. In summary, this chapter has provided an overview of the estimation methods and model selection criteria to be used in the subsequent analyses in the thesis.

Chapter 4

Continuous-Time Model Identification of the Subglottal System

This chapter presents the key outcomes of estimating models for the subglottal system, specifically the body cover model, impedance-based model, and linear-time invariant models. The estimations are based on real input-output data obtained in the laboratory. However, in one case we use the frequency response of a section of the IB model for estimation. A detailed overview of the dataset pre-processing is provided. Additionally, the chapter delves into the problem of model order selection for linear time-invariant models by using information criteria. Moreover, the models are tested for inverse filtering, which is crucial for ambulatory assessment of the subglottal system.

4.1 Body Cover Model parameter estimation

For the body cover model (BCM) estimation, we opted to use simulated data for the input-output values. This approach allows us to avoid considering the effects of the trachea and focus on a simpler model that accounts for air pressure from the lungs and vocal fold collisions. In this work, the BCM is assumed to be a nonlinear system by incorporating nonlinearity coefficients into the springs models.

4.1.1 Model simulation

To simulate the model and address the state and parameter estimation challenges, we derive a discrete-time approximation of the model dynamics. This involves employing a truncated Taylor series expansion [131] of the Port-Hamiltonian formulation (2.21) to discretize the model, with a uniform sampling period denoted as Δ. The application of the truncated Taylor series expansion was previously demonstrated in [132], albeit for a simpler model than the one explored in this thesis. The order of the Taylor series expansion is selected individually for each state variable, extending up to the derivative at which the input explicitly appears. This process yields a discrete-time state-space (4.1) model in which each state is influenced by the input signal over a single time step.

$$
\begin{aligned}
q_u(k\Delta+\Delta) =& q_u(k\Delta) + \Delta\dot{q_u}(k\Delta) + \frac{\Delta^2}{2!}\ddot{q_u}(k\Delta) + +\frac{\Delta^3}{3!}\dddot{q_u}(k\Delta) + \frac{\Delta^4}{4!}\ddddot{q_u}(k\Delta)\\
q_{col_u}(k\Delta+\Delta) =& q_{col_u}(k\Delta) + \Delta\dot{q}_{col_u}(k\Delta) + \frac{\Delta^2}{2!}\frac{s_u}{m_u}\dot{p_u}(k\Delta) + +\frac{\Delta^3}{3!}\frac{s_u}{m_u}\ddot{p}_u(k\Delta) +\\
& \frac{\Delta^4}{4!}\frac{s_u}{m_u}\dddot{p}_u(k\Delta)\\
p_u(k\Delta+\Delta) =& p_u(k\Delta) + \Delta\dot{p}_u(k\Delta) + \frac{\Delta^2}{2!}\ddot{p}_u(k\Delta) + \frac{\Delta^3}{3!}\dddot{p_u}(k\Delta)\\
q_l(k\Delta+\Delta) =& q_l(k\Delta) + \Delta\dot{q}_l(k\Delta) + \frac{\Delta^2}{2!}\frac{\dot{p}_l(k\Delta)}{m_l}\\
q_{col_l}(k\Delta+\Delta) =& q_{col_l}(k\Delta) + \Delta\dot{q}_{col_l}(k\Delta) + \frac{\Delta^2}{2!}\frac{s_l}{m_l}\dot{p}_l(k\Delta)\\
p_l(k\Delta+\Delta) =& p_l(k\Delta) + \Delta\dot{p}_l(k\Delta)\\
q_b(k\Delta+\Delta) =& q_b(k\Delta) + \Delta\dot{q}_b(k\Delta) + \frac{\Delta^2}{2!}\frac{1}{m_b}\dot{p}_b(k\Delta) + \frac{\Delta^3}{3!}\frac{1}{m_b}\ddot{p}_b(k\Delta)\\
p_b(k\Delta+\Delta) =& p_b(k\Delta) + \Delta\dot{p}_b(k\Delta) + \frac{\Delta^2}{2!}\ddot{p}_b\\
y(k) =& G^T x(k) + v(k)
\end{aligned}
\tag{4.1}
$$

where the time derivatives on the right-hand side are replaced using equations (2.1)-(2.11).

Note that the approximate sampled-data model (4.1) is more accurate than applying Euler discretization [133, 131].

We further assume the noise measurement $v(k)$ in (4.1) is an i.i.d. sequence, with variance σ_v^2.

In this work, the notations $y(k)$ and y_k, $v(k)$ and v_k are used indistinctly to represent the same variable. Both notations denote the value of the variable y or v at the discrete time step k. This interchangeable use is consistent throughout the document.

An important issue for model simulation is the choice of the input function. In this case, the model input is a result of the interaction between sub-glottal pressure, vocal fold physiology and dynamics, and other effects such as wave propagation through the trachea during phonation. In this work, the effect of the trachea and other structures are not considered, and a simple approach for the interaction between fluid (air pressure from the lungs) and vocal folds is considered. Supra-glottal pressure, related to F_{in_u} in the above formulation is assumed to be 0 (i.e. atmospheric pressure). If a model of the trachea is included, this assumption does not hold since a wave reflection model is needed. On the other hand, sub-glottal pressure is modeled by a two-regime approximation of the air-vocal fold interaction, following the work of [134]. When the vocal folds are closed, sub-glottal pressure is not affected. When the folds are open, a Bernoulli-based approximation of their interaction with the air pressure is considered, i.e.,

$$\begin{aligned} u_k &= P_{l,k} L_g T_l \\ P_{l,k} &= P_{s,k} \left[1 - \Theta(a_{min,k}) \left(\frac{a_{min,k}}{a_{l,k}} \right)^2 \right] \Theta(a_{l,k}) \end{aligned} \tag{4.2}$$

where $L_g = 1[cm]$ is the glottis length, $P_{s,k} = 8[kPa]$ is the sub-glottal pressure, $T_l = 0.15[cm]$ is the lower vertical thickness of the vocal fold model, $a_{min,k} = \min(a_{u,k}, a_{l,k})$, $a_{u,k} = 2q_{u,k}L_g$ and $a_{l,k} = 2q_{l,k}L_g$ are the upper and lower glottal areas, respectively, $u_k = F_{in_l}$ and

$$\Theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{4.3}$$

An appropriate time step Δ must be used in order to accurately simulate the model. Model parameters, as well as initial conditions have been taken from [1], with $q_{u0} = 0.0179[cm]$, $q_{l0} = 0.018[cm]$, $q_{b0} = 0.3[cm]$ and $x_0 = \begin{bmatrix} q_{u0} & 0 & 0 & q_{l0} & 0 & 0 & q_{b0} & 0 \end{bmatrix}^T$. The parameter vector is defined as follows:

$$\theta = [k_u, k_l, k_b, k_c, \eta, \eta_{col}, b_u, b_l, b_b, b_{col_u}, b_{col_l}, m_u, m_l, m_b]^T$$

and $\eta_{col_u} = \eta_{col_l} = \eta_{col} = 500$, $\eta_u = \eta_l = \eta = 100$, $k_{col_u} = 3k_u$ and $k_{col_l} = 3k_l$.

In the fallowing, two estimation problems are presented: state estimation from a noisy output measurement sequence, and parameter estimation from both true state and estimated state variables.

4.1.2 State Estimation

In order to estimate the state from measurements of the system output, we use a Kalman Filter [135]. This algorithm is formulated for the state space representation of a linear system, subject to both state and measurement noise, mutually independent, Gaussian distributed and having known variance and zero mean. The key aspect to the implementation corresponds to the linearization of the model around the current posterior estimate of the state, resulting in an Extended Kalman Filter (EKF). The recursive equations defining the estimation process are given in Algorithm 2: where

Algorithm 2 Extended Kalman Filter Algorithm

procedure EKF$(\theta, \hat{x}_0, \Sigma^-_{x,0}, G, \sigma_v, y, u_k)$
$K_k \leftarrow \Sigma^-_{x,k} G(\sigma_v^2 + G^T \Sigma^-_{x,k} G)^{-1}$
$\hat{x}^+_k \leftarrow \hat{x}^-_k + K_k(y_k - G^T \hat{x}^-_k)$
$\Sigma^+_{x,k} \leftarrow (I - K_k G^T)\Sigma^-_{x,k}$
$A_k \leftarrow I + \Delta \nabla_x f(\theta, \hat{x}^+_k)$
$\tilde{u}_k \leftarrow \Delta \left[f(\theta, \hat{x}^+_k) - \nabla_x f(\theta, \hat{x}^+_k)\hat{x}^+_k + G u_k \right]$
$\hat{x}^-_{k+1} \leftarrow A_k \hat{x}^+_k + \tilde{u}_k$
$\Sigma^-_{x,k+1} \leftarrow A_k \Sigma^+_{x,k} A_k^T$
end procedure

$\tilde{u}_k$ is a deterministic term resulting from the combination of the linearization and the input of the system, A_k corresponds to the equivalent state transition matrix of the system, $\nabla_x f$ is the Jacobian matrix resulting from partial differentiation of f with respect to the state x_k, $\hat{x}_k$ is the state estimate at time k, $\Sigma_{x,k}$ is the covariance matrix associated to the estimate and the superscripts + and - denote posterior and prior estimation. Note that this formulation assumes the parameter vector $\theta_k = \theta$ to be constant and known.

The EKF requires an initial estimate of the state $\hat{x}$, and its associated covariance matrix $\Sigma^-_{x,0}$. The initial estimate is initialized to the equilibrium position x_0 and a diagonal covariance matrix with elements equal to 10^{-3}.

4.1.3 Parameter Estimation

It is usually stated that "all models are wrong, but someones are usefu" [136]. As a consequence, a key goal in the study of the subglottal system is to fit a (general) vocal fold model to each (particular) patient. The information available for the parameter estimation problem are the input and output signals (in this case from a simulation of the model). The EKF estimation process can be also exploited for parameter estimation, considering the following augmented system:

$$\begin{aligned} \theta_{k+1} &= \theta_k \\ \bar{y}_k &= f(\theta, \hat{x}_k^+) + Gu_k + e_k, \end{aligned} \tag{4.4}$$

where $\hat{x}_k^+$ corresponds to the posterior state estimate at time k, and e_k models the propagation of state estimation error. As the model is non-linear, the derivation of an explicit error propagation formula is out of the scope of this work. Hence, e_k will be modeled as a white, Gaussian, independent and identically distributed process, whose covariance matrix Σ_e is unknown and treated as an tuning parameter.

In this framework, and following a similar procedure to that of the formulation of the EKF for the state estimation problem, the output function is linearized on each iteration with respect to the parameter vector and around the prior estimate at time k: $\hat{\theta}_k^-$. Note that $\hat{x}_k^+$ is constant in this context. The recursive equations defining the estimation process are presented in Algorithm 3:

Algorithm 3 Recursive Estimation Algorithm

procedure RE($\hat{\theta}_0^-, \Sigma_{\theta,0}^-, G, \Sigma_e, u_k, y_k$)
$C_k \leftarrow \nabla_\theta f(\hat{\theta}_k^-, \hat{x}_k^+)$
$\bar{u}_k \leftarrow f(\hat{\theta}_k^-, \hat{x}_k^+) - \nabla_\theta f(\hat{\theta}_k^-, \hat{x}_k^+)\hat{\theta}_k^- + G^T u_k$
$K_k \leftarrow \Sigma_{\theta,k}^- C_k^T (\Sigma_e + C_k \Sigma_{\theta,k}^- C_k^T)^{-1}$
$\hat{\theta}_k^+ \leftarrow \hat{\theta}_k^- + K_k(\bar{y}_k - C_k\hat{\theta}_k^- - \bar{u}_k)$
$\Sigma_{\theta,k}^+ \leftarrow (I - K_k C_k)\Sigma_{\theta,k}^-$
$\hat{\theta}_{k+1}^- \leftarrow \hat{\theta}_k^+$
$\Sigma_{k+1}^- \leftarrow \Sigma_{\theta,k}^+$
end procedure

where $\bar{u}_k$ is a deterministic term resulting from the combination of the linearization and the original input of the system, C_k corresponds to the equivalent state-output matrix of the system, $\nabla_\theta f$ is the Jacobian matrix resulting from partial differentiation of f with respect to the parameter vector θ_k, $\hat{\theta}_k$ is the parameter estimate at time k,

$\Sigma_{\theta,k}$ is the associated covariance matrix and the superscripts + and - denote posterior and prior estimation. Note that the structure of the parameter model (4.4) does not change in the update step that maps posterior to prior estimates on the last two steps of Algorithm 3.

An alternative approach for parameter estimation problem is to perform maximum likelihood estimation, i.e., considering the likelihood function (3.5) given by:

$$L(\theta) = \ln(p(y_{0:N-1}|\theta)) \tag{4.5}$$

where $p(y_{0:N-1}|\theta)$ is the probability of the observed data $\{y_k\}_{k\in 0:N-1}$, given a parameter vector θ. The maximum likelihood estimate is then given by (3.26):

$$\hat{\theta}_{ML} = \arg\min_{\theta} L(\theta). \tag{4.6}$$

Note that the formulation of the discretized port-Hamiltonian vocal fold model (4.1) does not consider a process noise term, and the state transition is therefore deterministic. Hence, the probability distribution of the output is characterized by $p(y_k|y_{0:k},\theta) \sim \mathcal{N}(g(x_k,\theta),\sigma_v^2)$. If the initial output is independent of noise process v_k (i.e. the Markov's property holds), then we have the following:

$$\begin{aligned} L(\theta) &= \ln p(y_{0:N-1}|\theta) = \ln \prod_{k=0}^{N-1} p(y_k|y_{0:k-1},\theta) \\ &= \sum_{k=0}^{N-1} \ln p(y_k|y_{0:k-1},\theta) \\ &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{k=0}^{N-1}(y_k - g(x_k,\theta))^2 \end{aligned} \tag{4.7}$$

Note that, in order to obtain the maximum-likelihood estimate, it is sufficient to minimize the sum of the squared error term (3.26), or, equivalently, the mean squared error.

4.1.4 Numerical results

In this section, we implement the state and parameter estimation methodologies outlined in Sections 4.1.2 and 4.1.3. A discrete sampling interval of $\Delta = 10^{-5}[s]$ is utilized. The true parameter values are provided in Table 4.1, which also displays

the parameter estimates obtained through Maximum Likelihood (ML) and Extended Kalman Filter (EKF).

Table 4.1 True ($\bar{\theta}$) and Estimated Parameters using EKF ($\hat{\theta}_{EKF}$) and Maximum-Likelihood ($\hat{\theta}_{ML}$) estimation

θ	$\bar{\theta}$	$\hat{\theta}_{EKF}$	$\hat{\theta}_{ML}$
k_u	3500	3498.31	3430
k_l	5000	5000.13	4900
k_b	20000	20000.32	19600
k_c	500	471.74	490
η	100	93.49	98
η_{col}	500	496.91	490
b_u	4.7329	4.09	4.6382
b_l	5.6569	2.614	5.5437
b_b	12.6491	19.22	12.3961
b_{col_u}	4.7329	5.00	4.6382
b_{col_l}	5.6569	5.65	5.5437
m_u	0.01	0.00079	0.0098
m_l	0.01	0.0002	0.0098
m_b	0.05	0.073	0.0490

Additionally, Table 4.2 shows the mean-square error between the simulated state of the discrete-time port-Hamiltonian model (4.1) and the state estimation obtained by the EKF.

Figures 4.1 - 4.4 show the displacement of masses (vocal folds) in the model, with the parameters fitted using ML and EKF. The input force resulting from the interaction between the model dynamics and the sub-glottal pressure is also shown.

Figures 4.5 - 4.9 show the collisions of the upper and lower masses and the momentum of the masses.

Table 4.2 Mean-square error of simulated state $x(k\Delta+\Delta)$ and estimated state $\hat{x}(k\Delta+\Delta)$

$x(k\Delta+\Delta)$	**MSE**$(x(k\Delta+\Delta),\ \hat{x}(k\Delta+\Delta))$
q_u	0.001
q_{col_u}	9.916×10^{-06}
p_u	2.686×10^{-06}
q_l	0.004
q_{col_l}	7.572×10^{-06}
p_l	6.317×10^{-05}
q_b	0.0002
p_b	2.0202×10^{-06}

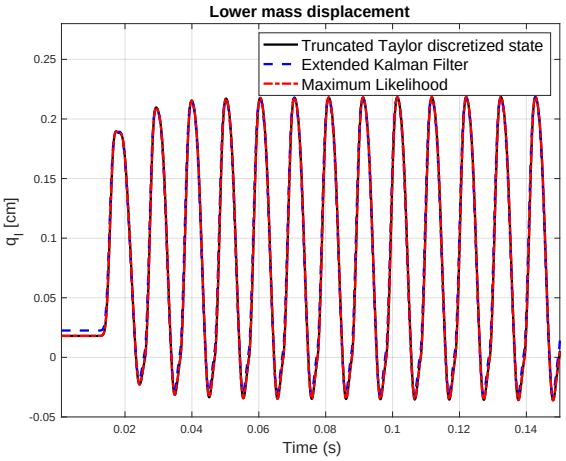


Figure 4.1 Lower mass displacement system

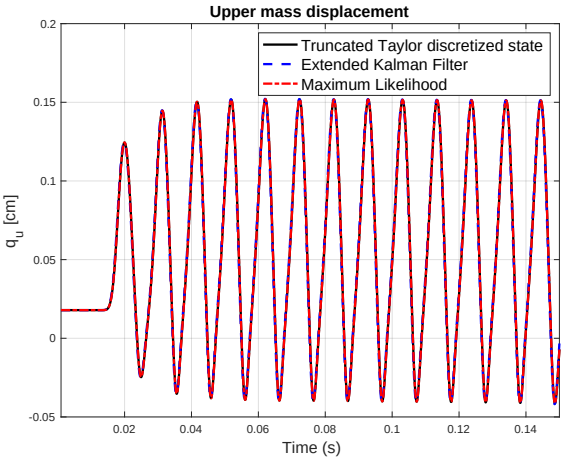


Figure 4.2 Upper mass displacement system

4.2 Impedance-base model parameter estimation

In this section, we present the estimates of a continuous-time model for the subglottal system. We present the results of estimating the parameters of the transfer function $F(p)$ given in (2.32) from its frequency response, and the parameters of $T(p)$ in (2.33) from time-domain sampled-data. In both cases, the SRIVC methodology is employed. Additionally, comments are provided on the criteria considered for selecting the order of $F(p)$.

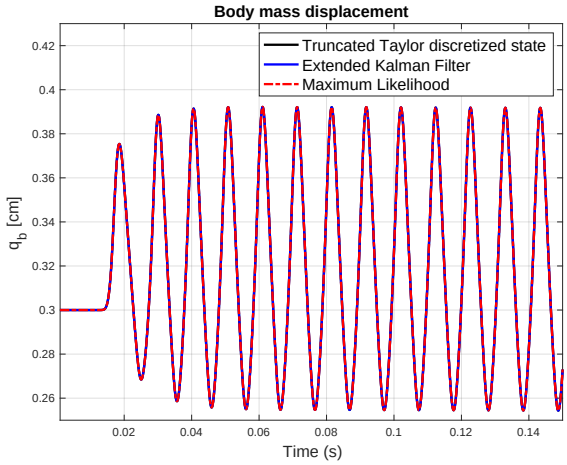


Figure 4.3 Body mass displacement system

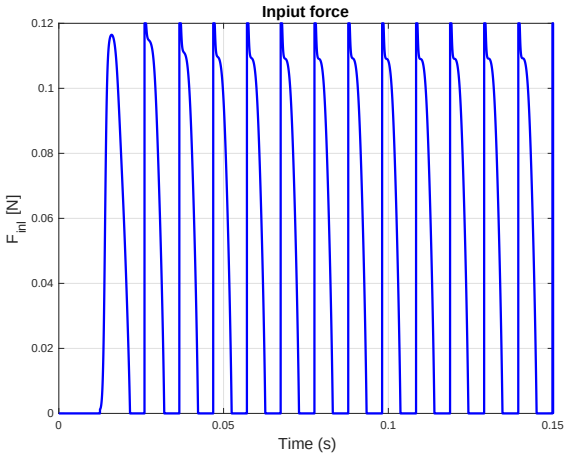


Figure 4.4 Input force

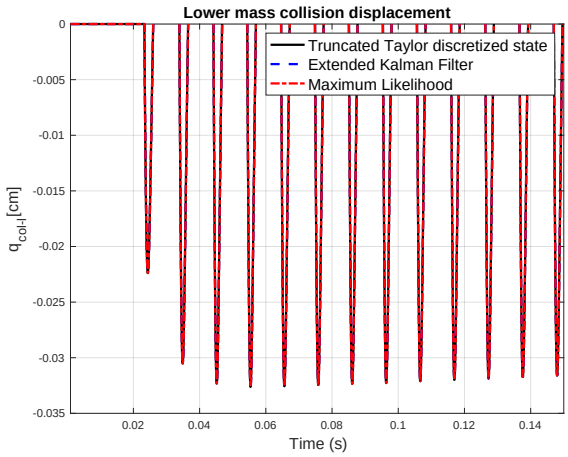


Figure 4.5 Lower mass collision displacement

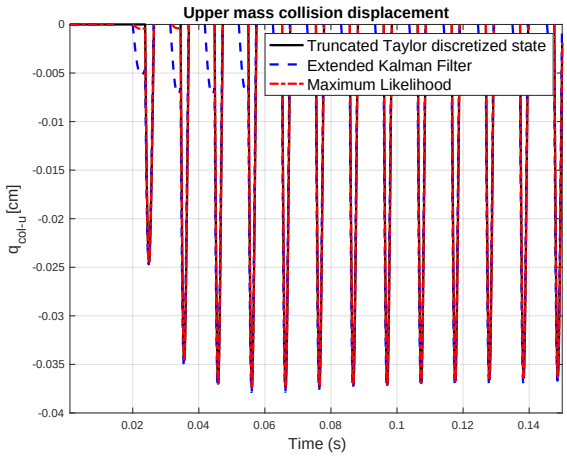


Figure 4.6 Upper mass collision displacement

4.2.1 Data pre-processing

We have the data of several patients for the estimation process, with each dataset comprising approximately 200,000 samples of input-output pairs, sampled at $20kHz$. As an example, we present below, in Figures 4.10 to 4.12, the power spectrum and the input-output data corresponding to the pronunciation of the vowel *a* five times. 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 [137].

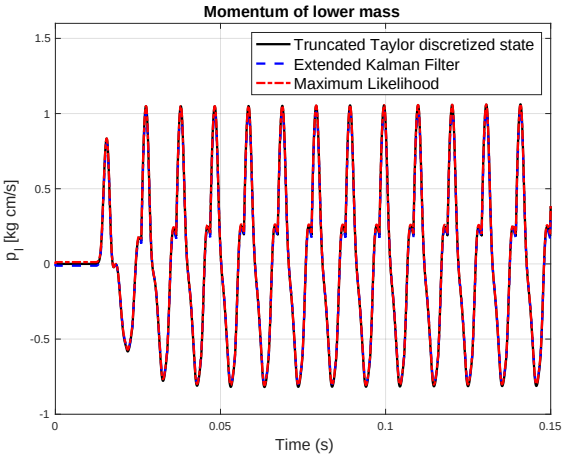


Figure 4.7 Lower mass collision displacement

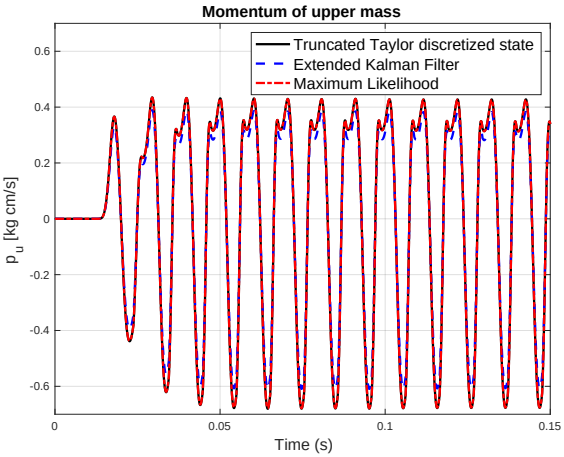


Figure 4.8 Upper mass collision displacement

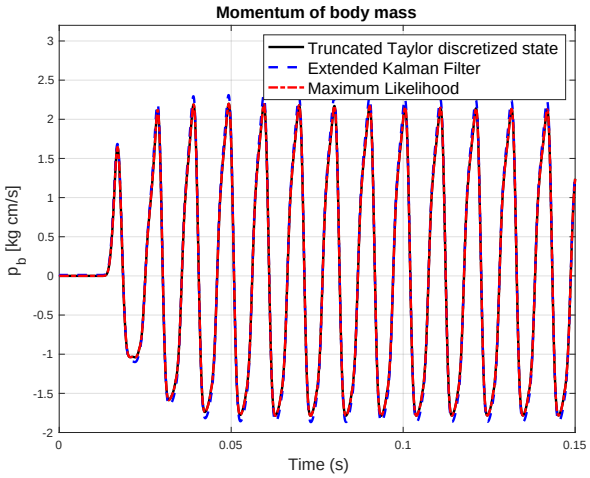


Figure 4.9 Body mass collision displacement

Moreover, downsampling plays a crucial role in alleviating the impact of aliasing, a phenomenon that distorts the frequency characteristics of signals during the sampling process [138]. The reduction in sampling rate not only may help to reduce the impact of aliasing artifacts but also enhance the accuracy of system identification. Hence, downsampling stands as an essential step in the estimation process, ensuring the precise identification of system parameters.

Considering the impact of high-frequency data on parametric estimation, we resample the ACC and GVV signals to a frequency of $F_{sub} = 8192Hz$. A similar down-sampling strategy has been applied, for example, in [23]. The power spectrum of the resampled data is shown in Figure 4.10. The analysis of the power spectrum is

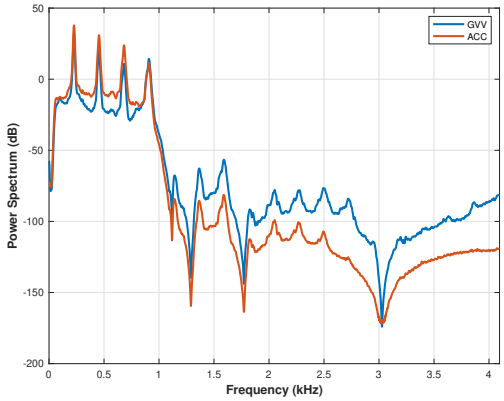


Figure 4.10 Input-output resampled dataset power spectrum

crucial for understanding signal characteristics such as dominant frequencies, energy distribution, and gain. Moreover, it may enhance experimental identification results.

The resampled GVV and ACC signals are shown in Figures 4.11 and 4.12, respectively. In these figures, data points in red correspond to the down-sampled signal, and blue lines represent the original signal sampled at high frequency (20kHz).

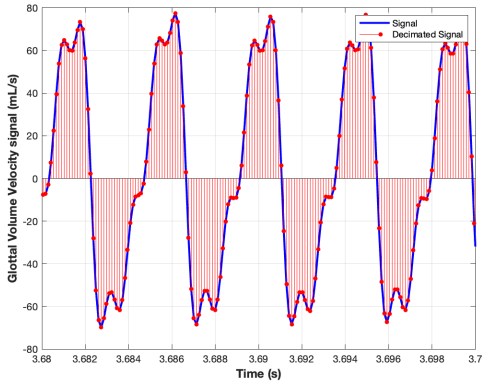


Figure 4.11 Glottal volume velocity (GVV) original and resampled signal

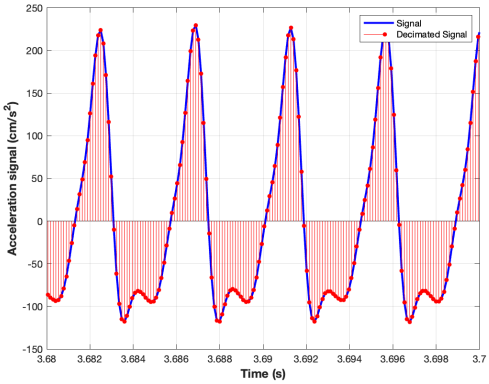


Figure 4.12 Acceleration (ACC) signal original and resampled signal

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 [5]. In this section, we test this assumption using the coherence method. The analysis shows that, in fact, the assumption of linearity between the measurements of ACC and GVV could be valid in the frequency bandwidth of interest.

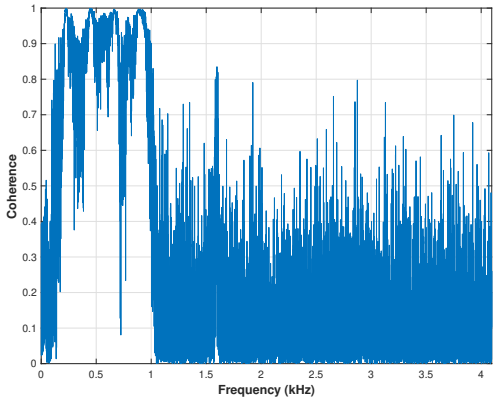


Figure 4.13 Coherence between ACC and GVV

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. Figure 4.13 shows the coherence between the two signals as a function of frequency. It can be noticed that, in the frequency range from approximately 220Hz 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 Figure 4.10, where the peaks appear within the same frequency bandwidth.

4.2.2 Continuous-time model from frequency-domain data

For the estimation of $F(p)$ defined in (2.32), the transfer functions $H_{sub1}(p)$ and $Z_{sub2}(p)$, are obtained from the frequency response and applying the procedure in Section 3.2.1. For each transfer function, a set of candidate models was estimated by varying the model order from 1 to 14 and the relative degree from 0 to 13. The analysis for 3 patients is presented below.

For $H_{sub1}(p)$, Figures 4.14 to 4.16 show statistics such as the mean square error (MSE), the coherence between the estimated model and the data used for estimation, and the model stability. After studying the obtained results, the model that exhibits the maximum coherence is used for estimation. In this case, the chosen structure has 6 poles and 3 zeros.

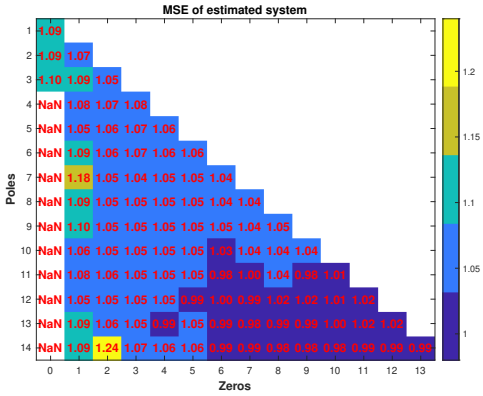


Figure 4.14 MSE between estimated $H_{sub1}(p)$ model and the data used for estimation

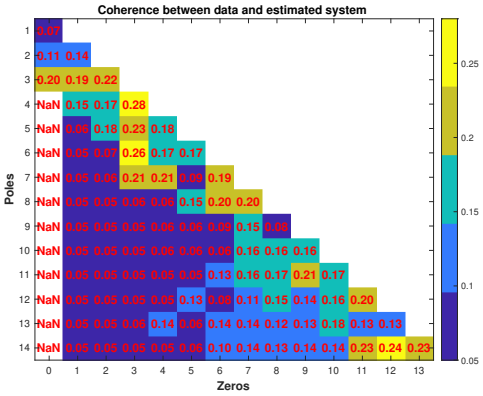


Figure 4.15 Coherence between simulated data from estimated $H_{sub1}(p)$ model and the data used for estimation

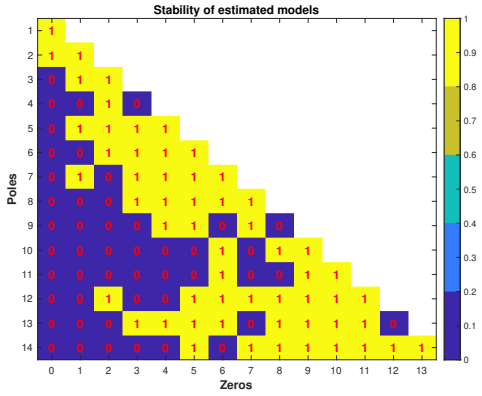


Figure 4.16 $H_{sub1}(p)$ model stability up to order 14

The results for the estimation of $Z_{sub2}(p)$ are shown in Figures 4.17 to 4.19. In the case of this transfer function, the model selected is of order 3 and with relative degree 1, i.e., it has 3 parameters in the denominator and 3 parameters in the numerator. It is interesting to note in Figure 4.19 that for $Z_{sub2}(p)$ very few stable models are obtained.

Once $H_{sub1}(p)$ and $Z_{sub2}(p)$ have been estimated, we obtain an estimate of the transfer function $F(p)$ defined in (2.32). Note that the resulting transfer function is of order 9 and relative degree 4. In the next subsection this transfer function is used to filter GVV to obtain the signal GVV^* (see Figure 2.7), which then is used for the estimation of $T(p)$.

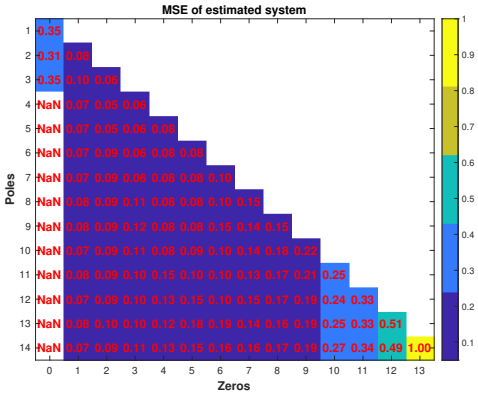


Figure 4.17 MSE between estimated $Z_{sub2}(p)$ model and the data used for estimation

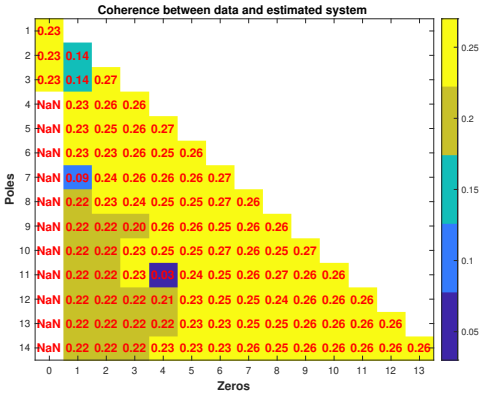


Figure 4.18 Coherence between simulated data from estimated $Z_{sub2}(p)$ model and the data used for estimation

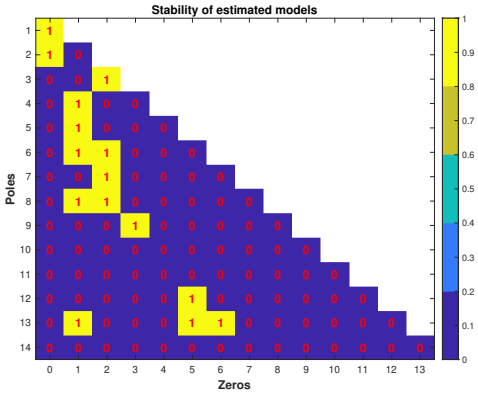


Figure 4.19 $Z_{sub2}(p)$ model stability up to order 14

4.2.3 Continuous-time model from sampled time-domain data

In this subsection, the results of the parameter estimation for the continuous-time transfer function $T(p)$ are defined in (2.33). These parameters can then be used to obtain the parameters Q_i from (2.34). However, the SRIVC method presented in Section 3.2 does not consider constraints in the parameter estimates and, thus, on the parameters Q_i. Similarly, the method cannot guarantee that the estimated model for $T(p)$ has two zeros at the origin as in the structure of transfer function (2.33).

In this section, the continuous-time model is estimated using an input-output dataset from 3 individuals, each containing more than 200000 samples—split into five phonations—sampled at 20 kHz. The phonation of vowel 'a' in the three scenarios under investigation is matched in this dataset. The data was downsampled at a frequency

of 8192 Hz to reduce the associated computing cost [139]. The resampled dataset contains 2000 time data points for estimate. There is a high percentage of good fit in the computed models. The 3^{rd} vocalization of the second subject is shown in Figures 4.20 and 4.21. However, since the estimated model has non-zero roots in the numerator polynomial, and the model in terms of practical physiological factors has two zeros at zero, further approximation is needed.

In fact, when applying SRIVC to any of the input-output datasets (GVV^* and ACC, respectively) corresponding to a single phonation of vowel *"a"*, a model is obtainde as the one shown in (4.8). This precise model belongs to the 3^{rd} vocalization of the second subject.

$$\hat{T}(p) = \frac{-3.638p^2 - 1586p - 6.509e06}{p^2 + 3734p + 8.694e06} \tag{4.8}$$

The model obtained provides a satisfactory fit percentage as shown in Figure 4.20. However, to interpret this model in terms of meaningful physiological parameters $Q_{1,2,3}$ a further approximation is required.

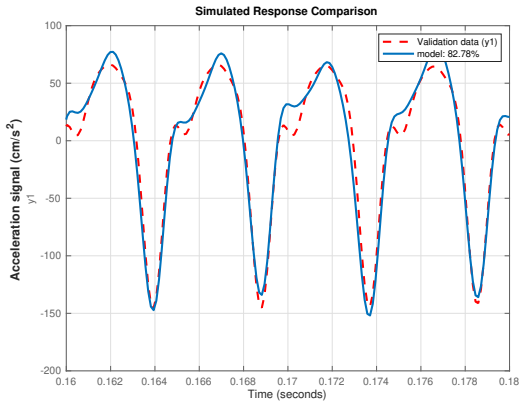


Figure 4.20 Comparison of $G(p) = F(p)T(p)$ model against validation data

To approximate the model (4.8) to the structure in (2.33), we obtain an approximate transfer function $\hat{T}_{approx}(p)$ solving the following optimization problem:

$$\arg\min_{b,a_1,a_2} \left\| \kappa(x) \left[\hat{T}(j\omega) - \hat{T}_{approx}(j\omega) \right] \right\|_2^2 \tag{4.9}$$

where $\kappa(x)$ is a frequency window function. This coefficient was computed using a tapered cosine window, with cosine fraction $r = 0.1$ [140].

$$\kappa(x) = \begin{cases} \frac{1}{2}\left\{1 + \cos\left[\frac{2\pi}{r}\left(x - \frac{r}{2}\right)\right]\right\} & 0 \leq x < \frac{r}{2} \\ 1 & \frac{r}{2} \leq x < 1 - \frac{r}{2} \\ \frac{1}{2}\left\{1 + \cos\left[\frac{2\pi}{r}\left(x - 1 + \frac{r}{2}\right)\right]\right\} & 1 - \frac{r}{2} \leq x \leq 1 \end{cases} \tag{4.10}$$

where x is an L-point linearly spaced vector and $0 \leq r \leq 1$.

The resulting approximate transfer function is given by (4.11), and the comparison of frequency responses is illustrated in Figure 4.21.

$$\hat{T}_{\text{approx}}(p) = \frac{-3.637p^2}{p^2 + 3406p + 5.581e06} \tag{4.11}$$

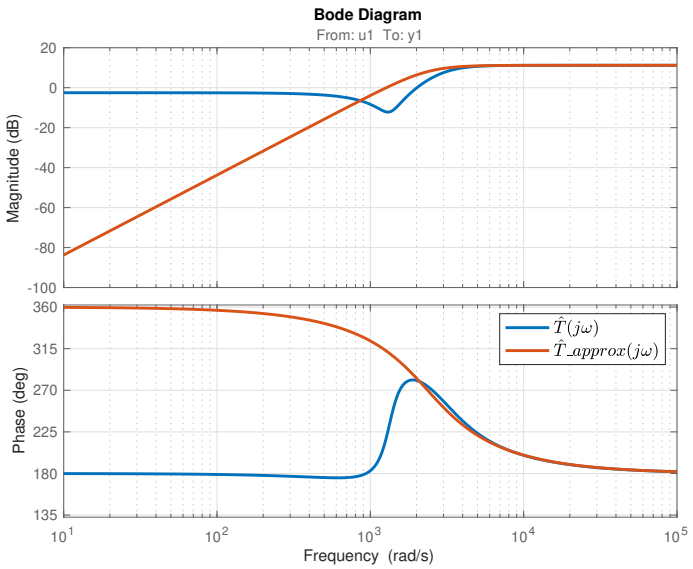


Figure 4.21 Frequency response comparison

The obtained models for the subglottal system show good fit and the results in Table 4.3 for three different subjects highlight that the method proposed in this paper allows to obtain specific parameters for each one of them. However, the obtained Q_i parameters in most cases have no direct physiological meaning in the IB model, and in the cases highlighted in red in Table 4.3 lead to unstable models. This suggests that deriving the Q_i parameters from the estimated IB model using the techniques applied in this paper may not be useful. The figures shown above, as well as the transfer function (4.11), were obtained by using data from the third phonation of subject 2.

Table 4.3 Estimates of the parameters Q_i for five vocalizations of three different subjects

Vocalization	1	2	3	4	5	Mean	Std
Subject 1							
Q_1	-0.36	-0.35	0.20	-0.60	-0.62	-0.35	0.33
Q_2	-1.01	-1.03	-1.33	-0.82	-0.84	-1.00	0.21
Q_3	-24.02	-25.10	-38.95	-15.66	-16.81	-24.12	9.30
Subject 2							
Q_1	-4.15	-1.83	-1.68	-1.72	-1.14	-2.10	1.17
Q_2	-0.62	-0.73	-0.93	-0.73	-1.19	-0.84	0.23
Q_3	6.37	-10.61	-12.98	-7.87	-32.85	-11.59	14.07
Subject 3							
Q_1	-0.76	-0.77	-1.01	-0.62	-0.87	-0.80	0.15
Q_2	-1.16	-1.16	-0.87	-1.24	-0.94	-1.07	0.16
Q_3	-25.39	-23.94	-13.58	-31.18	-18.79	-22.58	6.70

Figure 4.22 shows the frequency response for all the estimated models for the third subject, highlighting the similarity among all of them.

4.3 Linear Time-Invariant model parameter estimation

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 4.2.1.

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 [141]. Parameter estimation is performed for all possible combinations of candidate models, using orders 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.

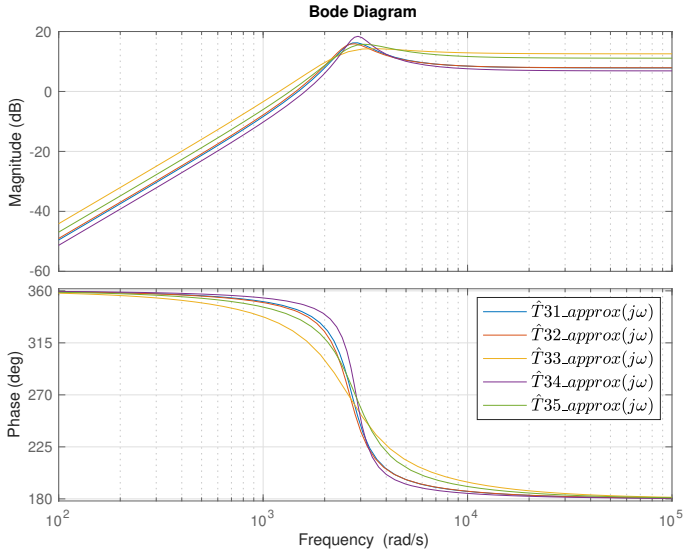


Figure 4.22 Frequency response comparison

The study is based on vocalization data gathered from three individuals under study, with each subject's dataset comprising five vocalizations. The Figures 4.23 to 4.30 presented below correspond to the results of the model estimation for the second vocalization of the first subject.

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. Figures 4.23, 4.24, 4.25 show the AIC, BIC, and YIC matrices, respectively, obtained when continuous-time system identification is performed using SRIVC. In these figures, it can be noticed that the three 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)

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y(t_k) - \hat{y}(t_k, \hat{\theta}^n))^2} \tag{4.12}$$

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 (Figure 4.26 shows the normalized RMSE).

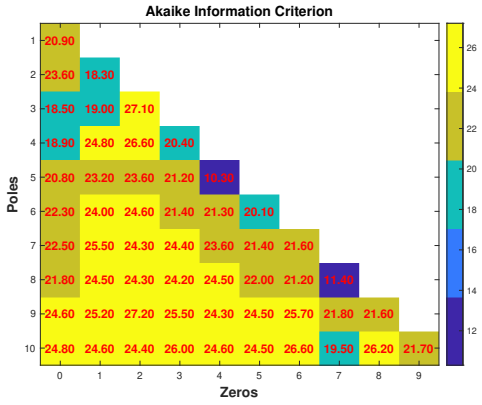


Figure 4.23 Heatmap of AIC coefficients

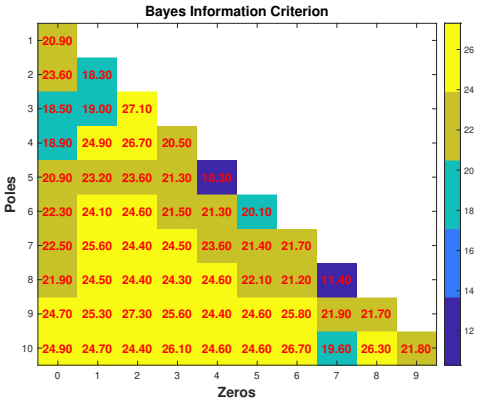


Figure 4.24 Heatmap of BIC coefficients

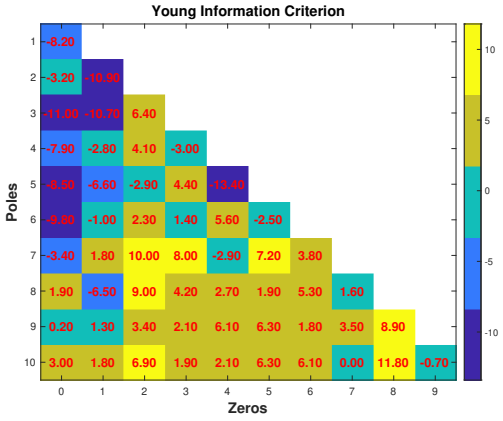


Figure 4.25 Heatmap of YIC coefficients

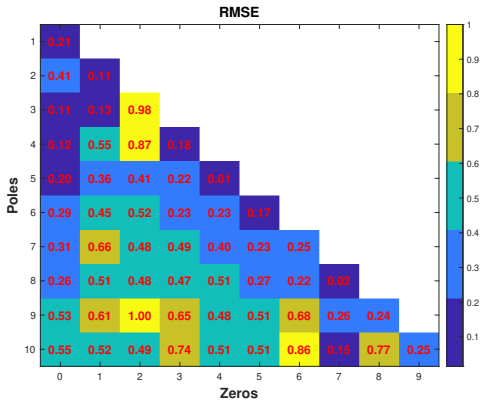


Figure 4.26 Root mean squared error (RMSE) of the candidates models

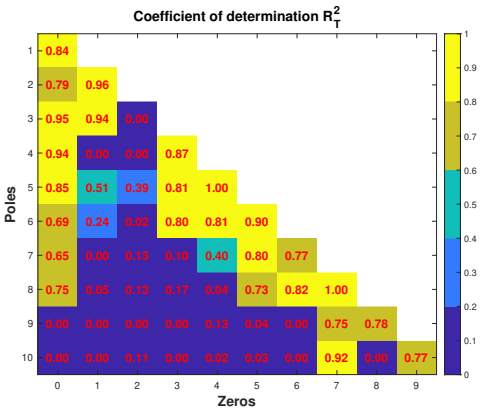


Figure 4.27 Heatmap of R_T^2 coefficients

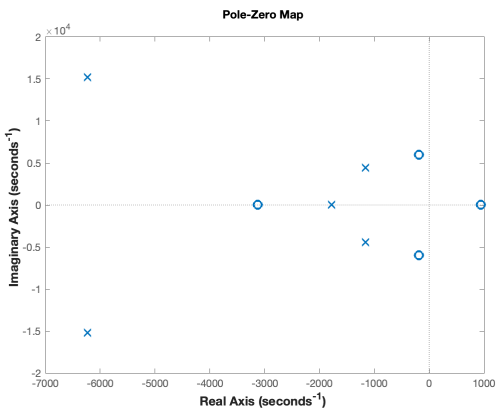


Figure 4.28 Poles and zeros map of the selected model estimated using SRIVC

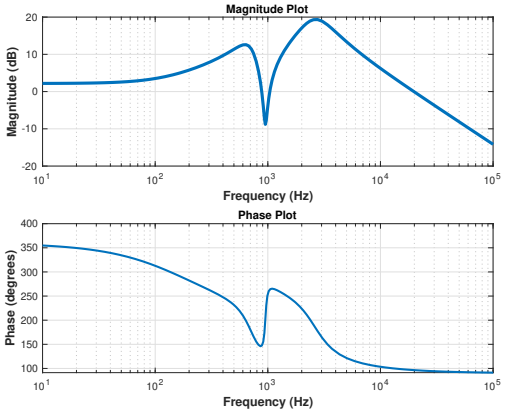


Figure 4.29 Bode diagram of the selected model estimated using SRIVC

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, Figure 4.27 shows the coefficient of determination obtained for the estimated models. It can be noticed that the highest value of the 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, Figures 4.28 and 4.29 show the pole/zero map and Bode magnitude plot of the selected model (with 5 poles and 4 zeros), and Figure 4.30 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:

$$FIT = \left(1 - \frac{\|y(t_k) - \hat{y}(t_k, \hat{\theta}^{\eta})\|}{\|y(t_k) - \bar{y}(t_k)\|}\right) 100\% \tag{4.13}$$

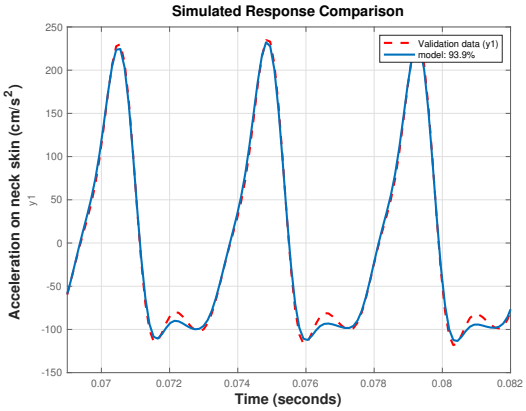


Figure 4.30 Comparison of the estimated model with SRIVC and resampled data simulation against validation data

Estimation results for models with more than 20 parameters are not reported. This decision is based on the observation that the achieved model fits for the specific application are already relatively high, effectively capturing characteristics and behaviors of the subglottal system. Attempting to improve the fit by increasing the model order would only compromise the model parsimony. Therefore, the focus is on maintaining the balance between model complexity and goodness of fit, prioritizing the parsimonious nature of the model over marginal improvements in fit.

4.3.1 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 measured acceleration. However, a key objective of obtaining a model for the subglottal system is to exploit 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 [5], the estimated glottal airflow is used to obtain indicators of the speech health of the patients.

In this section, inverse filtering is performed applying a Kalman smoother strategy (see algorithm (4)) as presented in [74]. A key difference 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.

Algorithm 4 Kalman Smoother Algorithm

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

The results of the Kalman smoother estimation of the glottal airflow (using the optimal model obtained by SRIVC) are shown in Figure 4.31. 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 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) [73].

Table 4.4 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 4.4 obtained from ACC are similar to previous results in the literature when applying Kalman filtering, such as, for example Table 2 in [142] and Table 4 in [74]. However, compared to those previous results, in this study we have used a lower-order model (5 poles and 4 zeros) estimated

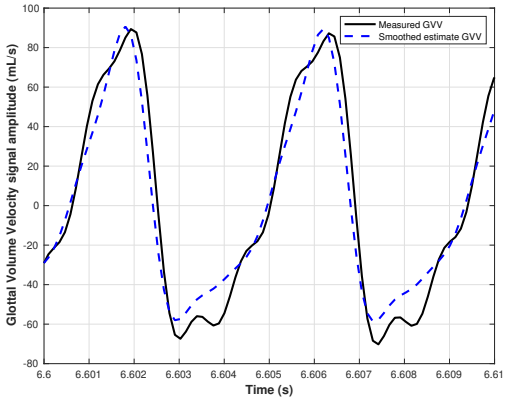


Figure 4.31 GVV estimation with Kalman Smoother and SRIVC estimated model

Table 4.4 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

directly from time-domain data. This implies a reduction in model order of 77% (5/22) against recent efforts [74] and 98.6% (5/350) against previous ones [142]. 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.

Table 4.5 shows the aerodynamic features for the 5 vocalizations of the 3 subjects under study. This also shows the fit obtained for each model, and the RMSE between GVV_OVV and GVVV_ACC.

Table 4.5 Aerodynamic Features computed from GVV_OVV for three subjects under study

Model \ Features		FIT (%)	RMSE (mL/s)	H1-H2 (dB)	HRF (dB)	MFDR (L/s^2)	ACFL (mL/s)	NAQ
Subject 1	Vocalization 1	88.85	10.30	8.85 ± 2.62	7.68 ± 3.42	138.62 ± 81.04	97.21 ± 55.10	0.16 ± 0.03
	Vocalization 2	93.9	10.62	8.43 ± 2.50	7.19 ± 3.26	142.10 ± 83.84	98.00 ± 56.24	0.15 ± 0.03
	Vocalization 3	90.78	10.81	9.42 ± 5.00	7.92 ± 5.77	138.43 ± 76.37	94.52 ± 50.58	0.15 ± 0.04
	Vocalization 4	74.45	11.24	8.21 ± 2.42	6.81 ± 3.11	142.07 ± 82.72	98.31 ± 55.99	0.15 ± 0.03
	Vocalization 5	87.82	8.29	8.38 ± 2.49	7.12 ± 3.25	154.59 ± 91.87	108.63 ± 63.06	0.16 ± 0.03
Subject 2	Vocalization 1	92.26	11.94	6.56 ± 22.56	4.77 ± 3.27	90.90 ± 54.88	62.39 ± 37.17	0.15 ± 0.04
	Vocalization 2	92.30	10.88	6.69 ± 2.59	4.94 ± 3.34	95.33 ± 56.64	65.44 ± 38.30	0.15 ± 0.04
	Vocalization 3	92.64	11.05	6.75 ± 2.56	5.04 ± 3.35	93.98 ± 55.81	64.91 ± 37.95	0.15 ± 0.04
	Vocalization 4	93.16	11.17	6.61 ± 2.75	4.97 ± 3.57	88.92 ± 56.27	61.51 ± 38.42	0.15 ± 0.05
	Vocalization 5	73.37	11.36	7.01 ± 2.70	4.89 ± 3.52	92.75 ± 54.64	64.99 ± 37.87	0.14 ± 0.04
Subject 3	Vocalization 1	83.61	8.03	5.99 ± 5.32	3.05 ± 5.84	80.19 ± 68.84	49.57 ± 41.38	0.13 ± 0.04
	Vocalization 2	84.81	7.82	5.66 ± 2.21	2.91 ± 2.93	78.90 ± 64.41	49.66 ± 41.46	0.13 ± 0.03
	Vocalization 3	81.68	7.46	6.46 ± 4.23	3.81 ± 4.70	75.40 ± 64.11	48.07 ± 39.76	0.14 ± 0.04
	Vocalization 4	90.19	7.23	8.19 ± 7.69	5.40 ± 7.81	76.35 ± 66.98	47.24 ± 40.18	0.15 ± 0.05
	Vocalization 5	83.78	8.21	6.27 ± 4.28	3.46 ± 4.79	75.39 ± 64.52	47.66 ± 39.29	0.13 ± 0.04

4.4 Chapter summary

The results of parameter estimates for different models of the subglottal system and for the vocal folds have been presented in this chapter. First, the Body Cover Model is the focus of attention. Bond graph analysis was used to first depict the vocal fold structure in this model, which is built within Hamiltonian ports. The model states and parameters are then computed using maximum likelihood and an extended Kalman filter. After that, a simplified form of the impedance-based model is presented, with an emphasis on utilizing SRIVC to estimate three of the five important parameters in the IB model. However, the obtained Q_i parameters do not provide a clear physiological interpretation.

In the chapter, the results of the LTI model estimation was also presented. These estimations were performed assuming a black-box approach, and the results achieved allowed for a reduction in model order of 77% (5/22) compared to recent results in [74] and 98.6% (5/350) compared to previous study in [142]. This shows a significant reduction in computational cost and model complexity, which is crucial for real-time monitoring applications.

Chapter 5

Conclusions

The current thesis has considered the modeling and characterization of vocal fold and subglottal systems through parameter estimation methodologies. The three case of studies presented in this work have provided distinctive conclusions, which are summarized below:

First, a study of the vocal folds body cover model was presented. The non-linear nature of mechanical elements suggested the derivation of a discrete port-Hamiltonian system, showcasing oscillatory behavior. Furthermore, system state estimation using an Extended Kalman filter and model parameter estimation through Maximum Likelihood underscored the robustness of the proposed model.

Secondly, continuous-time impedance-based model estimation of the subglottal system was studied to obtain subject-specific parameter estimates. With this goal, acoustic and mechanical models were separated. For the estimation of the acoustic model $F(p)$, the frequency response of subglottal section 1 and the frequency-dependent driving-point impedance of tract 2 was utilized. The estimation of both models was set for 105 candidate models selecting as the best model order and relative degree, the one with highest coherence with the data used for estimation. Then the mechanical model $T(p)$ was estimated with SRIVC, and a further approximation was applied to guarantee the model structure with two zeros at the origin. However, in some cases, the obtained Q_i parameters did not provide a clear physiological interpretation.

In the third study, continuous-time, linear time-invariant models were obtained for the subglottal system. The black box 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, significantly reducing computational cost and model complexity.

The obtained models, especially the LTI model, are validated not only using validation data but also by performing inverse filtering and estimating the glottal volume velocity from accelerometer data.

In synthesis, these works collectively advance our understanding of vocal and subglottal dynamics. While the first study emphasizes non-linear modeling and precise estimation, the second study navigates impedance-based modeling challenges. The third study, through identification techniques, highlights the potential for real-time assessment by using low-order LTI models. Together, they contribute to the sophisticated landscape of vocal system modeling, showing a new way for enhanced insights and applications in ambulatory vocal function assessment.

5.1 Future work

The work in this thesis opens up new areas of interest for the study of the subglottal system. The primary recommendations to be made are related to the estimation of parameters in the impedance-based model and the LTI model.

Further study is required for the estimation of the Q parameters of the impedance-based model by using system identification techniques in order to obtain parameter values with physiological interpretation. To achieve this, it is proposed to determine the specific structure, considering all the branches of the lungs, for $F(p)$ model in (2.31). Once the specific structure of $F(p)$ is obtained, the estimation of the $Q_{4,5}$ parameters could be carried out, which was not obtained in this research because the model was assumed to be a black box. This work could significantly improve the fit of the IB model, parameterized using experimental model identification techniques.

Additionally, as future work, the characterization of the uncertainty of the subglottal IB model parameter estimates by estimating confidence intervals and standard errors. Furthermore, it aims to determine the trade-off between model order and the quality of parameter estimates for the subglottal IB system, based on information criteria, when introducing modifications to the standard IB model.

It is also proposed to estimate the model $F(p)$, considering the fractional model suggested in [143], and to make the necessary adjustments to integrate it with the $T(s)$ model. This approach aims to achieve a comprehensive representation of the subglottal system, potentially enhancing the model's fit and thereby improving the accuracy of the estimated GVV flow.

In the case of the LTI model, it is parameterized using GVV and ACC, but it has the drawback that GVV is not obtained from a direct measurement; rather, it results from filtering OVV. For future work, it is proposed to estimate a Hammerstein model structure, performing the estimation directly with the measured OVV and ACC signals.

Bibliography

[1] M. Encina, J. Yuz, M. Zañartu, and G. Galindo, "Vocal fold modeling through the port-Hamiltonian systems approach," *2015 IEEE Conference on Control Applications (CCA)*, pp. 1558–1563, 2015.

[2] N. M. Kime, M. J. Ryan, and P. S. Wilson, "A bond graph approach to modeling the anuran vocal production system," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. 4133–4144, 2013.

[3] J. G. Fontanet, J. I. Yuz, J. A. Torres, M. A. Gordon, and H. Ramirez, "Port-Hamiltonian modeling of the vocal folds using bond-graph representation," *2021 IEEE International Conference on Automation/24th Congress of the Chilean Association of Automatic Control, ICA-ACCA 2021*, 2021.

[4] J. Chen, *Elements of Human Voice*. World Scientific, 2016.

[5] 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 Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1929–1939, 2013.

[6] M. Hirano, "Morphological structure of the vocal cord as a vibrator and its variations," *Folia Phoniatrica et Logopaedica*, vol. 26, no. 2, pp. 89–94, 1974.

[7] B. H. Story and I. R. Titze, "Voice simulation with a body-cover model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1249–1260, 1995.

[8] G. Fant, *Acoustic Theory of Speech Production*. Mouton & Co. N.V., 1971.

[9] K. N. Stevens, *Acoustic Phonetics*. The MIT Press, 2000.

[10] S. M. Lulich, A. Bachrach, and N. Malyska, "A role for the second subglottal resonance in lexical access," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, 2007.

[11] I. R. Titze, "Nonlinear source–filter coupling in phonation: Theory," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, 2008.

[12] I. Titze, T. Riede, and P. Popolo, "Nonlinear source–filter coupling in phonation: Vocal exercises," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, 2008.

[13] J. C. Ho, M. Zañartu, and G. R. Wodicka, "An anatomically based, time-domain acoustic model of the subglottal system for speech production," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1531–1547, 2011.

[14] M. Zañartu, L. Mongeau, and G. R. Wodicka, "Influence of acoustic loading on an effective single mass model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, 2007.

[15] Z. Zhang, J. Neubauer, and D. A. Berry, "The influence of subglottal acoustics on laboratory models of phonation," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, 2006.

[16] X. Chi and M. Sonderegger, "Subglottal coupling and its influence on vowel formants," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, 2007.

[17] S. M. Lulich, "Subglottal resonances and distinctive features," *Journal of Phonetics*, vol. 38, no. 1, pp. 20–32, 2010.

[18] K. Verdolini, C. Rosen, R. C. Branski, and others, *Classification manual for voice disorders-I.* Psychology Press, 2006.

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

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

[21] C. Morales, "¿De qué se enferman las trabajadoras chilenas?" *Ciencia y Trabajo*, vol. 23, pp. 20–24, 2007.

[22] R. E. Hillman, E. B. Holmberg, J. S. Perkell, M. Walsh, and C. Vaughan, "Objective assessment of vocal hyperfunction: An experimental framework and initial results," *Journal of Speech, Language, and Hearing Research*, vol. 32, no. 2, pp. 373–392, 1989.

[23] 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," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 8, pp. 2159–2169, 2017.

[24] M. Rothenberg, "New Inverse-Filtering Technique for Deriving the Glottal Air Flow Waveform during Voicing," *The Journal of the Acoustical Society of America*, vol. 48, no. 1A_Supplement, 1970.

[25] H. A. Cheyne, "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," *Proceedings of the 3rd IEEE-EMBS International Summer School and Symposium on Medical Devices and Biosensors, ISSS-MDBS 2006*, pp. 118–121, 2006.

[26] D. D. Mehta, J. H. Van Stan, M. Zañartu, M. Ghassemi, J. V. Guttag, V. M. Espinoza, J. P. Cortés, H. A. Cheyne, and R. E. Hillman, "Using ambulatory voice monitoring to investigate common voice disorders: Research update," *Frontiers in Bioengineering and Biotechnology*, vol. 3, no. OCT, pp. 1–14, 2015.

[27] J. P. Cortés, V. M. Espinoza, M. Ghassemi, D. D. Mehta, J. H. Van Stan, R. E. Hillman, J. V. Guttag, and M. Zañartu, "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.

[28] D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12 PART2, pp. 3090–3096, 2012.

[29] G. R. Wodicka, K. N. Stevens, H. L. Golub, E. G. Cravalho, and D. C. Shannon, "A Model of Acoustic Trnsmission in the Respirtor System," *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 9, 1989.

[30] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, "An acoustic model of the respiratory tract," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, 2001.

[31] V. P. Harper, H. Pasterkamp, H. Kiyokawa, and G. R. Wodicka, "Modeling and measurement of flow effects on tracheal sounds," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 1, 2003.

[32] M. Zañartu, D. D. Mehta, G. R. Wodicka, and R. E. Hillman, "Subglottal impedance-based inverse filtering for the ambulatory monitoring of vocal function: Initial results," *International Conference on Voice Physiology and Biomechanics*, p. 2012, 2012.

[33] V. Espinoza and M. Zañartu, "Estudio Dinámico de Parámetros de Filtrado Inverso para el Seguimiento Ambulatorio de la Función Vocal," in *IX Congreso Iberoamericano de Acústica*, no. December, 2014, pp. 1–10.

[34] J. P. Cortés, V. M. Espinoza, M. Ghassemi, D. D. Mehta, J. H. Van Stan, R. E. Hillman, J. V. Guttag, and M. Zañartu, "Ambulatory assessment of phonotraumatic vocal hyperfunction using glottal airflow measures estimated from neck-surface acceleration," *PloS one*, vol. 13, no. 12, p. e0209017, 2018.

[35] J. S. Perkell, R. E. Hillman, and E. B. Holmberg, "Group differences in measures of voice production and revised values of maximum airflow declination rate," *The Journal of the Acoustical Society of America*, vol. 96, no. 2, pp. 695–698, 1994.

[36] 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*, vol. 36, no. 5, pp. 623–650, 2011.

[37] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," *Computer Speech & Language*, vol. 28, no. 5, pp. 1117–1138, 2014.

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

[39] B. Henry and T. J. Royston, "A multiscale analytical model of bronchial airway acoustics," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, 2018.

[40] N. Hanna, J. Smith, and J. Wolfe, "How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, 2018.

[41] T. Söderström and P. Stoica, *System Identification*. Prentice Hall International, 1989.

[42] L. Ljung, *System identification: Theory for the user*, 2nd ed. Pearson, 12 1998.

[43] H. Garnier and L. Wang, *Identification of Continuous-time Models from Sampled Data*. Springer, 2008.

[44] H. Garnier, M. Gilson, P. C. Young, and E. Huselstein, "An optimal IV technique for identifying continuous-time transfer function model of multiple input systems," *Control Engineering Practice*, vol. 15, no. 4, pp. 471–486, 2007.

[45] H. Garnier and P. C. Young, "The advantages of directly identifying continuous-time transfer function models in practical applications," *International Journal of Control*, vol. 87, no. 7, pp. 1319–1338, 7 2014.

[46] H. Garnier, "Direct continuous-time approaches to system identification. Overview and benefits for practical applications," in *European Journal of Control*, vol. 24, 2015, pp. 50–62.

[47] L. A. Zadeh, "On the Identification Problem," *IRE Transactions on Circuit Theory*, vol. 3, no. 4, pp. 277–281, 1956.

[48] L. Ljung, *System identification-Theory for the user*. Prentice Hall New Jersey, 1987, vol. 25, no. 3.

[49] ——, "Perspectives on system identification," in *Annual Reviews in Control*, vol. 34, no. 1, 2010.

[50] S. D. Fassois and D. E. Rivera, "Applications Of System Identification," *IEEE Control Systems Magazine*, vol. 27, no. 5, pp. 24–26, 2007.

[51] M. Hussain, C. Fidge, E. Foo, and Z. Jadidi, "System Identification Methods for Industrial Control Systems," in *Secure and Trusted Cyber Physical Systems: Recent Approaches and Future Directions*, Pal Shantanuand Jadidi and Zahraand Foo Ernest, Eds. Cham: Springer International Publishing, 2022, pp. 25–50.

[52] S. Del Favero, G. Pillonetto, C. Cobelli, and G. De Nicolao, "A novel nonparametric approach for the identification of the glucose-insulin system in Type 1 diabetic patients," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 44, no. 1 PART 1, 2011.

[53] K. Turksoy, L. Quinn, E. Littlejohn, and A. Cinar, "Multivariable Adaptive Identification and Control for Artificial Pancreas Systems," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 883–891, 2014.

[54] T. Peyser, E. Dassau, M. Breton, and J. S. Skyler, "The artificial pancreas: Current status and future prospects in the management of diabetes," *Annals of the New York Academy of Sciences*, vol. 1311, no. 1, 2014.

[55] R. Mukkamala and R. J. Cohen, "A forward model-based validation of cardiovascular system identification," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 281, no. 6, pp. H2714–H2730, 12 2001. [Online]. Available: https://doi.org/10.1152/ajpheart.2001.281.6.H2714

[56] X. Xiao, Y. Li, and R. Mukkamala, "A model order selection criterion with applications to cardio-respiratory- renal systems," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 3, pp. 445–453, 3 2005.

[57] L. E. Alkurawy, M. S. Saleh, I. S. Fatah, and A. H. Saleh, "Modeling and identification of human heart system," *International Journal on Engineering Applications*, vol. 7, no. 4, 2019.

[58] M. P. Vlaar and A. C. Schouten, "System identification for human motion control," in *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, vol. 2015-July, 2015.

[59] D. Schmitthenner and A. E. Martin, "Comparing system identification techniques for identifying human-like walking controllers," *Royal Society Open Science*, vol. 8, no. 12, 2021.

[60] J. Lee, K. Zhang, and N. Hogan, "Identifying human postural dynamics and control from unperturbed balance," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, 2021.

[61] H. Liang, W. Xie, P. Wei, Y. Zhou, and Z. Zhang, "Identification of Human Body Dynamics from a Human-Structure System: An Experimental Study," *Experimental Techniques*, vol. 47, no. 2, pp. 449–470, 2023. [Online]. Available: https://doi.org/10.1007/s40799-022-00548-7

[62] M. Gevers and G. Bastin, "What Does System Identification Have to Offer?" *IFAC Proceedings Volumes*, vol. 15, no. 4, pp. 77–84, 1982. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667017629677

[63] H. Larsson, S. Hertegård, P. A. Lindestad, and B. Hammarberg, "Vocal fold vibrations: High-speed imaging, kymography, and acoustic analysis: A preliminary report," *Laryngoscope*, vol. 110, no. 12, 2000.

[64] D. A. Berry, D. W. Montequin, and N. Tayama, "High-speed digital imaging of the medial surface of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, 2001.

[65] Y. Zhang, E. Bieging, H. Tsui, and J. J. Jiang, "Efficient and Effective Extraction of Vocal Fold Vibratory Patterns from High-Speed Digital Imaging," *Journal of Voice*, vol. 24, no. 1, 2010.

[66] F. L. Lecluse, M. P. Brocaar, and J. Verschuure, "The electroglottography and its relation to glottal activity," *Folia Phoniatrica*, vol. 27, no. 3, 1975.

[67] D. G. Childers, D. M. Hicks, G. P. Moore, L. Eskenazi, and A. L. Lalwani, "Electroglottography and vocal fold physiology," *Journal of Speech and Hearing Research*, vol. 33, no. 2, 1990.

[68] P. Kitzing, "Clinical applications of electroglottography," *Journal of Voice*, vol. 4, no. 3, 1990.

[69] C. T. Herbst, W. T. S. Fitch, and J. G. Švec, "Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively," *The Journal of the Acoustical Society of America*, vol. 128, no. 5, 2010.

[70] C. T. Herbst, "Electroglottography – An Update," 2020.

[71] H. A. Cheyne, "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*, 2006, pp. 118–121.

[72] P. S. Popolo, J. G. Švec, and I. R. Titze, "Adaptation of a Pocket PC for use as a wearable voice dosimeter," *Journal of Speech, Language, and Hearing Research*, vol. 48, no. 4, pp. 780–791, 2005.

[73] 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," *Journal of Voice*, vol. 17, no. 3, 2003.

[74] A. Morales, J. I. Yuz, J. P. Cortés, J. G. Fontanet, and M. Zañartu, "Glottal Airflow Estimation using Neck Surface Acceleration and Low-Order Kalman Smoothing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 2055–2066, 2023.

[75] L. Mora, H. Ramirez, J. Yuz, L. G. Y, and M. Zañartu, "Energy based Fluid Structure Model of the Vocal Folds," *IMA Journal of Mathematical Control and Information.*, 2020. [Online]. Available: https://doi.org/10.1093/imamci/dnaa031

[76] A. Cinar and K. Turksoy, *Advances in Artificial Pancreas Systems: Adaptive and Multivariable Predictive Control*. Springer, 2018.

[77] C. K. Boughton and R. Hovorka, "Advances in artificial pancreas systems," *Science translational medicine*, vol. 11, no. 484, 2019.

[78] T. A. Smith, M. M. Seron, G. C. Goodwin, A. M. Medioli, B. R. King, C. E. Smart, M. Harris, D. N. Oneal, J. Rafferty, and P. Howley, "1058-P: The Use of Metabolic Digital Twins to Personalize Mealtime Insulin Dosing in Type 1 Diabetes Clinical Management," 2019.

[79] M. Abdi, A. Karimi, M. Navidbakhsh, G. Pirzad Jahromi, and K. Hassani, "A lumped parameter mathematical model to analyze the effects of tachycardia and bradycardia on the cardiovascular system," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 28, no. 3, pp. 346–357, 2015.

[80] P. Kochová, R. Cimrman, M. Štengl, B. Ošt'ádal, and Z. Tonar, "A mathematical model of the carp heart ventricle during the cardiac cycle," *Journal of theoretical biology*, vol. 373, pp. 12–25, 2015.

[81] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: Implementation issues and performance evaluation," *International Journal of Control*, vol. 76, no. 13, pp. 1337–1357, 2003.

[82] R. A. González, C. R. Rojas, S. Pan, and J. S. Welsh, "On the Relation Between Discrete and Continuous-Time Refined Instrumental Variable Methods," *IEEE Control Systems Letters*, vol. 7, pp. 2233–2238, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10143357/

[83] V. Pascu, H. Garnier, L. Ljung, and A. Janot, "Benchmark problems for continuous-time model identification: Design aspects, results and perspectives," *Automatica*, vol. 107, 2019.

[84] P. C. Young, *Recursive Estimation and Time-Series Analysis*, 2nd ed. London: Springer Science & Business Media, 2012, no. 2.

[85] S. Pan, R. A. González, J. S. Welsh, and C. R. Rojas, "Consistency analysis of the Simplified Refined Instrumental Variable method for Continuous-time systems," 2020.

[86] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell system technical journal*, vol. 51, no. 6, pp. 1233–1268, 1972.

[87] A. M. Vahabzadeh-Hagh, Z. Zhang, and D. K. Chhetri, "Hirano's cover–body model and its unique laryngeal postures revisited," *The Laryngoscope*, vol. 128, no. 6, pp. 1412–1418, 6 2018.

[88] J. Flanagan and L. Landgraf, "Self-oscillating source for vocal-tract synthesizers," *IEEE Transactions on Audio and Electroacoustics*, vol. 16, no. 1, pp. 57–64, 1968.

[89] I. R. Titze, "The human vocal cords: A mathematical model. Part I," *Phonetica*, vol. 28, no. 3-4, 1973.

[90] ——, "The human vocal cords: A mathematical model. Part II," *Phonetica*, vol. 29, no. 1, pp. 1–21, 1974.

[91] D. C. Karnopp, D. L. Margolis, and R. C. Rosenberg, *System dynamics: modeling, simulation, and control of mechatronic systems*. John Wiley \& Sons, 2012.

[92] A. J. der Schaft and A. J. Van Der Schaft, *L2-gain and passivity techniques in nonlinear control*. Springer, 2000.

[93] 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," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 9, 2020.

[94] P. Alku, C. Magi, S. Yrttiaho, T. Bäckström, and B. Story, "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," *The Journal of the Acoustical Society of America*, vol. 125, no. 5, 2009.

[95] G. P. Kafentzis, Y. Stylianou, and 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.

[96] V. M. Espinoza, J. P. Cortés, M. Zañartu, G. A. Alzamendi, J. I. Yuz, D. D. Mehta, and R. E. Hillman, "Modeling and Estimation of Neck-Surface Skin and Subglottal Systems with an ARMA-based model." in *The 51st Voice Foundation Symposium*, Philadeflphia, U.S.A., Jun. 2022.

[97] P. Mokhtari, H. Takemoto, and T. Kitamura, "Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches," *Speech Communication*, vol. 50, no. 3, pp. 179–190, 2008.

[98] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 222, no. 594-604, pp. 309–368, 1922.

[99] G. C. Goodwin and R. L. Payne, *Dynamic system identifications: Experiment design and data analysis*. Academic Press, 1977.

[100] K. J. Astrom, "Maximum likelihood and prediction error methods," *Automatica*, vol. 16, no. 5, pp. 551–574, 1980.

[101] B. T. Poljak and J. Tsypkin, "Robust identification," *Automatica*, vol. 16, no. 1, pp. 53–63, 1980. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0005109880900862

[102] D. Bertsimas and O. Nohadani, "Robust maximum likelihood estimation," *INFORMS Journal on Computing*, vol. 31, no. 3, pp. 445–458, 2019.

[103] J. I. Yuz and G. C. Goodwin, "Robust Identification of Continuous-time Systems from Sampled Data," in *Identification of Continuous-time Models from Sampled Data*, H. Garnier and L. Wang, Eds. London: Springer London, 2008, ch. 3, pp. 67–89.

[104] H. Unbehauen and G. P. Rao, "A review of identification in continuous-time systems," *Annual Reviews in Control*, vol. 22, 1998.

[105] P. Young and A. Jakeman, "Refined instrumental variable methods of recursive time-series analysis: Multivariable systems," *International Journal of Control*, vol. 31, no. 4, pp. 741–764, 1980.

[106] P. C. Young, H. Garnier, and M. Gilson, "Refined instrumental variable identification of continuous-time hybrid box-jenkins models," in *Identification of continuous-time models from sampled data*. Springer, 2008, no. 9781848001602, ch. 4, pp. 91–131.

[107] P. Young, "Parameter estimation for continuous-time models-A survey," *Automatica*, vol. 17, no. 1, 1981.

[108] R. A. González, C. R. Rojas, S. Pan, and J. S. Welsh, "Theoretical and practical aspects of the convergence of the SRIVC estimator for over-parameterized models," *Automatica*, vol. 142, 2022.

[109] H. Garnier, M. Gilson, T. Bastogne, and M. Mensler, "The CONTSID toolbox: A software support for data-based continuous-time modelling," in *Identification of Continuous-time Models from Sampled Data*. Springer, 2008, no. 9781848001602, pp. 249–290.

[110] A. Padilla, H. Garnier, and M. Gilson, "Version 7.0 of the CONTSID toolbox," *IFAC-PapersOnLine*, pp. 757–762, 2015.

[111] S. M. Halat, J. I. Yuz, J. C. Ag, and S. Mar, "On frequency-domain maximum likelihood identification of state-space time-varying systems," in *UKACC International Conference on Control*, 2010, pp. 1–6.

[112] G. C. Goodwin, J. C. Agüero, J. S. Welsh, J. I. Yuz, G. J. Adams, and C. R. Rojas, "Robust identification of process models from plant data," *Journal of Process Control*, vol. 18, no. 9, pp. 810–820, 2008.

[113] M. Gilson, J. S. Welsh, and H. Garnier, "Frequency-domain instrumental variable based method for wide band system identification," in *Proceedings of the American Control Conference*, 2013.

[114] ——, "A Frequency Localizing Basis Function-Based IV Method for Wideband System Identification," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 1, pp. 329–335, 2018.

[115] P. Van den Hof and S. Douma, "An IV-based iterative linear regression algorithm with optimal Output Error properties," in *TU Delft*, 2009.

[116] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1689–1695, 2001.

[117] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[118] K. Burnham and D. Anderson, "Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd edn. Springer, Berlin." 2002.

[119] K. P. Burnham and D. R. Anderson, *Multimodel inference: A Practical Information-Theoretic Approach.* Sociological Methods and Research, 2004.

[120] J. E. Cavanaugh and A. A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements," *WIREs Computational Statistics*, vol. 11, no. 3, p. e1460, 2019. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.1460

[121] P. Stoica and Y. Selén, "Model-order selection: A review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.

[122] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, 1951.

[123] H. Akaike, "On the Likelihood of a Time Series Model," *The Statistician*, vol. 27, no. 3/4, 1978.

[124] P. Young, "Recursive Estimation, Forecasting, and Adaptive Control," *Control and Dynamic Systems*, vol. 30, pp. 119–165, 1989.

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

[126] P. Young and H. Garnier, "Identification and estimation of continuous-time rainfall-flow models," *IFAC Proceedings Volumes*, vol. 39, no. 1, 2006.

[127] A. Gómez González, J. Rodríguez, X. Sagartzazu, A. Schuhmacher, and 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.

[128] A. Klein, T. Sauer, A. Jedynak, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 2, 2006.

[129] P. Stoica and R. L. Moses, *Spectral analysis of signals.* Pearson/Prentice Hall, 2005.

[130] R. Johansson, *System modeling and identification.* Prentice-Hall, 1993.

[131] J. Yuz and G. Goodwin, "Sampled-Data Models for Linear and Nonlinear Systems," *Springer*, 2013.

[132] G. E. Galindo, M. Zanartu, and J. I. Yuz, "A discrete-time model for the vocal folds," in *IEEE EMBS International Student Conference*, 2014, pp. 74–77.

[133] J. I. Yuz and G. C. Goodwin, "On sampled-data models for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1477–1489, 10 2005.

[134] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocalfold model." *The Journal of the Acoustical Society of America*, 1995.

[135] P. J. Hadwin and S. D. Peterson, "An extended \textsc{K}alman filter approach to non-stationary Bayesian estimation of reduced-order vocal fold model parameters," *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. 2909–2920, 2017.

[136] G. E. P. Box, "Science and Statistics," *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 791–799, 1976. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10480949

[137] N. K. Rout, D. P. Das, and G. Panda, "Computationally efficient algorithm for high sampling-frequency operation of active noise control," *Mechanical Systems and Signal Processing*, vol. 56, 2015.

[138] A. V. Oppenheim and G. C. Verghese, *Signals, systems and inference.* Pearson Higher Ed, 2016.

[139] J. G. Fontanet, J. I. Yuz, H. Garnier, A. Morales, J. P. Cortés, and M. Zañartu, "Continuous-time model identification of the subglottal system," *Biomedical Signal Processing and Control*, vol. 95, p. 106394, 2024.

[140] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*, 2nd ed. New York: John Wiley & Sons, 2000.

[141] R. A. González, C. Rojas, S. Pan, and J. S. Welsh, "Parsimonious Identification of Continuous-Time Systems: A Block-Coordinate Descent Approach," *arXiv preprint arXiv:2304.03259*, 2023.

[142] J. P. Cortés, G. A. Alzamendi, A. J. Weinstein, J. I. Yuz, V. M. Espinoza, D. D. Mehta, R. E. Hillman, and M. Zañartu, "Kalman Filter Implementation of Subglottal Impedance-Based Inverse Filtering to Estimate Glottal Airflow during Phonation," *Applied Sciences (Switzerland)*, vol. 12, no. 1, pp. 1–20, 2022.

[143] J. F. Duhe, S. Victor, P. Melchior, Y. Abdelmoumen, and F. Roubertie, "Online system identification of heat transfers in lungs with the LMRPEM-2 method," in *2023 International Conference on Fractional Differentiation and Its Applications, ICFDA 2023*, 2023.