2017

BAYESIAN ESTIMATION OF A SUBJECT-SPECIFIC MODEL OF VOICE PRODUCTION FOR THE CLINICAL ASSESSMENT OF VOCAL FUNCTION

GALINDO FLORES, GABRIEL EDUARDO

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTMENT OF ELECTRONIC ENGINEERING

Bayesian Estimation of a Subject-Specific Model of Voice Production for the Clinical Assessment of Vocal Function

A dissertation submitted by

Gabriel E. Galindo

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy
in Electronic Engineering

Thesis Advisor
Matías Zañartu, Ph.D.

Valparaíso, 2017.

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA

Bayesian Estimation of a Subject-Specific Model of Voice Production for the Clinical Assessment of Vocal Function

Documento entregado por

Gabriel E. Galindo

como requerimiento parcial para la obtención del grado de

Doctor en Ingeniería Electrónica

Profesor Guía
Matías Zañartu, Ph.D.

Valparaíso, 2017.

TITLE:

Bayesian Estimation of a Subject-Specific Model of Voice Production for the Clinical Assessment of Vocal Function

AUTHOR:

Gabriel E. Galindo

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electronic Engineering at Universidad Técnica Federico Santa María.

Documento enviado como requerimiento parcial para la obtención del grado de Doctor en Ingeniería Electrónica de la Universidad Técnica Federico Santa María.

Matías Zañartu, Ph.D. _____

Juan I. Yuz, Ph.D. _____

Sean D. Peterson, Ph.D. _____

Kyle J. Daun, Ph.D. _____

Valparaíso, July 2017.

To ***María Paz***

ACKNOWLEDGMENTS

I want to thank professor Matías Zañartu for his dedication to this work, and for providing me the opportunity to pursue this thesis research. His continuous guidance, extra effort, and support made this work possible. I extend my thank to professor Sean Peterson for his continuous encouragement, enthusiasm and feedback. I also want to thank professors Juan Yuz, Edson Cataldo, Byron Erath, Robert Hillman, and Kyle Daun, which all contributed in the growth and development of this thesis idea.

Many other collaborators contributed to the production of this work, and I want to special thanks Juan Pablo Cortés, Christian Castro, Víctor Espinoza, Paul Hadwin, Andrés Llico, and the entire Voicelab. Which all influenced many aspects of my work.

I cannot forget to acknowledge and thanks my good friends Daniel, Ivan, and Alejandro, who are always questioning my strange ideas and that encouraged me to pursue this Ph.D. In special I want to thank Jorge for showing me what would later become my passion.

I am specially thankfully to my parents and family for the constant support and love. The foundations of my life cannot be separated of their teachings.

Last, but not least, my deepest gratitude goes to María Paz, the most magical woman I have ever known. Her constant support and courage provided me with the confidence to follow my dreams. Without her endless love, patience, and coffee supply, this thesis would have never seen the day light.

I gratefully acknowledge the financial support provided by scholarships from Universidad Técnica Federico Santa María, CONICYT, and MECESUP, as well as support from FONDECYT, and PIIC programs.

Bayesian Estimation of a Subject-Specific Model of Voice Production for the Clinical Assessment of Vocal Function

Presentado por
Gabriel E. Galindo
en el cumplimiento parcial del requisito de grado de
Doctor en Ingeniería electrónica

Profesor Guia
Matías Zañartu, Ph.D.

ABSTRACT

Los problemas de voz son conocidos por afectar negativamente a la comunicación, la interacción social, a las funciones laborales y a la calidad de vida. Se cree que los problemas vocales son condiciones recurrentes o crónicas que están relacionadas a patrones perjudiciales, lo que es conocido como hyperfuncionalidad vocal. Se cree que las principales causas biomecanicas de la formación de lesiones benignas en las cuerdas vocales (ej. nódulos y pólipos) y de las disfonia tenso-musculares son consecuencia de la hyperfuncionalidad vocal. A pesar de la prevalencia de estos trastornos, se conoce muy poco acerca de los mecanismos subyacentes de la hiperfunción vocal, los cuales argumentamos que pueden describirse cualitativamente a través del modelado numérico basado en la física de la función vocal. Exploramos esta idea desarrollando un modelo mejorado de producción de voz y un marco probabilista para estimar sus parámetros. Si bien se ha prestado cada vez más atención a los modelos de producción de voz que pretenden describir el comportamiento fisiológico general, se ha prestado poca atención a los modelos de un sujeto especifico. Se espera que el desarrollo del modelado sujeto-específico mejore el análisis clínico de la función vocal. En la primera parte de esta tesis, se propone un modelo triangular de cobertura y cuerpo de la cuerdas vocales para la producción de voz, permitiendo captura las características clave de la hiperfunción vocal. Se consideraron importantes características anatómicas para mejorar la relación modelo-fisiología del modelo estándar *body-cover model*. El modelo propuesto describe eficazmente el cierre glotal incompleto membranoso y cartilaginoso. La relevancia fisiológica del modelo triangular propuesto se explora en el contexto de la hiperfunción vocal, en la que los mecanismos compensatorios y la retroalimentación auditiva aumentan la presión de contacto de los pliegues vocales junto con las medidas clínicas relacionadas (por ejemplo, flujo inestable y tasa de declinación de flujo máximo).

Utilizando el modelo propuesto, se desarrolló y probó un método de estimación Bayesiano multimodal utilizando tanto datos clínicos como datos generados sintéticamente. El uso de dicho mecanismo de identificación de sistemas permite construir un modelo de sujeto específico basado en un marco probabilístico, extendiendo así esfuerzos previos que sólo producen estimaciones deterministas. El estimador propuesto permitió con éxito la identificación de los parámetros del modelo y sus intervalos de credibilidad. Además, dado que el metodo pose una variación natural

en el tiempo, también se obtuvieron señales relacionadas con el modelo que no son observables en la configuración clínica (por ejemplo, presión de contacto de los pliegues vocales).

Para la evaluación de los datos clínicos, se utilizó como fuente de información la videoendoscopia de alta velocidad y las mediciones de flujo glotal. La predicción y el contraste se hicieron usando una señal de micrófono adquirida en sincronía con las otras mediciones. La señal de micrófono estimada asemejó la medida clínica dentro de sus bandas de credibilidad. Por lo tanto, se ilustra el potencial del método propuesto para construir y extraer información adicional relacionada con el modelo de las configuraciones clínicas. Las estimaciones de la presión de contacto obtenida con la técnica propuesta se ajustaron estrechamente a estudios previos. Así, el marco de modelaje propuesto permite investigar los vínculos causales entre la patología y/o compensación del paciente y la hiperfunción vocal, tanto por elucidar algunos de los mecanismos físicos subyacentes como también por proporcionar nuevas medidas de la función vocal, las cuales son difíciles, si no imposibles, de obtener en ambientes clínicos.

Keywords – Cuerdas vocales, pliegues vocales, identificación de sistemas, estimación Bayesiana, Hyperfuncionalidad vocal.

Bayesian Estimation of a Subject-Specific Model of Voice Production for the Clinical Assessment of Vocal Function

submitted by
Gabriel E. Galindo
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy in Electronic Engineering

Thesis Advisor
Matías Zañartu, Ph.D.

ABSTRACT

Voice problems are known to negatively affect communication, social interaction, work-related functions and quality of life. Common voice disorders are believed to be chronic or recurrent conditions related to detrimental voice patterns, which are referred to as vocal hyperfunction. It is believed that the main biomechanical causes of benign vocal folds lesions formation (e.g., nodules and polyps) and muscle tension dysphonia are a consequence of vocal hyperfunction. In spite of the prevalence of these disorders, very little is known about the underlying mechanisms of vocal hyperfunction, which we argue that can be qualitatively described through physic-based numerical modeling of vocal function. We explore this idea by developing an enhanced model of voice production and a probabilistic framework to estimate its parameters. While increasing attention has been placed on voice production models that aim to describe general physiological behavior, little attention has been placed on single-subject models. The development of subject-specific modeling is expected to enhance clinical analysis of vocal function. In the first part of this thesis, a triangular body-cover model of voice production that capture the key features of vocal hyperfunction is proposed. Several important anatomical characteristics were considered to improve the model-physiology relation of the standard body-cover model. The proposed model effectively describes membranous and cartilaginous incomplete glottal closure. The physiological relevance of the proposed triangular model is explored in the context of vocal hyperfunction, where compensatory mechanisms and auditory feedback yield increased contact pressure of the vocal folds, and related clinical measures (e.g., unsteady flow, and maximum flow declination rate).

Using the proposed model, a multi-modal Bayesian estimation method was developed and tested using synthetically generated and clinical data. The use of such system-identification mechanism allows for constructing a subject-specific model based on a probabilistic framework, thus extending previous efforts that only produce deterministic point-based estimates. The proposed estimator successfully allowed for the identification of model parameters, and their associated credibility intervals. In addition, given the time-varying nature of the method, model-related signals that are non-observable in clinical setup were also obtained (e.g., vocal folds contact pressure).

For the assessment of clinical data, high-speed videoendoscopy and glottal flow measurements were used as information source. Prediction and contrast was made using a separate microphone signal acquired in synchrony with the other measurements. The estimated microphone signal matched the clinical measurement within its credibility bands, illustrating the potential of the proposed method to construct and extract additional model-related information from clinical setups. Estimates of contact pressure obtained with the proposed technique closely matched previous studies. Thus, the proposed numerical modeling framework enables the investigation of causal links between patient pathology/compensation and vocal hyperfunction by both elucidating some of the underlying physical mechanisms and providing new measures of vocal function that are difficult, if not impossible, to obtain in clinical setups.

Keywords – Vocal folds, System identification, Bayesian estimation, Vocal Hyperfunction.

TABLE OF CONTENTS

List of Figures

List of Tables

List of symbols and abbreviations

Abbreviations

f_0 fundamental frequency 69, 71, 72

AC-Flow amplitude of unsteady glottal flow 4, 65, 69, 75, 77, 79

ARMA auto regresive moving average 20

BCM body-cover model 2–4, 11–13, 15, 17, 27–29, 31, 33, 35–37, 43, 45, 47, 48, 63, 84, 107, 115

CA cricoarytenoid 67

CT cricothyroid 8, 9, 17, 53, 65, 72, 73, 89, 90, 94, 95, 97, 101, 105, 107, 108, 112

HMM hidden Markov model 85, 92

HNR harmonic to noise ratio 69, 71–73, 75, 77, 79, 81

HRF harmonic richness factor 71, 72, 75, 77, 79, 81

HSV high speed videoendoscopy 21, 94, 103, 105, 109, 111, 116, 117

IA interarytenoid 8, 9

LCA lateral cricoarytenoid 8, 9, 17, 37, 53

LPC linear predictive coding 20

MAP maximum a *posteriori* 106, 107, 109–111, 116

MCP maximum contact pressure 69, 75

MFDR maximum flow declination rate 4, 33, 65, 69, 75, 79

MGO membranous glottal opening 36, 37, 45, 48, 63, 67, 69, 72, 107

MRI magnetic resonance imaging 19, 36, 63

MTD muscle tension dysphonia 27, 28, 37, 91

Min-Flow minimum glottal flow 65

NET net energy transferred 69, 75

ODE ordinary differential equation 49, 53, 56, 63

PCA posterior cricoarytenoid 8, 9

PGD posterior glottal displacement 45, 47–49, 53, 72, 73, 75, 77, 105

PGO posterior glottal opening 27, 28, 31, 33, 35–37, 43, 45, 48, 63, 65, 67, 69, 71, 72, 95, 103, 107, 111, 113, 115

QD quality distance 71–73, 75

RMS root mean square 53, 66, 109

SD standard deviation 75, 89, 94, 107

SMCMC sequential Markov chain Monte Carlo 92, 93, 97, 99, 101, 116

SMC sequential Monte Carlo 92
SNR signal to noise ratio 69
SPL sound pressure level 65, 69, 71–73, 75, 77, 79, 81
SSM state space model 23, 59, 63
TA thyroarytenoid 8, 9, 17, 37, 53, 67, 72, 73
TBCM triangular body-cover model 3, 37, 39, 47–49, 51, 53, 54, 63, 67, 69, 87, 89, 94, 95, 101, 107, 111, 113, 115, 116
TLS transmission line scheme 19
TMM two mass model 21
TTS truncated Taylor series 49, 51, 53, 55, 56, 63, 87, 115
VF vocal folds 1–5, 7–9, 11, 12, 15, 17–21, 24, 25, 28, 33, 36, 37, 39, 41, 45, 47, 49, 51, 53, 55, 67, 69, 71, 77, 79, 81, 83–86, 89, 95, 101, 103, 105, 113, 115, 117
VH vocal hyperfunction 1–5, 67, 69, 71, 75, 77, 79, 81, 101, 109, 115
WRA wave reflection analog xvii, 19, 27, 29, 51, 55, 57, 59, 63, 64, 66, 109
iid independent and identically distributed 87, 92
pdf probability density function 4, 21, 83, 85–87, 91, 92, 97, 98, 107, 111, 116, 117
pmf probability mass function 87
rv random variable 21, 23, 84, 85

Symbols

A^* Effective trans-glottal area 15, 31, 35
A_{MGO} Onset area of the membranous portion of the glottis 45
A_{PGO} Onset area of the cartilaginous portion of the glottis 45
A_e^c supra-glottal compression of the epilaryngeal tube section 72, 73
A_g Glottal area 15, 29, 31, 35, 36
A_l lower-mass glottal area 13, 15, 43, 47
A_m Minimum area of membranous glottal channel 29, 33, 35, 45, 47
A_p Area of posterior glottal channel 29, 33, 35, 45
A_s Sub-glottal tract area 15, 31
A_e Supra-glottal tract area 15, 29, 31
A_u Upper-mass glottal area 13, 15, 43, 47
a_r^d Arytenoid displacement 45, 53, 65, 72, 105
ℓ_a Arytenoid length 45, 65, 105
a_r° Arytenoid rotation 45, 48, 53, 65, 68, 72, 105
$\alpha_{\mathbf{x}}$ Normalized collision factor on element $\mathbf{x}$ 43
$d_{\mathbf{x}}$ Damping coefficients on element $\mathbf{x}$ 41
ρ Air density 15, 29, 31, 35, 36
δ Distance of a determinate voice vector 71
$\mathbf{x}$ Conceptual space of grouped independent elements xvi–xviii, 39, 41, 43
f_0 Fundamental frequency 36, 65, 71, 73, 75, 77, 79, 81
$F_{\mathbf{x}}$ Force due to the $\mathbf{x}$ phenomena 39
f_s Sampling frequency 56, 63
Q_g Volumetric glottal flow 15, 29, 31, 33, 35, 36

v_m Flow velocity of membranous glottal channel 29
Q_m Membranous volumetric glottal flow 33, 47
Q_n Turbulent volumetric glottal flow 35, 36
v_p Flow velocity of posterior glottal channel 29
v_s Flow velocity on sub-glottal tract 29
v_e Flow velocity on supra-glottal tract 29
Q_T Total volumetric glottal flow (including noise sources) 36
$\mathcal{H}$ Heaviside step function 35, 41, 43, 45, 47, 48
D_H Hydraulic diameter 36
k_t Trans-glottal kinetic loss coefficient 15, 29, 31, 35
$\delta_{i,j}$ Kronecker delta function 59
ℓ_c Closed membranous glottal length 47, 48
ℓ_d Divergent membranous glottal length 47, 48
ℓ_g Membranous glottal length 13, 15, 17, 18, 33, 41, 43, 45, 47, 48
a_{CT} Activation of the cricothyroid muscle 17, 18, 65, 73, 95, 105
a_{LC} Activation of the cricoarytenoid muscles 17, 18, 105
a_{TA} Activation of the thyroarytenoid muscles 17, 18, 73, 105
$m_{\mathbf{x}}$ Mass of element $\mathbf{x}$ 39
μ_ν Mean value of a variable 35
ν Withe noise source 35
δ_{x_l} Posterior rest displacement of the lower mass 41, 43, 45, 48
δ_{x_u} Posterior rest displacement of the upper mass 41, 43, 45, 48
x_{PGD} Posterior Glottal Displacement 45
P_d pressure at the exit glottal plane 29
P_l Lower mass driving pressure 13, 15, 47
δ_{P_m} Membranous trans-glottal pressure drop due to viscous effects 33
p_e^+ Departing Supra-glottal dynamic pressure 31
p_e^- Incident Supra-glottal dynamic pressure 15, 31
p_s^+ Incident Sub-glottal dynamic pressure 15, 31
p_s^- Departing Sub-glottal dynamic pressure 31
$x_{\mathbf{x}}$ Position of element $\mathbf{x}$ 39
P_s Sub-glottal static pressure 15, 29, 31, 35, 47, 65, 73, 105
P_e Supra-glottal static pressure 15, 29, 31, 35, 47, 105
δ_{P_g} Total trans-glottal pressure drop due to viscous effects 33
δ_P Trans-glottal pressure 31, 35
P_u Upper mass driving pressure 13, 15, 47
$\mathcal{F}$ Common coupling function for wave reflection analog (WRA) tract bifurcations 59, 61
N Number of tubes used in a WRA tract 57, 59, 61
$\mathbf{p}$ Acoustic pressures vector used in WRA 57, 59
Δ_z Length of each tube in WRA 57
α Loss coefficients for acoustical traveling waves 55, 57, 59, 61
p^+ Forward acoustic pressure 55, 61

p^- Backward acoustic pressure 55, 61
r Juncture reflection coefficient 55, 57, 59, 61
$\mathbf{Pr}(X)$ Probability of random variable X 85
R_g Glottal flow resistance 33, 35, 48
R_m Membranous glottal flow resistance 33, 47, 48
R_p Posterior glottal flow resistance 33
Re Reynolds number 35, 36
γ_a Proportion factor of arytenoid length open to the glottal flow channel 45
γ_m Membranous resistance glottal flow factor 15, 33, 35
γ_p Posterior resistance glottal flow factor 33
r_e Supra-glottal reflection coefficient 31
r_s Sub-glottal reflection coefficient 31
c Speed of sound 15, 31, 35
$\zeta_{\mathbf{x}}$ Damping ratio on element $\mathbf{x}$ 41
$k_{\mathbf{x}}$ Non-linear spring coefficients on element $\mathbf{x}$ 39
$\eta_{\mathbf{x}}$ Non-linear spring coefficients on element $\mathbf{x}$ 39
σ_ν Standard deviation value of a variable 35
T_g Membranous glottal thickness 17, 18, 33, 47, 48
T_l lower-mass glottal thickness 15, 18, 47
τ Tolerance factor 71, 72
T_u Upper-mass glottal thickness 15, 18, 47
ω Vector of variables used in optimization cost functions 71
μ Air viscosity 33, 36, 47, 48
ω_T Target vector of variables used in optimization cost functions 71, 72
β Weight vector for different measurements in a cost function 71, 72

Chapter 1

INTRODUCTION

1.1 Overview

Voice problems are known to negatively affect communication, social interaction, work-related functions, and quality of life.[139] These disorders affect a significant percentage of the population, and have been reported as one of the main occupational health problems.[124,125] In Chile, voice related problems affect approximately 75% of the school teachers, and are the second cause of occupational health problems among the total working-age population.[15]

The most common voice disorders are believed to be chronic or recurrent conditions related to repeated detrimental vocal patterns, known as vocal hyperfunction (VH).[60] This includes the impairment of laryngeal muscle control without structural abnormalities as seen in muscle tension dysphonia (referred as non-phnotraumatic VH),[125] and the formation of benign vocal folds (VF) lesions such as nodules and polyps (referred as phonotraumatic VH).[60]

One of the main advantages of the numerical models of voiced speech is the ability to obtain data that is, in many cases, almost impossible to accurately obtain in a clinical setup. In addition, they allow for the study of a wide range of controlled case scenarios, with a fraction of the cost and time of other methods (e.g., physical models, ex-vivo experiments, etc.). The use of numerical models for the investigation, diagnosis, and treatment of voice disorders has been applied broadly in mimicking and predicting complex physical phenomena.[89,197] In particular, numerical models have been used to characterize phonotraumatic VH, and its relation with the pathogenesis of phonotraumatic lesions through *ad-hoc* modifications of model parameters, mimicking similar kinematic behaviors from those observed in *in-vivo* and *ex-vivo* larynx experiments.[58,146,182,196]

It is proposed in this thesis that the main biomechanical causes of VH can be quantitatively described through physics-based modeling. That is, that numerical models of voice production can be accurately defined to mimic VH mechanisms as well as normal phonation. The proposed modeling framework consist of a series of modifications to the current VF models to include common anatomical characteristics, such as posterior glottal openings, incomplete glottal closure and onset conditions. The inclusion of such characteristics will allow for a more accurate representation of measurements of interest, such as collision forces, muscle activation parameters, and kinematic behavior in general. Thus, the proposed additions will help to better understand the underlying phenomena associated with healthy and pathological vocal function. Even though increasing attention has been placed on utilizing numerical models to describe general physiological behaviors, little attention has been placed on models that aim to describe a single person,

namely subject-specific modeling.

The development of subject-specific modeling is expected to enhance the clinical analysis in the assessment of vocal function. In this thesis, a mathematical method based on system identification tools is developed to adjust a numerical model by finding stochastic model parameters that allow for matching individual clinical recordings with known uncertainty. In addition, it is proposed that such estimation will extend the capabilities of current objective methods for voice assessment, providing access to additional measures of vocal function that are difficult to obtain in clinical setups. In this way, the clinical analysis will not be limited to a single result, but rater to a probability of occurrence, yielding an improved interpretation of the estimated information.

1.2 Aims

The general aim of this thesis is to develop numerical tools for the investigation of general and individual vocal behaviors, particularly focusing on VH. The main assumption is that key biomechanical characteristics can be quantitatively described through data-based modeling with physical knowledge (also known as greybox modeling). Therefore, to generate such framework, we first evaluate the capability of a generic numerical model of voice production to reproduce certain VH conditions. Subsequently, we develop an estimation procedure that allows for the determination of a subject-specific model parameters based on clinical measurements. Thus, the following specific aims are proposed:

- **Specific Aim 1:** *To design a numerical model to describe normal and hyperfunctional vocal behavior, in a computationally efficient framework.*
  A series of anatomical improvement are proposed over the body-cover model (BCM),[154] a reduced-order numerical VF model broadly used in the analysis of human phonation.[44] These anatomical improvements consider the inclusion of posterior glottal opening, arytenoid posturing, auditory feedback, flow enhancement, and computational efficiency. We expect to produce a model with accurate anatomical and physical representations to study VH and normal phonation.

- **Specific Aim 2:** *To develop a Bayesian framework to construct time-varying estimates of a subject-specific model.*
  The proposed approach is expected to produce a subject-specific model capable of estimating the underlying parameters with known uncertainty. This is initially tested with synthetic signals as benchmark. The method should allow to estimate time-varying signals that are otherwise hidden from clinical setups, thus increasing the clinical assessment through model-related virtual signals.

- **Specific Aim 3:** *To assess the proposed Bayesian framework using clinically available data.*
  Using the numerical voice model of Specific Aim 1 and the Bayesian estimation method of Specific Aim 2, a subject-specific model and virtual sensors signals are obtained using clinical recordings. This assessment requires overcoming a series of difficulties, such as different sampling rates, measurement noise, signal synchrony, among others. To provide a

proof of concept, two different estimations are performed using the same configuration. The first, simulates clinical conditions in a synthetically generated data set, allowing a direct comparison between ground truth and the estimated system. Subsequently, using the same estimation configuration, the simulated data is replaced with the clinically acquired data, yielding a subject-specific model using clinical data. As a consequence, virtual sensor data will also be available for comparison and further analysis of the method.

In summary, the aims of the project can be synthesized in the following hypothesis: *"If a set of calibrated time-varying measurements is provided, then a voice production model can be specially tuned to reproduce such measurements using non-stationary Bayesian estimation"*. As a corollary: *"If a non-stationary Bayesian subject-specific-model estimation is provided, then through it time-varying clinically-hidden signals can be accurately estimated"*.

1.3 Overview of the proposed methodology

The methodology can be separated in four different steps that are performed consecutively, these are: (1) model modification for anatomical considerations, (2) computation improvement in numerical models, (3) Bayesian estimation on VFs model, and (4) clinical application of Bayesian estimation.

The BCM[154] was chosen, as a starting point, given its broad use setups to the one required for the proposed research, and its advantages over other numerical models of voice production.[44] This model offers a separation of body a cover layer of the VFs, which allows for better describing the system kinematics with low computational cost. In addition, the use of control rules associated with muscle activation,[173] provides the BCM with a convenient way of setting model parameters by directly relating anatomical characteristics. This model was further developed into a triangular body-cover model (TBCM) in order to allow for a posterior gap and arytenoid rotation, which are critical features to model VH.

In addition to the anatomical modifications of the model, a few improvements were made to reduce computational requirements. The analytical solution of the TBCM requires solving a series of differential equations over each computational step. For this purpose, it is usual to obtain such solution through an iterative numerical method such as Runge-kutta. However, these numerical solvers are computationally expensive, making them less attractive when running several models in parallel (e.g., particle filtering). To avoid such problems, a Truncated Taylor Series[189] approach was developed. Moreover, a state space implementation of the acoustic propagation method was implemented, thus further reducing the computational demands.

Later on, an inverse problem using Bayesian estimation was explored to estimate model parameters from synthetic data, thus allowing to quantify the estimation error. Two Bayesian estimation methods were compared, stationary and non-stationary estimation. Subsequently, a clinical analysis was performed combining the model proposed modifications, Bayesian estimation, and clinically acquired data. In this process, an edge detection technique was applied to a calibrated high speed video-endoscopy recording. From the time-varying glottal edge, a times series signal of the glottal area was extracted and used to produce a Bayesian subject-specific model.

1.4 Significance and expected contributions

One of the causes of phonotraumatic VH is believed to be linked to the application of incorrect compensation mechanisms, such as increased sub-glottal pressure to compensate a reduced loudness, or increasing intrinsic laryngeal muscles to improve spectral characteristics of the voice. It has been stated that these compensatory mechanism lead to a vicious cycle that perpetuates vocal deficiency.[60,61] However, and in spite of the significant prevalence of these common voice disorders,[125] little is known on the underlying mechanisms that lead to VH. This thesis work opens the path for a broader study on the underlying causes of VH.

On the other hand, the use of inverse analysis has been used in lumped-mass models to determine subject-specific parameters with aims of pathology classification,[136,137] using high speed video-endoscopy of normal[126,185] and pathological voices.[137] As part of this thesis, a Bayesian framework to produce subject-specific models, using high speed video-endoscopy was developed.[56] Herein, several synchronized clinical recordings were used to stochastically identify a non-stationary model of the VFs, thus producing a set of model parameters that describe the observed data with an associated credibility interval. In addition, the use of non-stationary scenarios yields additional data estimate that can be considered as a "virtual sensor" since they produce a time-varying signal that is model-related to the observed clinical data.

As part of this thesis, and following up from previous efforts in our group,[192] additional modifications to the BCM were done to improve the comprehension of VH[50] mechanism. The contributions included new compensatory mechanisms that were used to offset glottal phonatory insufficiency, in particular, posterior glottal openings and incomplete membranous closure, reducing glottal efficiency, radiated sound pressure level, and harmonic richness in the resulting voiced sounds. Consequently, instead of fixing the underlying cause of the deficiency, compensatory mechanisms were triggered to achieve "normal phonation" characteristics, which translate in the recovery of vocal measures to values obtained when no modifications were made. The effect of compensating the incomplete glottal closure with compensatory mechanism resulted in a significant increase of acoustic measures such as maximum flow declination rate (MFDR) and amplitude of unsteady glottal flow (AC-Flow), closely mimicking previous findings in clinical studies.[60,116] Furthermore, the observations of collision pressure and energy transfer to the VFs were found highly correlated with the increase in MFDR and AC-Flow, illustrating the insight that low lumped-mass models can provide for measurements and data that is otherwise hidden in clinical setups.

Using the proposed model, a novel method to extract VFs model parameters, obtaining not only the most probable configuration, but also acquiring the probability density function (pdf) for all possible configurations was developed. In addition, and given the non-stationary nature of the estimation method, a completely new estimated signal can be recovered from the system identification step, allowing to estimate a virtual sensing by the direct observation of a model-related measurement. This virtual sensor acquisition can provide useful information to clinical analysis that required the time-series data, and not just a feature extraction (e.g., Ambulatory assessment of vocal function).

Therefore, the proposed framework is expected to provide insights into the underlying biomechanical descriptions of VFs behavior, thus allowing for a more comprehensive assessment of

human phonation. In addition, the increases capability to estimate VFs contact provides a new insight to the pathogenesis of VFs lesions with a minimally invasive procedure. The proposed numerical modeling framework is expected to provide the first causal links between patient pathology/compensation and VH by elucidating the physical mechanisms of the pathophysiology of VH and providing new measures of vocal function that are difficult, if not impossible, to obtain in clinical setups. Consequently, this thesis work contributes to delineate the pathogenesis of voice related problems while also providing improved methods for diagnosis of hyperfunctional behaviors.

1.5 Document structure

The chapters in this thesis document are structured as follows: Chapter 2 is a background chapter that includes the state of the art, the general description of voice production models, the current phonation theory, the base of system identification, and the basic ideas of Bayesian estimation theory. Chapter 3 provides insights on the model developed in this research by introducing a series of technical and anatomical modifications made to improve model efficiency and make the model more physiologically relevant. Chapter 4 presents the foundation for stochastic system identification on general VFs models. Herein, the theoretical capabilities of the proposed research are presented with a survey of two different methods. Chapter 5 presents a clinical case study to illustrate the functionality of the methods described earlier. Chapter 2 provides conclusions and discussion of the capabilities that this research brings into the clinical assessment of voice production, as well as the outline for future work.

Chapter 2

BACKGROUND

In this chapter, relevant aspects of voice theory are presented along with the outline of the system identification process. The current state of the art for both topics, voice theory and system identification of vocal folds (VF), is discussed in detail.

2.1 Theory of phonation

Phonation theory has evolved from early magical and mythic explanations for the occurrence of voice,[4] to self-sustained theories of phonation that account for kinematic, acoustic and aerodynamic components.[25, 75, 156, 167, 179] Currently, the Myoelastic theory[167] is the most accepted theory of phonation, stating that the phonation is produced by the vibration of the VF interacting with sub-glottal pressure along with vocal tension. The Myoelastic theory also includes the muco-ondulatory theory,[115] which emphasize the importance of the mucosal wave and the degree of glottal adduction. Analogies with other systems have been used to explain phonation, such as the negative resistance theory,[4, 22] that states the VF movement is produced by a complex oscillator, and that has been used to explain vibratory patterns observed in mass lesions of the VF.[4, 22]

There are several anatomic elements involved in phonation, which can be grossly separated into respiratory tract and the larynx, where the later is the organ acting as a voiced sound source. The respiratory tract can be further separated, from inferior to superior elements, into: lungs, bronchi, trachea, larynx, pharynx, and oral/nasal cavities. On the other hand, inside the larynx, there is a set of layered tissues and muscles that have the ability to narrow the glottal cavity in order to produce self-oscillatory movement by the interaction of airflow, sub-glottal pressure, and sound waves.

As in every oscillatory system, the production of movement requires an energy source, which for the voice production system is provided by the lungs. The energy source of voice production is the air reserved in the lungs on inhalation, where the lung capacity increases, producing a negative pressure that drives the air into the bronchi and lungs. During this gesture, the VF are usually fully abducted to allow the air to pass freely through the larynx. To phonate, an expiration gesture is produced while adducting the VF, producing a narrowing in the larynx area, thus increasing the pressure inside the lungs and trachea. This excess of pressure can lead to self-oscillatory movement of the VF, inducing a pulsatile airflow, that is later modulated (filtered) by the superior vocal tract. During this process, the adduction movement of the VF produces

a convergent geometry of the VF superior-inferior edges. This shape enables a Bernoulli flow phenomenon that drives the VF apart, up to a divergent configuration, where a jet configuration takes place. This new regime reduces the pressure and, along with the elastic forces of the tissue, restores the initial configuration. This recursive process of convergent-divergent positions produces a time-varying opening, referred to as glottis, that pulsates the airflow. The inferior-superior movement of the glottal edges is defined as mucosal wave, and is a key component of healthy vocal fold behavior.[180]

The airflow produced by the oscillatory movement of the VF is later modulated (filtered) by the superior vocal tract, which has the ability to change its shape to produce different articulations. This process allows for the production of the so-called "formants" (resonance) in speech, which is no other than the amplification of certain frequencies which produces a given voiced sound. This resulting airflow is also filtered by the lips and nose, and is radiated into the surrounding air as voiced sound. During this process, reflected sound waves still remain in the superior vocal tract altering the incident pressures of the VF, which in turn affects the driving pressures and alter the characteristics of the voice source. This interaction is known as source-filter interaction, and has been shown to be a key component in voice production.[166, 191, 195] Finally, all this process of flow-tissue-sound interaction can be grouped as the Myoelastic-Aerodynamic theory of phonation,[167] which is the base of the phonation theory used in this research.

2.2 Vocal folds and vocal tract physiology

To understand the behavior and properties of voiced sounds, it is essential to understand the composition of the VF, and its surrounding structure. The larynx is composed of a set of muscle and cartilages mechanically linked to produce tension and displacement of its inner structure. The laryngeal cartilages can be separated into thyroid cartilage, cricoid cartilage, and arytenoid cartilages. The posturing of these 3 cartilages (considering the arytenoid cartilages as one), set the base for the phonatory and respiratory function of the larynx, intrinsically controlling voice pitch, breathiness, loudness and VF contact.

The muscles in the larynx can be divided into intrinsic and extrinsic, depending on the strict attachment to cartilages within the larynx (intrinsic) or their attachment to other surrounding structures (extrinsic). The intrinsic muscles are: thyroarytenoid (TA) muscle, that connect the two arytenoid cartilages with the thyroid cartilage; the cricothyroid (CT) muscle, which connects the anterior part of the thyroid with the anterior portion of the cricoid cartilage; the lateral cricoarytenoid (LCA) muscle, which connects the side of the arytenoid with the sides of the cricoid cartilage; the posterior cricoarytenoid (PCA) muscle, which is antagonist of the LCA muscle, connecting the side of the arytenoids to the posterior portion of the cricoid cartilage; and the interarytenoid (IA) muscle, which connects both arytenoid cartilages together.[164]

The TA, CT, LCA, PCA are paired muscles. This means that each muscle is separated into sided muscles that are dependent to the same neuronal control. In theory, this means that for a normal subject the activation rate for each paired side should be in synchrony with the other. In addition, the TA and IA can be further separated into two additional muscles. For the TA, a sub division on the thyrovocalis muscle (also referred as *vocalis* muscle), and the thyromuscularis

is suggested.[164] This division, however, is still the subject of debate since the functional division of two bundles has not been completely clarified. On the other hand, the IA is clearly separated into a transverse part (unpaired) and an oblique part (paired). Allowing for the adduction and inclination of the arytenoid cartilages.

The combined efforts of the muscular process, to control the positioning of the different cartilages, can be set in three different configurations known as "posturing", "resting", and "respiration". These configurations refer to positioning the VF in pre-phonatory, rest, and respiratory conditions, respectively. The IA and LCA muscles serve as an adductor of the VF, while the PCA is the principal abductor. On the other hand, the TA and CT, allow for controlling the tension and length of the VF, respectively.

The morphology of vocal fold soft tissue can be separated in a muscular layer (TA muscle) and lamina propria, and epithelium. note that the lamina propria can be further divided into a series of layers due to the properties of their cells.[164]

The outer layer, the epithelium, is a thin skin layer that cover the whole structure of the VF with an approximate thickness of 0.05 to 0.1$[mm]$.[65] The superficial layer, beneath the epithelium, is the first layer of the lamina propria and is constituted of elastin fibers and interstitial fluids. The intermediate layer is also constituted of elastin fibers, but with fibers that are uniformly aligned in the anterior-posterior direction with the addition of some collagen fibers. The inner layer of the lamina propria is called "deep layer", and is constituted from nearly inextensible collagen fibers.[164]The three layered lamina propria have an approximated thickness of 1.5 to 2.5$[mm]$, which is small in comparison to the muscular layer, which is approximately 7 to 8$[mm]$ thick. The mucosal wave is generally attributed to the lamina propria structure and composition[64, 164, 167] rather than other layers.

Table 2.1 shows the different groups used to decompose the VF structure. Figure 2.1 has a graphical representation of the VF composition.

The muscular activation has several effects on the tension and onset condition of phonation (the instant right before phonation is produced). The activation of the TA muscle produces an increment in the tension of the VF, which also increases its rigidness and also shorten its length. This produces, as a result, an increment of the fundamental frequency and a decay in the amplitude of vibration of the VF. The activation of the CT muscle produces an elongation of the VF, decreasing their thickness, and increasing their tension, and also increasing the fundamental frequency. The IA muscle activation, along with the combined effect of the LCA and PCA muscles, produces a narrow channel with variant convergence degree depending on the equilibrium between the forces of these muscles; at the same time, the action of these

Table 2.1. Comparison of the common vocal folds division[164]

Layers	Five Layers Scheme	Tree Layers Scheme	Two Layers Scheme
Epithelium	Epithelium	Mucosa	Cover
Lamina Propria	Superficial Layer		
	Intermediate Layer	Ligament	
	Deep Layer		Body
Muscle	Muscle	Muscle	

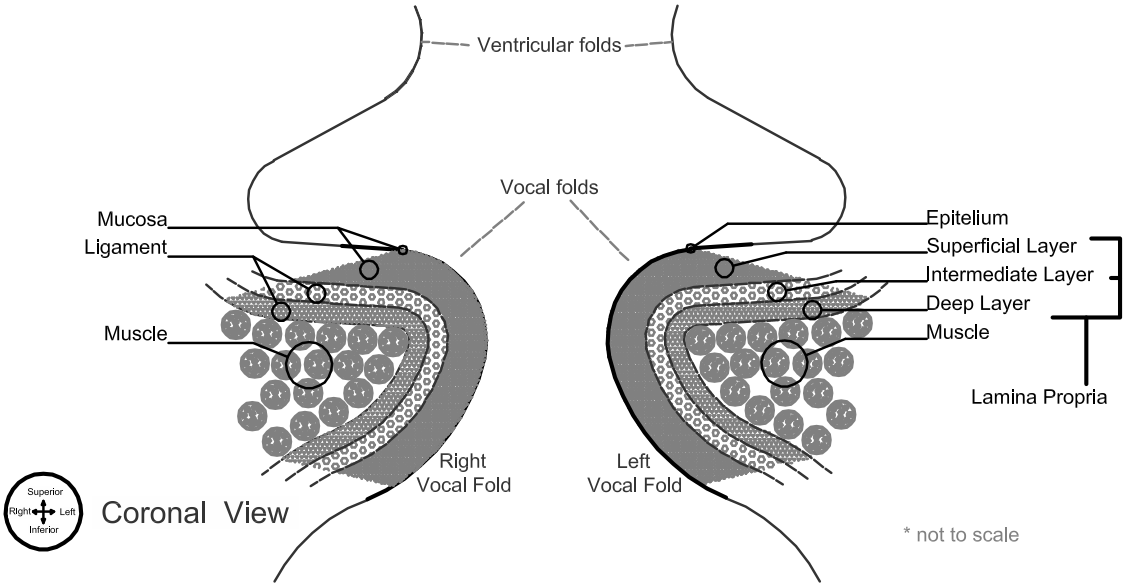


Figure 2.1. Vocal folds layered structure

muscles produces an elongation of the vocalis muscle, increasing its tension and narrowing the glottal channel.[74, 169, 173] Furthermore, some configurations of the arytenoid cartilages can produce incomplete glottal closure at the cartillaginous posterior portion of the glottis and possibly at the membranous portion as well.[192] This incomplete closure has been related with normal and pathological phonation and has even been linked with the pathogenesis of vocal hyperfunction.[60, 61] More detailed information on the structure and physiology of the larynx can be found in technicals textbooks.[164, 167]

2.3 Brief review of vocal fold models

There are several numerical implementations that describe the behavior of the VF. The difference between these models comes from the need to better illustrate different aspects of human phonation. Therefore, there are models that have highly detailed kinematic description in detriment of accurate tissue flow interaction. In contrast, there are models that have all interactions considered important in voice production, but with rather unrealistic glottal shapes. According to Titze's model classification,[167] numerical models of voice production can be separated into four groups: (1) low dimensional models, (2) high dimensional models, (3) continuum models, and (4) finite element models. Low dimensional models represent the movement of the VF by the kinematic interaction of lumped elements. These models lack spatial precision and representation of biological structures. However, their simplicity allow for the reproduction of scenarios that would be highly computationally expensive for other types of models (e.g., optimization and identification procedures). High dimensional models are the logical extension of the low dimensional models, herein, the amount of masses are increased to the point where the structures can be considered as points of masses bounded by kinematic interactions with their neighbors (usually linked with springs and dampers). Thus allowing for a better anatomical representation, with a configuration usually more complex than that of low dimensional models. Continuum models are, an extension of the increasing dimensionality of lumped element models. In this case, the infinitesimal springs and damper can be used to approximate an elastic continuum, which can later be analytically defined using normal-modes of vibration.[167] Finally, finite element (and finite differences) models start from the continuum solution to produce a grid that resembles the biological structure. This grid takes properties from the continuum definition and discretizes them into a set of nodes and binding properties.[167]

The use of lumped element models has been shown to be useful to reproduce the phenomena under VF production,[44, 78, 154, 195] even producing natural-sound voices,[9, 88] and serving for the clinical assessment of VF.[192] During this research, a particular branch of lumped element models, known as BCM,[154] will be used. These models use a two-layer configuration (the body and cover) with varying number of masses in the inferior-superior direction, and a single mass on the anterior-posterior direction.

As a particular case for this research, a modification on the shape of the body-cover model (BCM) masses are performed in order to better describe contact forces and transient phenomena in the closing phase. However, it is important to note that several other voice production models are available, and that a review of all the alternatives is out of the scope of this thesis.

Nevertheless, over the past years, significant efforts have been made to classify the models by its structure, functionality and evolution. For further information the reader can review Titze and Alipour 2006,[167] Birkholz and Kröger 2011,[11] and Erath et al. 2013.[44]

Given that the model developed in this research is an extension of the BCM of Story and Titze,[154] a detailed analysis of its construction will be presented for completeness.

2.4 Body-cover model and its physiological rules

The BCM[154] is a double-layered model of VF with two masses in the cover layer and one mass on the body layer(see Figure 2.2). The movement of each mass is limited to the lateral axis only, with each mass linked as shown in Figure 2.2. This produces the following forces for each mass:

$$F_u = m_u\ddot{x}_u = F_{ku} + F_{du} - F_{kc} + F_{eu} + F_{uCol}, \tag{2.4.1}$$

$$F_l = m_l\ddot{x}_l = F_{kl} + F_{dl} + F_{kc} + F_{el} + F_{lCol}, \tag{2.4.2}$$

$$F_b = m_b\ddot{x}_b = F_{kb} + F_{db} - (F_{ku} + F_{du} + F_{kl} + F_{dl}), \tag{2.4.3}$$

where, F_u, F_l, F_b, are the forces in the upper, lower and body masses, respectively, with mass m and position x. The applied forces then are composed by the following elements:

- F_{ku}, F_{kl}, and F_{kb} are the forced due to the lateral springs,
- F_{du}, F_{dl}, and F_{db} are the forced due to the lateral dampers,
- F_{kc} is the coupling force of the cover masses,
- F_{uCol} and F_{lCol} are the contact forces, which are present only on the closure of the VF.
- F_{eu} and F_{el} are the forces due to the pressure in the glottal channel,

By defining the equilibrium position of each mass to be x_{u0}, x_{u0}, and x_{u0} for the upper, lower, and body mass, respectively, Then each of the forces acting in the system can be represented by the following equations:

Forces due to lateral springs:

$$F_{ku} = -k_u\left\{[(x_u - x_{u0}) - (x_b - x_{b0})] + \eta_u\left[(x_u - x_{u0}) - (x_b - x_{b0})\right]^3\right\}, \tag{2.4.4}$$

$$F_{kl} = -k_l\left\{[(x_l - x_{l0}) - (x_b - x_{b0})] + \eta_l\left[(x_l - x_{l0}) - (x_b - x_{b0})\right]^3\right\}, \tag{2.4.5}$$

$$F_{kb} = -k_b\left[(x_b - x_{b0}) + \eta_b\left(x_b - x_{b0}\right)^3\right], \tag{2.4.6}$$

where k_u, k_l, and k_b are the linear spring constants, while η_u, η_l, and η_b are the upper, lower,

and body, non linear spring constants, respectively.

Forces due to dampers:

$$F_{du} = -d_u \left(\dot{x}_u - \dot{x}_b\right), \tag{2.4.7}$$

$$F_{dl} = -d_l \left(\dot{x}_l - \dot{x}_b\right), \tag{2.4.8}$$

$$F_{db} = -d_b \dot{x}_b, \tag{2.4.9}$$

$$\tag{2.4.10}$$

with the damping coefficients d_u, d_l, and d_b are calculated as follow,

$$d_u = 2\zeta_u \sqrt{m_u k_u}, \tag{2.4.11}$$

$$d_l = 2\zeta_l \sqrt{m_l k_l}, \tag{2.4.12}$$

$$d_b = 2\zeta_b \sqrt{m_b k_b}, \tag{2.4.13}$$

with ζ_u and ζ_l being damping ratios that variates when the masses are colliding, while the body mass has a fixed ratio ζ_b.

Forces due to cover coupling:

$$F_{kb} = -k_c \left[\left(x_l - x_{l0}\right) - \left(x_u - x_{u0}\right)\right] \tag{2.4.14}$$

with k_c the spring constant that links the cover together.

Forces due to collision:

$$F_{uCol} = -h_{uCol} \left[\left(x_u - x_{uCol}\right) + \eta_u \left(x_u - x_{uCol}\right)^3\right], \tag{2.4.15}$$

$$F_{lCol} = -h_{lCol} \left[\left(x_l - x_{lCol}\right) + \eta_l \left(x_l - x_{lCol}\right)^3\right], \tag{2.4.16}$$

where h_{uCol} and h_{lCol} are the contact spring coefficients, and x_{uCol} and x_{uCol} are the point of displacement where collision occurs.

The BCM considers a variable glottal area defined by the position of the masses, and given that the model is considered symmetric, the area can be defined by:

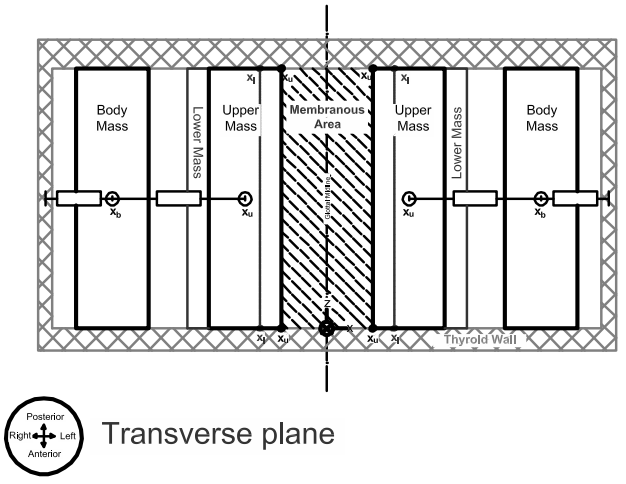
$$A_u = \max\left(0, x_u \ell_g\right), \tag{2.4.17}$$

$$A_l = \max\left(0, x_l \ell_g\right) \tag{2.4.18}$$

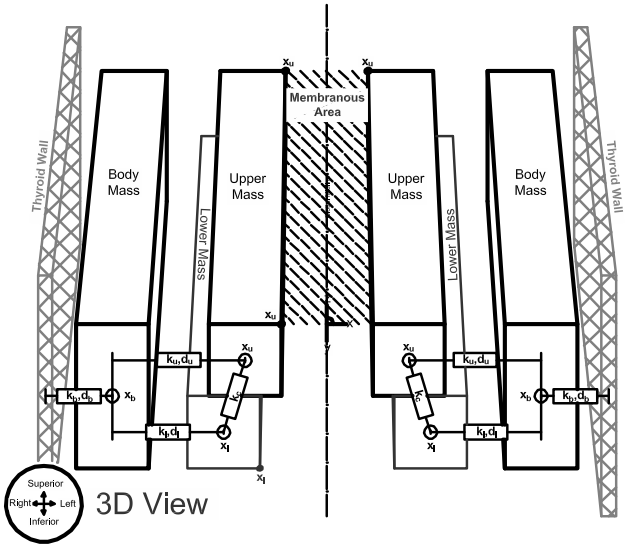
where A_u and A_l are the upper and lower areas, and ℓ_g is the effective length of the glottis.

The change of areas in the inferior-superior direction is the key component for the self-sustain oscillation of the model, since it allows for the flow to change its behavior from a Bernoulli channel, to a jet channel (See Figure 2.3), thus changing the driving pressure as well.

The restoring pressures obtained from the different configuration are presented in Table 2.2, where P_u is the pressure applied to the upper mass, and P_l is the pressure applied to the lower



(a)



(b)

Figure 2.2. Body Cover Model Structure. (a) Top View, (b) 3D view

mass. Thus, the force due to the pressure is obtained by the following equations,

$$F_{eu} = P_u \ell_g T_u, \tag{2.4.19}$$

$$F_{el} = P_l \ell_g T_l, \tag{2.4.20}$$

where T_u and T_l are the upper and lower thickness of the masses, that combined makes the total thickness of the VF.

The flow solution for the original BCM is a flow model dependent of the incident pressure and the minimum glottal area. The original solution of the model is given by:

$$Q_g = c\frac{A_g}{k_t}\left\{-\frac{A_g}{A^*} \pm \left[\left(\frac{A_g}{A^*}\right)^2 + k_t \frac{2}{c^2\rho}\left(p_s^+ - p_e^-\right)\right]^{1/2}\right\} \tag{2.4.21}$$

where A_g is the minimum area between A_u and A_l, k_t is a trans-glottal pressure factor defined by Scherer et al.,[134] c is the speed of sound, ρ is the air density, p_e^- and p_s^+ are the incident acoustic pressures into the glottal area produced by the upper and lower vocal tracts, and A^* is a trans-glottal area defined by

$$A^* = \left(A_s^{-1} + A_e^{-1}\right)^{-1} \tag{2.4.22}$$

where A_s and A_e are the sub-glottal and supra-glottal areas, respectively.

This equation has shown to have problems when the glottal area is close to zero, or when the acoustic interaction is strong.[96] A correction to this equations was made by Lucero and Schoentgen,[96] replacing the current flow solution to include viscous effects, producing the following flow model for the BCM

$$Q_g = \pm c\frac{A_g}{k_t}\left\{\left(-\frac{A_g}{A^*} + \frac{\gamma_m}{\rho c A_g^2}\right) + \left[\left(-\frac{A_g}{A^*} + \frac{\gamma_m}{\rho c A_g^2}\right)^2 + \frac{4k_t}{\rho c^2}\left|p_s^+ - p_e^-\right|\right]^{1/2}\right\}. \tag{2.4.23}$$

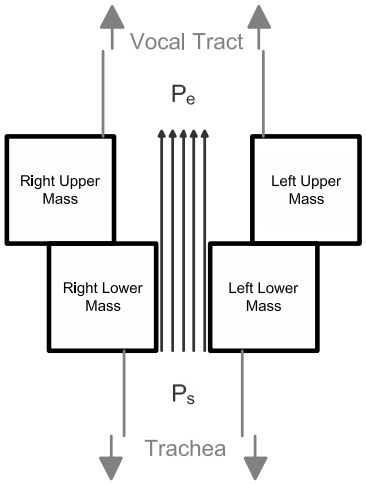
where γ_m is a viscous coefficient dependent of glottal characteristics of the onset condition.[96]

It is worth mentioning that this is not the only modification proposed for the BCM to correct the flow and pressure distribution, some corrections also include the Coanda effects,[42,158] modified contact forces,[113] and waveform alterations.[135]

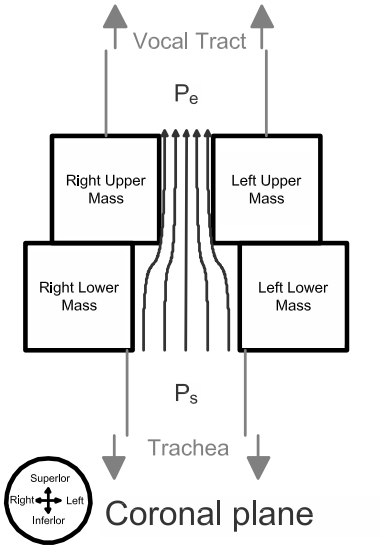
Aside from the flow inconsistencies in the original implementation of the BCM, there are additional complications associated with the selection of adequate parameters for the springs and dampers of the model. This problem is not unique to the BCM and is observed in many,

Table 2.2. Pressures components on each glottal configuration [P_e: pressure in the supra-glottal tract, P_s: pressure in the sub-glottal tract]

Pressure	Fully Closed	Upper Closed	Lower Closed	Convergent	Divergent
P_u	0	0	P_e	P_e	P_e
P_l	0	P_s	0	$P_s - (P_s - P_e)\left({}^{A_u}/_{A_l}\right)^2$	P_e



(a)



(b)

Figure 2.3. Flow channel patterns for the body cover model. (a) Divergent, (b) Convergent

if not all, numerical models of voice production. In the initial implementation of numerical models, some of the associated parameters where selected in an *ad hoc* fashion to reproduce certain outputs.[78] The two mass model[78] uses a series of numerical values to find the proper parameter configuration that makes phonation sustainable. This was also the case of the original implementation of the BCM,[154] where 4 sets of parameters where selected to imitate Hirano's four modes of phonation.[64] More recently, Steinecke and Herzel[146] used an ad-hoc bilateral parameter modification to emulate nerve paralysis. Lately, Peter Birkholz[9] heuristically selected parameter values to produce six different types of phonation. The obvious drawback of such approach is that physiological relevance is compromised.

A significant advance on this matter has been made by several different researchers,[7,170,173] by creating a set of physiological rules to adjust a model parameter according to the activation of intrinsic muscles of the larynx. further advances were introduced using transient time variation,[169] three dimensional estimation movement,[74] and a combination of active and passive strength of the muscles.[140] In addition, muscle activation parameters have also been used on inverse problems,[16,56] and to further investigate causes of hyperfunction.[50,192] The simplest, yet effective scheme that combines BCM and activation parameters is presented by Titze and Story.[173] Here, a set of rules from the physiology of the VF were applied to obtain the biomechanical tissue properties of the model. The activation of the CT, TA, and LCA muscles (a_{CT}, a_{TA}, and a_{LC} parameters respectively) are used to control all the mechanical properties of the BCM such as spring constants, VF length, onset conditions, etc. As an outline summary, the rules include the following:

- **Elongation rule:** Estimates the vocal fold length from the resting length (ℓ_0), and the combined activation of the CT, TA, and LC muscles, obtaining,

$$\ell_g = \ell_0 \left[1 + G\left(a_{CT}R - a_{TA}\right) - a_{LC}H\right]. \tag{2.4.24}$$

  The elongation gain (G), the adductor strain factor (H), and the torque ratio (R), are empirically obtained values defined as $G = 0.2$, $H = 0.2$, and $R = 3.0$ for human phonation.

- **Nodal point rule:** Establishes a relation of the existing difference between the upper and lower edges of the VF due to muscular activation. This phenomena produces different amplitudes in the upper and lower edges that translate to the model as a nodal point of vibration given by

$$z_n = \left(1 + a_{TA}\right)\frac{T_g}{3}, \tag{2.4.25}$$

  where T is the effective thickness of the VF

- **Thickness rule:** The thickness of the VF is obtained solely from the passive force applied, thus obtaining the following

$$T_g = \frac{T_0}{1 + 0.8\left(G\left(a_{CT}R - a_{TA}\right) - a_{LC}H\right)}, \tag{2.4.26}$$

  which is the ratio between the vibrating thickness at resting length (T_0) and the longitudinal

VF strain used to calculate the VF length.

- **Depth rule:** The depth of the body and cover layers are obtained from

$$D_{body} = \frac{a_{TA} D_{mus} + 0.5 D_{lig}}{1 + 0.2\left(G\left(a_{CT} R - a_{TA}\right) - a_{LC} H\right)}, \tag{2.4.27}$$

$$D_{cover} = \frac{D_{muc} + 0.5 D_{lig}}{1 + 0.2\left(G\left(a_{CT} R - a_{TA}\right) - a_{LC} H\right)}, \tag{2.4.28}$$

  where D_{mus}, D_{muc}, and D_{lig} are the muscular, mucosa, and ligament depths, respectively. These values are obtained from average clinical observations, and can particularly suffer from generic modeling fallacy (Average parameters do not necessary lead to average results).[23]

- **Adduction rule:** The adduction rule establishes the degree of separation of the VF with respect to the midline position. This has been obtained from fibroscopic imaging, and is represented by the following equation:

$$\xi_{02} = 0.25 \ell_0 \left(1 - 2 a_{LC}\right), \tag{2.4.29}$$

  where ξ_{02} is half separation of the upper edge of the VF.

- **Convergence rule:** The convergence rule produces the separation of the lower portion of the VF by the relation $\xi_{01} = \xi_c + \xi_{02}$, where ξ_{01} is the lower edge of the VF, and ξ_c is given by

$$\xi_c = T_g \left(0.05 - 0.15 a_{TA}\right) \tag{2.4.30}$$

These rules are then used to translate muscular activation values (that range between 0 and 1 for a_{CT} and a_{TA}, and -1 to 1 for a_{LC}), into model parameters $(\ell_g, T_u, T_l, m_u, m_l, m_b, k_u, k_l, k_b, k_c, x_{u0}, x_{l0})$. For further details, please refer to Titze and Story 2002.[173]

2.5 Acoustic propagation

As mentioned in the physiology section of this chapter, the sound generated in the glottis is modulated (filtered) by the vocal tract, mouth and nose before it is radiated into the air. The filtering process occurs due to the propagation and reflection across the areas in the trachea, larynx and pharynx. The tissue properties in the tract play an important role in the attenuation of the sound waves, and the interaction produced by surrounding sound waves in the larynx has been shown to be a key part of the VF oscillation.[171] In addition, it has been shown that the sub-glottal tract propagation produce important effects in phonation, introducing resonances[20, 59, 148] and straining, in some cases, vocal folds vibration.[166, 195, 199] Thus, accounting for sound propagation its critical for the development of this thesis.

The glottal source of sound can be characterized by a dipole source.[98, 201] This source is caused by the pulsatile flow that is produced due to the self-sustained VF oscillation, thus producing (at the glottis) inverted in phase sound-waves that travel up and downstream in the

vocal tract.[105] Aeroacoustic analysis establishes that noise components produced by narrow areas can induce perturbations of the flow thath are generally an order of magnitude smaller than the dipole source, but that can still be perceived on voiced speech.[111,201] These noise components can be produced either during the final closing stages of normal phonation, or due to the incomplete glottal closure caused by arytenoid posturing, onset conditions, or pathological anatomical formations (e.g., nodules and polyps).[192] Furthermore, the ventricular folds can eventually produce an additional dipole component[198] that could eventually be the only source of sound in certain pathological conditions.

There are two techniques used to model the sound propagation waves across the vocal tract: (1) WRA,[85,90,151] and (2) TLS.[47,48,148,149,181] Both methods use a one-dimensional planar wave representation of the sound wave pressure, which has been shown to be sufficiently to accurate for voice models.[45,151] The wave reflection analog (WRA) method uses a planar wave representation, where the solution of forward and backward pressures is solved in a time domain framework. Therefore, the total pressure in each section of the vocal tract is the superposition of both forward and backward traveling acoustic pressures. On the other hand, the transmission line scheme (TLS) is a representation of the tract section as an electrical T-section analog to the acoustic components present in that section. This analogy between electrical components and acoustic elements allows for the inclusion of complex scenarios that are more difficult to describe in the WRA, such as cavities and branches of the supra and sub-glottal tract, time variating shapes of vocal tract (e.g., running speech), and frequency dependent losses.[49,66,78,104] However, even when the TLS allows for better representation of the vocal tract, the WRA is still broadly used due to its simplicity and sufficiently good results on sustained vowels.[44] In this thesis, a small modification of the WRA representation is proposed to improve the computational time required to process all the tract pressures. A detailed description of WRA is provided in Chapter 3.

To obtain the values of the tract areas used in each method, the three dimensional sub and supra-glottal tracts are discretized into small concatenated tube representations. these tract area functions are typically obtained through MRI.[152,155,157]

A important element that relates the sound propagation with the dipole sound source is the acoustic interaction. This concept establishes the relationship between the propagated wave pressure and the intra-glottal pressures. In the linear source filter theory[48] the vocal tract and the glottal source were assumed to be independent elements, thus neglecting the effects that arise from the surronding sound waves on the airflow and tissue dynamics.[47] Subsequent studies revealed important effects regarding the interaction between these elements, thus leading to a non-linear source-filter theory of phonation.[49,57,77,87,120,122,123] Thanks to advances in magnetic resonance imaging (MRI),[69,152,155,157] it was shown that the vocal tract impedance is actually comparable to the glottal impedance, thus implying a stronger interaction than the initially stated.[165,166,172]

The non-linear source-filter coupling theory of phonation[166] establish a relation between the impedance of the tract and the glottis. at the same time, it allows to apply this knowledge of acoustic interaction of the VF to the numerical models of phonation. Two types of interactions can be described in this sense: “level 1”, where the dynamic pressures of the acoustical waves

affects the flow but not directly affect the kinematic behavior of the VF, and "level 2", where the incident acoustic pressures affect both flow and VF kinematic movement. Extensive research over this theory has been done in order to quantify the effect of different degrees of interaction, encountering variations on the phonation threshold pressure, instabilities, sound pressure radiation variances, bifurcations, aphonation, sub harmonics, and frequency jumps; all phenomena commonly seen in normal and pathological speech.[2,171,177,194]

2.6 System identification for speech production

Any modeling of physical systems has associated an inevitable loss of information due to simplifications, hidden information, contaminated information, etc. Accounting for these procedures in modeling is an important issue. In that regard, the common practice of averaging information for clinical and biomechanical analysis, has been shown to be insufficient when using model outputs for population analysis.[23] This inherent misinterpretation of the data is known as the "generic modeling fallacy", and it states that if a model is generated using averaged biomechanical parameters, the corresponding output is not necessarily an average result. In consequence, even when the model has high physical meaning, the usage of average parameter values wont necessarily reflect average behavior. To correct for this situation, the problem could be addressed from a different perspective, where the input parameters are "adequate" to produce average results. Thus, to address such problems, a system inverse analysis can be used.

In system theory, two different approaches can be used to model a determinate system: (1) analytic (or physic-based) modeling, and (2) parametric (or data-based) modeling. The analytical modeling uses information on the physical structure of the system to design a set of rules between parameters, inputs, and outputs. This method limits the behavior of numerically equivalent elements according to the analytic information gathered from the physical world. Most of the VF models fit under this category. Parametric modeling, on the other hand, uses a generic description that is not directly related with the physical world, and no specific physic-based information is extracted from the system. These types of models use generic mathematical formulations to describe the system output behavior, e.g., auto regresive moving average (ARMA) models,[21] linear predictive coding (LPC)[110] models, etc. To produce both types of models, it is possible to observe the desired output and estimate the model parameters to reproduce such observations. This process is commonly known as "system identification" since it aims to obtain the filter information from the observed signal. Even though the computational complexity of parametric models is lower than the physic-based models, its performance is highly dependent on the model structure identification, which is not straightforward. Given the physiological characteristics desired in the VF model, this thesis will use only analytic modeling based on physiology and physics of the voice production system.

To understand the system identification process, it becomes necessary to characterize "systems" as a standard concept. In this research, we use *system* to define an organized set of devices, rules, or items, that are capable of interacting and restructuring certain inputs, transforming them into modified outputs (see Figure 2.4(a)). Under this idea, the physiology and anatomy of the VF can be divided in several levels of abstraction as shown in Figure 2.4(b).

It has been previously stated that VF modeling has a vast set of options for several components.[11,30,44] Furthermore, even when only one model is selected, there is no guarantee that the parameters obtained from general population will accurately represent the vocal fold behavior.[23]

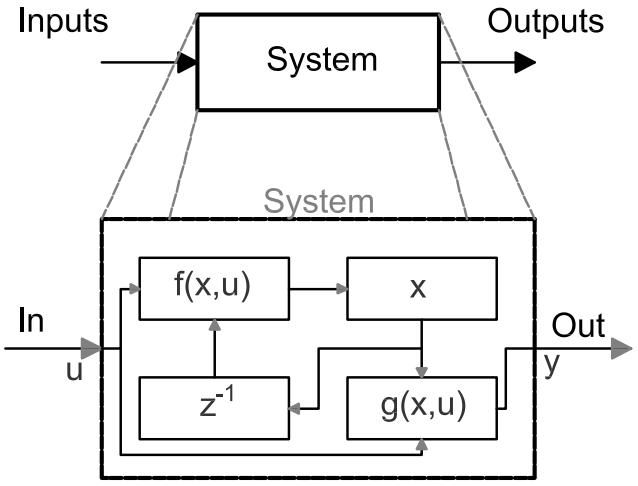
To improve the relation between the parameter estimation and the observed behavior, we could use outputs of the system to identify the parameters and inputs, of the model. This process is usually called "inverse problem", and it can be represented as the inverse function $u = f^{-1}(y)$ of the model $f(u)$ that relates the input u with the output y.

There are several identification schemes that have been utilized in VF models. Most of these methods assume that the system behaves deterministically in both model structure and control parameters. For instance, Tokuda and Herzel[176] used a non-linear two mass model (TMM)[78,146] of VF which was fit to a time series to extract the asymmetry factor of the VF. Similarly, unilateral vocal fold paralysis estimation was performed by Schwarz et al.,[137] using over 30 high speed videoendoscopy (HSV) recording of healthy and pathological subjects. Chaos synchronization was performed by Zhang et al.[197] to obtain asymmetry parameters using the TMM of Ishizaka and Flanagan.[78,146] Wurzbacher et al.[184] also used HSV to classify non-stationary vocal fold vibration in a TMM, using Adaptive Simulated Annealing optimization.[76] Tao et al.[161] used genetic algorithm[?] to directly obtain the model parameters such as masses, spring constants, dampers, etc. Schwarz et al. used a multi-mass model[182] to find spatio-temporal quantifications of vocal fold vibration,[136] using also a genetic algorithmic approach.[72] Yang et al.[185] used an *in vitro* VF recording as a target for the numerical optimization[187] of a three dimensional multi-mass model[186] to obtain physiologically relevant biomechanical features.

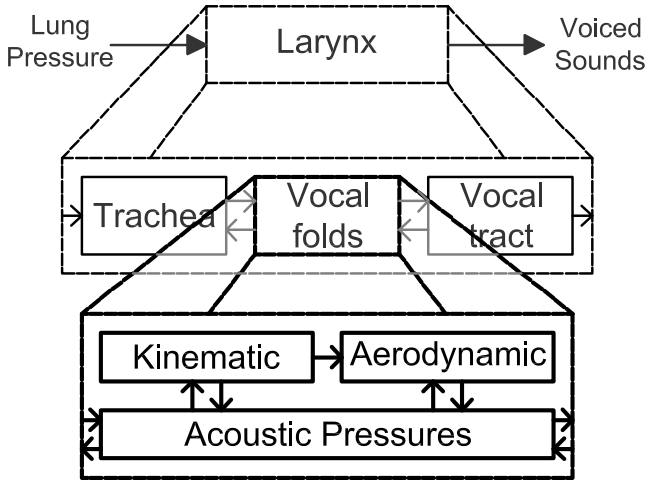
Almost all the work performed in the last ten years has been focused on finding deterministic parameters of VF models using an inverse problem approach. Significantly improving previous efforts of modeling, shifting the model paradigm from a causal modeling approach to non-causal modeling, which means the explanation of the model based on the resulting observations. This system identification approach reduce the error produced by averaging modeling,[23] allowing at the same time the application of parameter-based classification,[184] which can lead to improved vocal assessment.

The VF systems are characterized by a series of non-linear relations which could eventually produce multiple local minima, which can lead to a biased estimation of the underlying parameters. In addition, the uncertainty of the estimation process is not well quantified in the deterministic modeling of VF. The source of these uncertainties include problems like inaccurate measurements, inherent randomness of the system (e.g., noise produced by a posterior glottal opening), modeling assumptions, etc.

One solution to the deterministic modeling problem is to consider the parameters that rule the model behavior as random variable (rv). This means, that some model parameters do not have a given value in particular, but rather a probability of being in a determinate value-range. This mathematic characteristic is known as the probability density function (pdf) of a rv, and provides the ability to include credibility intervals, and other probabilistic descriptions for selected model parameters. Describing the model parameters as rv is a major change in the modeling of voice production, since it could better represent biological phenomena[35] than the deterministic approach. This stochastic approach has been considered by Cataldo et al.[16] using parameters pdf



(a)



(b)

Figure 2.4. System block representation. (a) Generic block system, and inner view of a state space arrangement, (b) Possible voice production system decomposition block

on the inverse filtering of a michrophone signal, with a stationary Bayesian estimation technique. In addition, Mehta et al.,[102] used a Kalman filter to track formant and anti-formant in an ARMA model, while Sahoo and Routray[127] used an extended Kalman filter to identify the glottal flow from the radiated sound pressure.

In this thesis we explore the use of Bayesian estimation techniques to obtain model parameters, thus in the following section, an overview of stochastic system identification is presented, with a focus on Bayesian estimation theory. These concepts will be later used in the development of a subject-specific model in Chapter 4.

2.7 Bayesian estimation theory

Bayesian signal processing aims to estimate the underlying probability density function of a random signal in order to provide statistical inference capabilities.[38] Therefore, assuming an unknown rv X that produces the corresponding noisy data Y, the estimation process should obtain $\mathbf{Pr}(X|Y)$, which is read as the probability of having the rv X given the observed variable Y. This process can be stated as[13]

$$\hat{X} = \arg\max_{X} \{\mathbf{Pr}(X|Y)\} \tag{2.7.1}$$

where $\hat{X}$ is the estimate of X. However, the Bayes rules states the following,

$$\mathbf{Pr}(X|Y) = \frac{\mathbf{Pr}(Y|X)}{\mathbf{Pr}(Y)}\mathbf{Pr}(X) \tag{2.7.2}$$

which basically says that the probability of having the rv X (with realization x) given the observed variable Y can be obtained by the previous information of X, noted as $\mathbf{Pr}(X)$, updated by the likelihood of Y given the knowledge of X. In this form, the term $\mathbf{Pr}(X)$ is simply referred to as *prior*, $\mathbf{Pr}(X|Y)$ is *posterior*, $\mathbf{Pr}(Y|X)$ is *likelihood*, and $\mathbf{Pr}(Y)$ is *evidence*. Thus, if we know the system stochastic output Y and the process that generated that output, we could estimate the rv X using Bayes theorem. When applied to the voice models, this could be interpreted as obtaining the flow signal (X) from the radiated pressure (Y), knowing the vocal tract used to produce that output $\mathbf{Pr}(Y|X)$.[127] Note that for this formulation no linearity conditions were imposed to the likelihood, and no restrictions were given to the distributions of the priors or the posteriors, thus not limiting the solution to only Gaussian distributions.

A particularly useful numerical representation of models are the state space model (SSM). These numerical representations are composed, as before, from a series of inputs and outputs, but they also define a *state*, which is defined as the info about the system in absence of external input. Therefore, the SSM is defines as a collection of variables that are mutually independent, and can sufficiently specify the dynamic system behavior.

The general formulation of a deterministic, non-linear, continuous-time SSM, can be expressed

as[13]

$$\dot{x}_t = \mathcal{A}\left(x_t, u_t\right), \tag{2.7.3}$$

$$y_t = \mathcal{C}\left(x_t, u_t\right), \tag{2.7.4}$$

where u_t, y_t, and x_t are the system input, output and state variables, respectively; $\mathcal{A}\left(\cdot\right)$ and $\mathcal{C}\left(\cdot\right)$ are nonlinear state and observation functions.

For a complete linear system, the formulation translates to the following,

$$\dot{x}_t = \mathbf{A}_t x_t + \mathbf{B}_t u_t, \tag{2.7.5}$$

$$y_t = \mathbf{C}_t x_t + \mathbf{D}_t u_t, \tag{2.7.6}$$

with $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$, matrices that define the relationships between the states, inputs, and outputs.

From the physical point of view, the state-space representation allows for the same numerical representation to capture dynamic behaviors of several different physical systems. Since different physical systems can be described in flow-equivalent representations of energy transmission.[112] For example, the same state-space model structure can be used to describe electrical and mechanical systems. Furthermore, the representation of physical systems in state-space models has been further developed in Port-Hamiltonian systems, allowing for representing interconnection of system and energy transmission dynamics in a more "naturally driven" form.[178] This approach has also been used to explore energy transfer in VF models.[39]

The state space model can be further developed in a discrete domain, using for this purpose a Markov chain representation. That is, each state depends only on the state attained in the previous event. Therefore, the evolution of the Markov chain state space representation is

$$x_{k+1} = \mathcal{A}\left(x_k, u_k\right), \tag{2.7.7}$$

$$y_k = \mathcal{C}\left(x_k, u_k\right), \tag{2.7.8}$$

where the initial condition x_0 is defined *a priori*. The subindex k denotes the sample corresponding to $t = T_s k$, for a uniformly sampled signal with sampling time period of T_s and $k \geq 0$. It is important to note that the state and observation functions $\mathcal{A}\left(\cdot\right)$, $\mathcal{B}\left(\cdot\right)$, do not necessarily match the formulation of the continuous time function of the same nomenclature.[189]

To incorporate uncertainty in the system, the state-space models must incorporate randomness in their formulation, either on the inputs, outputs, initial conditions, or a combination of all three. One of the most popular implementations of the uncertainty on state-space models are the Gauss-Markov models,[13] which for a discrete linear system is

$$x_{k+1} = \mathbf{A}_k x_k + \mathbf{B}_k u_k + \mathbf{W}_k \omega_k, \tag{2.7.9}$$

$$y_k = \mathbf{C}_k x_k + \mathbf{D}_k u_k + \nu_k, \tag{2.7.10}$$

where ω and ν are process and observation Gaussian noise, respectively, with initial state x_0

normally distributed. The same system in non-linear form can be represented as follows

$$x_k = \mathcal{A}(x_{k-1}, u_{k-1}, \omega_{k-1}), \tag{2.7.11}$$

$$y_k = \mathcal{C}(x_k, u_k, \nu_k), \tag{2.7.12}$$

which in a Bayesian form can be represented as[13]

$$\mathbf{Pr}\left(x_k|x_{k-1}\right) = \mathcal{A}\left(x_k|x_{k-1}, u_{k-1}, \omega_{k-1}\right) \tag{2.7.13}$$

$$\mathbf{Pr}\left(y_k|x_{k-1}\right) = \mathcal{C}\left(y_k|x_{k-1}, u_k, \nu_k\right) \tag{2.7.14}$$

where the propagation of the state is given by the probability $\mathcal{A}\left(x_k|x_{k-1}, u_{k-1}, \omega_{k-1}\right)$ and the likelihood distribution is obtained from $\mathcal{C}\left(y_k|x_{k-1}, u_k, \nu_k\right)$. This expression evidences the phenomena of the "unobserved state" at step $k-1$, since it is only propagated to the measurement likelihood at step k.

The Bayesian framework basically states that:

> **Given** a set of noisy uncertainty measurements $y_{0\ldots k}$, and input set $u_{0\ldots k}$, a prior distribution for the state $\mathbf{Pr}\left(x_0\right)$, process noise $\mathbf{Pr}\left(\omega_{k-1}\right)$, and measurement noise $\mathbf{Pr}\left(\nu_k\right)$ as well as the conditional transition probability distributions $\mathbf{Pr}\left(x_k|x_{k-1}\right)$, and likelihood probability distributions $\mathbf{Pr}\left(y_k|x_{k-1}\right)$, characterized by the state model $\mathcal{A}\left(x_k|x_{k-1}\right)$ and measurement model $\mathcal{C}\left(y_k|x_{k-1}\right)$, **find** the 'best' estimate of the filtering *posterior* $\hat{\mathbf{Pr}}\left(x_k|y_k\right)$, and is associated statistics.[13]

This approach can be further developed to obtain analytical solutions for linear-Gaussian models (Kalman filter), and non-linear Gaussian models (Extended Kalman filter and Unscented Kalman filter). For non-linear non-Gaussian models, there is no analytical solution that fits all model types, and commonly, other numerical approximations are used (e.g., particle filtering).

The VF system, and the voice production process is, by its nature, non-linear, time-varying, and stochastic. Thus, it becomes a challenge to completely identify all the parts involved in the mathematical modeling of voice production using the framework previously described. However, given the extensive research on biological properties of the VF, vocal tract, and phonatory process, many parts of the modeling structure and physiology are already sufficiently well defined and the problem can be simplified. Given the tools provided by Bayesian estimation, there is a clear possibility of estimating the parameters that control specific behaviors from clinical observations and mathematical modeling.

Chapter 3

PROPOSED VOICE PRODUCTION MODEL

Chapter 2 outlines phonation theory, where state of the art and basic background knowledge were discussed to provide the foundation for subsequent chapters. In order to properly address the system identification scheme described in Chapter 4, a few adjustments in the underlying model must be made to improve the physiological descriptions and to reduce the computational load. Therefore, this chapter introduces a theoretical description of numerical models of speech production that tackle these concerns. To achieve such improvements, structural modifications are performed on the body-cover model (BCM),[154] by adding incomplete glottal closure controlled by arytenoid posturing, yielding membranous and posterior glottal openings, with a triangular shaped glottis. In addition, a discretization of differential equations of vocal fold lumped mass model, allows for a significant reduction in computational time with a sufficiently low error. Furthermore, a state space representation of the vocal tract reduces even more the computational time used in each step of the simulation, when compared with the distributed implementation of the wave reflection analog (WRA).[90] At the end, this chapter concludes with a proposed model which will be used in the subsequent chapters as a base for simulations and estimations.

3.1 Body cover model with posterior glottal opening

Self-sustain lumped element models of voice production have been proven to provide information that is otherwise hidden from clinical observations, thus improving the insights of normal and pathological phonation.[44,192] However, even when important advances have been made to improve the physiological relevance of numerical models,[154,173] only limited efforts have been made to account for the effects of incomplete glottal closure.[10,192] The BCM, described in Chapter 2, provides several benefits over other types of numerical models,[44] including simplified representations of the kinematic behavior, differentiation of the body and cover layers, and low computational load. In addition, its three way interaction between airflow, tissue, and sound[44,163] provides acoustic feedback to accurately describe non-linear interaction.[166,194] However, the BCM lacks the ability to represent normal and pathological cases with presence of a posterior glottal opening (PGO).

Incomplete glottal closure is a common condition affecting both normal and pathological subjects,[124] with many pathologies (e.g., nodules, polyps, muscle tension dysphonia (MTD))

presenting incomplete glottal closure due to a PGO.[32,67,81,89,107,128,130,131] Some authors relate PGO with common laryngeal configurations in women[57,68,91] and children,[153] possibly leading to a cue for the clinical prevalence of hyperfunctional behavior on women.[128] In addition, evidence show a direct relationship between incomplete glottal closure in onset conditions, and additional stress and efforts to produce perceptual normal phonation,[50,192] thus providing a link between incomplete glottal closure, MTD, and hyperfunctional voices.

Some effort has been made to quantify the effects of incomplete closure of the vocal folds (VF) by simulating an imposed waveform in parametric models,[28,29] and in two mass model.[27] A more recent study proposed a triangularly shaped posterior glottal channel[133] to quantify AC and DC flow. Other studies in low-dimensional models, have explored this problem by incorporating pre-contact changes in stiffness,[113] restricted vibration of the VF,[99] nonlinear damping coefficients,[94] and triangularly shaped glottis.[10] These different approaches have been used to: mimic polyps and nodules,[89,113] to better reproduce the abduction phenomena in consonant-vowel-consonant gestures,[95,99] and to synthesize voice.[9] Despite these efforts, the study of PGO remains largely neglected when using full flow-sound-tissue interaction[166] with self-sustained numerical models, which has proved to be an important feature of phonation.[121]

The simplicity of the BCM, along with its direct structure-physiology relation, its low computational cost, and its relatively accurate dynamic representation, makes it an ideal candidate to be used in the context oa a subject-specific estimation that is a complex and computationally demanding process. However, before approaching the estimation process, the BCM requires some additions, such as, the inclusion of incomplete glottal closure. A comprehensive analysis of a parallel glottal channel has been performed, as part this thesis work and it was reported in Zañartu et al.,[192] where we found that the inclusion of a separate glottal flow channel can account for the lack of a PGO in the original BCM. In this section we sumarize and provide additional insights on the theoretical development of a separated flow channel added to the BCM, with the aim of obtaining a more physiologically relevant voice production model.

As shown in Chapter 2, the symmetric BCM is composed of three masses in each vocal fold, organized in two separated layers of the body and cover structure. The lumped masses are linked together by a structure of springs and dampers that characterize the viscosity and stiffness of tissue properties (see Figure 2.2).[154] The strictly parallel glottal edge configuration in the BCM (See Figure 3.1) implies that it only accurately describe the membranous glottal area but not the anterior-posterior edge dynamics.

The vocal fold posturing associated to arytenoid configuration can produce a PGO with or without necessarily producing incomplete closure in the membranous portion of the glottis.[107,133] A PGO can be caused, among other things, by the abnormal function of the posterior cricoarytenoid muscle, by VF paresis, or by the presence of nodules and polyps. In this study, the inclusion of PGO will only consider the abnormal posturing due to the cricoarytenoid muscle behavior, referred by Morrison as Type I of MTD.[107] This produces a rotation of the arytenoid cartilages leading to a separation of the cricoarytenoid joint, but a closure of the membranous glottal attachment (also referred as vocal process[164]). The resulting configuration yields a non-vibratory separate glottal channel that can be modeled independently from the membranous glottal channel.[107,133,192] (see Figure 3.2)

The basic assumption on the independent flow channel approach is that both membranous and posterior channels are part of a unique control volume, where the air-flow recombines at some point downstream in the supra-glottal area. In addition, the supra-glottal area is assumed sufficiently large for total recombination with a uniform flow pressure field (see Figure 3.3), where the air that passes through the channels is considered incompressible and inviscid, neglecting the losses happening inside the channels. While these initial assumptions of incompressibility and recombination are not completely accurate, they do not differ significantly from the original assumptions on the BCM,[154] and will be later improved by small adjustments.[96]

We start by considering flow velocity of membranous glottal channel (v_m) and flow velocity in posterior glottal channel (v_p) equal at the input sections of the control volume, then, conservation of mass on a steady state regime yields an exit flow velocity (v_e) of

$$v_e = v_s \frac{(A_m + A_p)}{A_e}, \tag{3.1.1}$$

where v_s is the flow velocity of the sub-glottal tract, A_e is the sectional area of the immediate supra-glottal tract, and A_m and A_p are the areas of the membranous and posterior channels respectively.

From conservation of linear momentum, the pressure differential from the supra-glottal tract section to the upper exit plane of the glottal area is,

$$(P_e - P_d) = \frac{1}{2}\rho \frac{Q_g \left|Q_g\right|}{(A_m + A_p)^2} \left[2\frac{A_m + A_p}{A_e}\left(1 - \frac{A_m + A_p}{A_e}\right)\right], \tag{3.1.2}$$

where ρ is the air density, P_e is the supra-glottal pressure, P_d is the pressure at the exit glottal plane, and Q_g is the total volumetric glottal flow.

Given that the volumetric glottal flow must be the same on the inferior and superior portions of the glottis, the following relationship can be drawn

$$P_d = P_s - \frac{1}{2}\rho \frac{Q_g \left|Q_g\right|}{(A_m + A_p)^2}, \tag{3.1.3}$$

where P_s is the total pressure in the sub-glottal section.

Finally, the trans-glottal pressure drop can be expressed as

$$(P_s - P_e) = \frac{\rho k_t}{2} \frac{Q_g \left|Q_g\right|}{(A_m + A_p)^2}, \tag{3.1.4}$$

where k_t is a trans-glottal kinetics loss coefficient defined by[192]

$$k_t = \left(1 - \frac{A_g}{A_e}\right)^2 + \left(\frac{A_g}{A_e}\right)^2, \tag{3.1.5}$$

with total glottal area of $A_g = A_m + A_p$.

to fit this development in te context of WRA,[90,154] both sub and supra-glottal pressure can

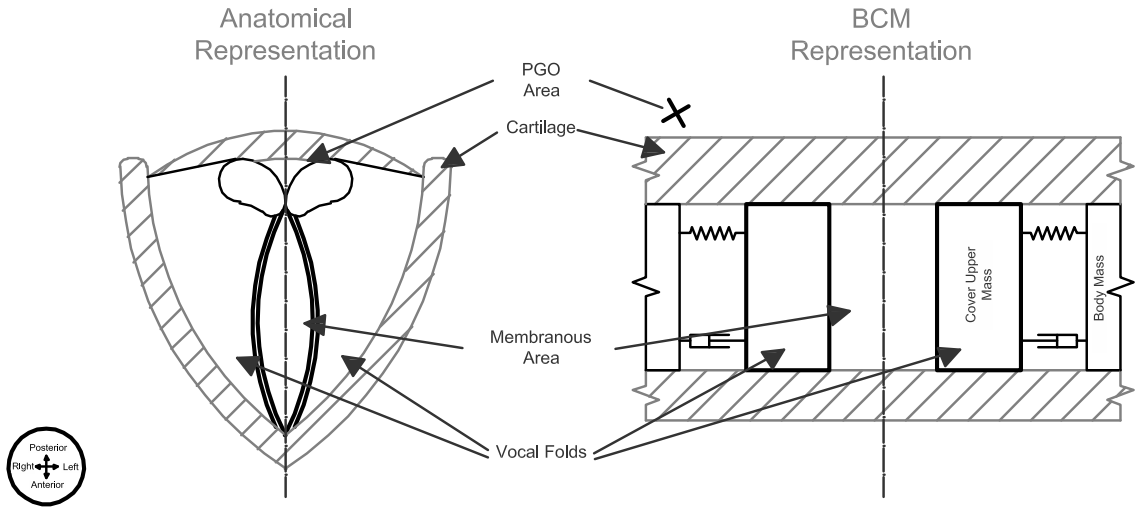


Figure 3.1. Comparison between the anatomy of the vocal folds and the structure of the symmetric BCM.

be expressed as the composition of incident and departing wave pressures,[90, 167] therefore,

$$P_s = p_s^+ + p_s^-, \tag{3.1.6}$$

$$P_e = p_e^+ + p_e^-, \tag{3.1.7}$$

where p_s^+, p_e^-, and p_e^+, p_s^-, are respectively the incident and departing wave pressures of the sub and supra-glottal tracts sections that are adjacent to the glottis. The departing waves can also be expressed as functions of the glottal flow by assuming a dipole source configuration, such that,[163, 175]

$$p_s^- = r_s p_s^+ - \rho c \frac{Q_g}{A_s}, \tag{3.1.8}$$

$$p_e^+ = r_e p_e^- + \rho c \frac{Q_g}{A_e}, \tag{3.1.9}$$

where c is the speed of sound, r_s and r_e are sub and supra-glottal reflection coefficients,[175] and A_s and A_e are the sub and supra-glottal areas, respectively. The reflection coefficients are given by

$$r_e = \frac{A_e - A_g}{A_e + A_g}, \tag{3.1.10}$$

$$r_s = \frac{A_s - A_g}{A_s + A_g}, \tag{3.1.11}$$

$$\tag{3.1.12}$$

and in original implementations of BCM are considered as $r_s \approx r_e \approx 1$.[163] However, the influence of reflection coefficients is later considered an important factor in the calculation of glottal flow, particularly when viscous behavior is applied.[96, 175]

Substituting the decomposition of the pressures in equation 3.1.4, and solving for the quadratic equation under[96] non viscous assumptions, the glottal flow can be expressed as

$$Q_g = \pm c \frac{A_g}{k_t} \left\{ -\frac{A_g}{A^*} + \left[\left(\frac{A_g}{A^*} \right)^2 + k_t \frac{2}{c^2 \rho} \left| \delta_P \right| \right]^{1/2} \right\}, \tag{3.1.13}$$

where δ_P is a trans-glottal pressure differential defined by $\delta_P = (1 + r_s)\, p_s^+ - (1 + r_e)\, p_e^-$, and A^* is an effective vocal tract area defined by

$$\frac{1}{A^*} = \frac{1}{A_e} + \frac{1}{A_s}, \tag{3.1.14}$$

The plus sign in Equation 3.1.13 is used when flow passes from the sub-glottal tract to the supra-glottal tract, meaning $\delta_P \geq 0$; The minus sign is used when a reversed flow is produced, meaning $\delta_P < 0$.

There are a few remarks on equation 3.1.13: (1) The stated solution for the flow is a corrected version of the original BCM, which had numerical issues when glottal pressures produced a reversed flow, (2) the glottal area with the proposed PGO addition is equal to the original BCM

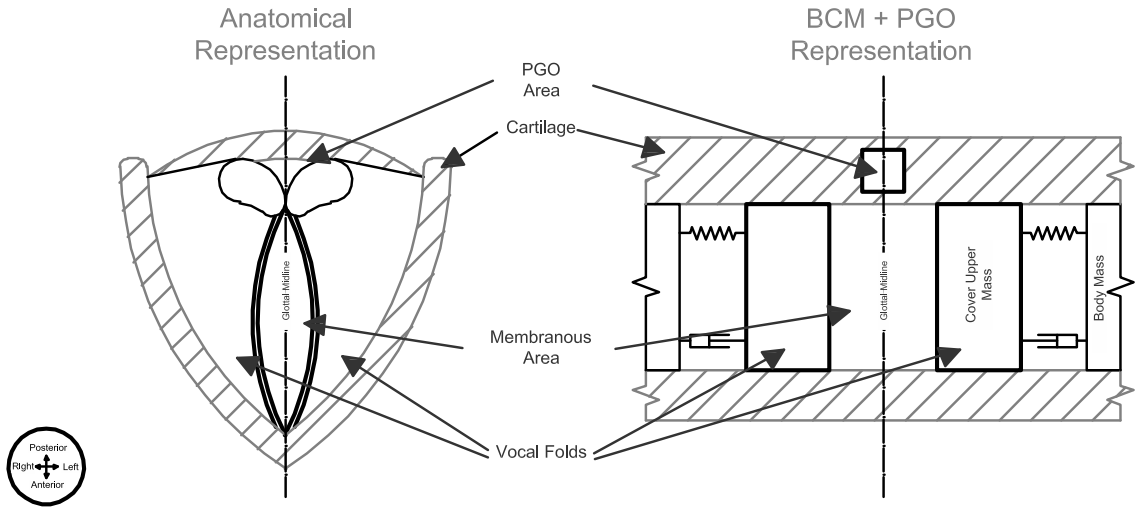


Figure 3.2. Comparison between the anatomy of the vocal folds and the structure of the symmetric BCM with posterior glottal opening inclusion.

area with no presence of PGO, and (3) the trans-glottal pressure differential is the same as the original BCM when full reflection of the tract coupling is considered.[154, 175] Therefore, the flow equation stated in 3.1.13 can be considered as an extension of the original corrected BCM flow equation expressed by Lucero and Schoentgen.[96]

The flow equation 3.1.13, while broadly accepted, has been proved to be non-differentiable on glottal closure when no PGO is present.[96] This non-differentiability comes from the full glottal closure, and can present problems in voice synthesis and voice analysis, where spectral characteristics are important, particularly affecting voice perception and related measurements such as the maximum flow declination rate (MFDR). To avoid these problems (in a no PGO scenario) Lucero and Schoentgen implemented a viscous pressure loss based on the Poiseuille formula for 2 parallel plates,[96] which is

$$\delta_{P_m} = \frac{12\mu\ell_g^2 T_g}{A_m^3} Q_m = \frac{\gamma_m}{A_m^3} Q_m, \tag{3.1.15}$$

where δ_{P_m} is the pressure loss between the parallel plates, μ is the air viscosity, ℓ_g is the length of the glottal channel, and T_g the thickness of the VF in the inferior-superior direction. Manipulating the expression, the viscous pressure drop can be also be presented as a resistive element to the volumetric air flow,

$$\delta_{P_m} = R_m Q_m = \frac{\gamma_m}{A_m^3} Q_m, \tag{3.1.16}$$

where R_m is the membranous glottal flow resistance due to viscous losses.

The posterior glottal channel, which is not included in the membranous glottal area, produces an alteration of the glottal resistance, therefore it forces a reformulation of the glottal pressure drop due to viscous effects, obtaining,

$$\delta_{P_g} = R_g Q_g, \tag{3.1.17}$$

where R_g, defined by Equation 3.1.18, is the equivalent glottal flow resistance of both membranous and posterior glottal channels.

$$R_g^{-1} = R_m^{-1} + R_p^{-1} \tag{3.1.18}$$

The total glottal resistance, in this configuration, is composed by the parallel arrangement of the membranous flow resistance (R_m) and the posterior flow resistance (R_p), which are defined by,

$$R_m = \frac{12\mu\ell_g^2 T_g}{A_m^3} = \frac{\gamma_m}{A_m^3}, \tag{3.1.19}$$

$$R_p = \frac{8\mu T_g \pi}{A_p^2} = \frac{\gamma_p}{A_p^2}, \tag{3.1.20}$$

where γ_m and γ_p are constant factors related to the shape and dimensions of the channels.

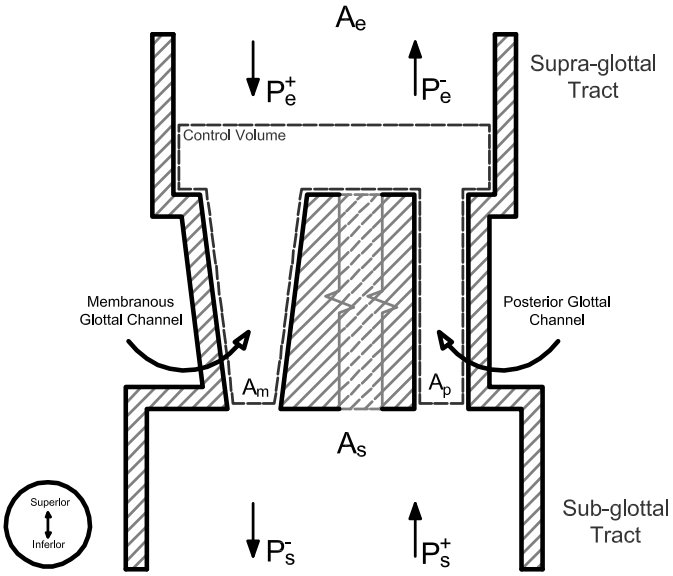


Figure 3.3. Control volume for separate posterior and membranous glottal flow channels.

The addition of viscous fluid effects in the trans-glottal pressure is therefore defined by

$$(P_s - P_e) = \frac{\rho k_t}{2}\frac{Q_g\left|Q_g\right|}{A_g^2} + R_g Q_g, \tag{3.1.21}$$

where replacing the left side with these components of incident and departing pressures waves, yields

$$\frac{\rho k_t}{2}\frac{Q_g\left|Q_g\right|}{A_g^2} + \left(\frac{\rho c}{A^*} + R_g\right)Q_g - \delta_P = 0. \tag{3.1.22}$$

This quadratic equation has a similar solution to the one applied by Lucero and Schoentgen,[96] obtaining

$$Q_g = \pm c\frac{A_g}{k_t}\left\{-A_g\left(\frac{1}{A^*} + \frac{1}{\rho c}R_g\right) + \left[A_g^2\left(\frac{1}{A^*} + \frac{1}{\rho c}R_g\right)^2 + \frac{2k_t}{\rho c^2}\left|\delta_P\right|\right]^{1/2}\right\}. \tag{3.1.23}$$

It is important to note that the division by small numbers is avoided when $A_p >> 0$. However, when the PGO is close to zero, Equation 3.1.23 becomes numerically unstable when the membranous glottal area A_m goes to zero, since the glottal flow resistance goes tends to infinity. Therefore, a numerically stable solution for the flow is in the case of PGO ≈ 0

$$Q_g \approx \frac{2\delta_P A_g^3}{\left(\frac{\rho c A_g^3}{A^*} + \gamma_m\right) + \left[\left(\frac{\rho c A_g^3}{A^*} + \gamma_m\right)^2 + 2k_t\rho\left|\delta_P\right|\right]}, \tag{3.1.24}$$

which eliminates the small number division produced by the sudden reduction of the membranous area on the closing phase when no PGO is present.

Given that the inclusion of a separate flow channel does not modify the equations of motion, and neither directly modifies the acoustic propagation of sound in the vocal tract, the general assumption of level 2 interaction[166] between vocal tract, glottal flow, and the kinematic behavior, remains unaltered. Thus, the inclusion of the PGO is an improvement to the regular BCM, and can eventually be included in several other lumped element models that work under similar assumptions.

Given that the area used for flow calculation is modified and a PGO is now considered, the turbulent glottal flow solution[167] must be revisited. The turbulent glottal flow has been defined as a random flow source defined by $Q_n = \mathcal{H}_{Re_c}^{(Re)}\nu$, where $\mathcal{H}_a^{(x)}$ is the Heaviside step function center in a, and ν is a Gaussian noise $\nu \sim \mathcal{N}(\mu_\nu, \sigma_\nu)$[174] characterized by

$$\mu_\nu = 0, \tag{3.1.25}$$

$$\sigma_\nu = Re^2 - Re_c^2, \tag{3.1.26}$$

where Re_c is the threshold Reynolds number where the noise source is activated (usually between

1200 and 1800 for voice applications[167]), and Re is the Reynolds number calculated by

$$Re = Q_g \frac{D_H}{A_g} \frac{\rho}{\mu}, \tag{3.1.27}$$

where D_H is the hydraulic diameter of the glottal section A_g, and the membranous and posterior glottal sections produce only one noise source. This approach is useful for subsequent developments, where the PGO will be considered linked to a pre-phonatory membranous glottal opening (MGO). In addition, it is important to note that this representation of the Reynolds number is slightly different and complexer than the one used by Titze,[167] where the area and geometry of the glottal section reduces this equation to a simpler expression.

Finally, the total glottal flow (Q_T) can be considered as the sum of the glottal flow and the turbulent flow, defined by

$$Q_T = Q_g + Q_n. \tag{3.1.28}$$

To compare the behavior of the BCM with and without PGO, a set of simulations were performed to quantify the variations of some measurements of glottal flow. The PGO varied from 0 to 0.1 $[cm^2]$, which spans from complete closure to abnormal phonation. A constant sub-glottal pressure of $800[Pa]$ with muscle activation parameters of 0.1, 0.8 and 0.5 for the cricothyroid, thyroarytenoid and lateral cricoarytenoid, respectively.[173] The supra-glottal tract was obtained from magnetic resonance imaging (MRI) measurements for a sustained letter /e/ in a male speaker.[157] Other model parameters (e.g., supra-glottal pressure, air temperature and viscosity, etc.) remain unaltered between the two model simulations. the results are presented in Figures 3.4 and 3.5.

Figure 3.4 shows the resulting flow signals for the BCM with and without a PGO of $0.02[cm^2]$. In this figure, the effect of having a PGO results in an increment of the minimum and maximum flow, a smoothing of the closed-open transition, and a flow perturbation during the closed phase. The perturbation on the closed phase can be attributed to the tract coupling due to the PGO, creating an open channel that modifies the acoustical wave pressures even during full membranous closure.

On Figure 3.5(a), the variation of the fundamental frequency is obtained as a function of the PGO. No variation of fundamental frequency (f_0) is present in the original BCM. However, the inclusion of the PGO produces a decay in f_0, which can be attributed to the diminished energy affecting the kinematic behavior of the VF. Figure 3.5(b) shows the minimum and AC components of the glottal flow as a function of PGO. As before, the BCM exhibits no variation, while the inclusion of PGO increasingly raises the minimum flow and reduces the AC component of the flow.

Based on these results it is clear that the inclusion of a PGO on the BCM is an improvement in terms of its physiological relevance, which allows to model DC offset in glottal airflow, which is a commonly seen behavior.[116] However, even when the inclusion of a PGO is an important improvement, we still neglects the relation between posterior and membranous portions of the glottis, thus allowing for unrealistic scenarios where a large PGO could be present without

showing MGO in VF posturing and onset conditions.

3.2 Triangular body cover model of the vocal folds

The inclusion of a PGO is an important improvement in the conception of a more realistic glottal configuration.[133, 192] However, the lack of a relation between the membranous configuration and the posterior glottal channel is unrealistic, since the posturing of the VF affects both membranous and posterior glottal configurations.[169] On the other hand, a lumped element model with triangular shaped glottis has been used for speech synthesis[9, 10] to include the effects of the zipper-like vocal fold closure, which is commonly observed in female voices[141] and certain types of MTD.[107] However, the model uses only 2 masses, it does not account for muscle activation, and does not include important features used in more recent proposals.[96, 192]

In this section, a lumped element model that includes pre-phonatory posturing is presented. The proposed model includes modifications mentioned in previous sections and characteristics of triangularly shaped glottis from Birkholz[9] models. The particular implementation of the model includes arytenoid posturing,[133, 169] triangular shaped glottis,[9, 10] PGO,[192] body-cover structure,[154] viscosity effects on trans-glottal pressure[96] and turbulent flow solution.[167] The proposed model, referred to as the triangular body-cover model (TBCM), also converges to the BCM when no rotation or displacement is present. Due to the link between arytenoid posturing and glottal openings, it is hypothesized that if a triangular glottal shape is assumed, then a more representative voice production model can be obtained. Simulations of different scenarios will be performed in order to compare these results with clinically observed phenomena.

One of the key features of the proposed TBCM is the ability to represent incomplete glottal closure with a zipper-like closing shape, allowing to relate incomplete closure in a posterior and membranous glottal sections, referred here as PGO and MGO. Controlling the pre-phonatory positioning of the arytenoid cartilages, both PGO and MGO can be related, mimicking a mechanism of MTD type I.[107]

The arytenoids are pyramid-shaped cartilages that sit on top of the cricoid lamina (cricoarytenoid joints), serving as a posterior attachment for vocal fold structures (see Figure 3.6(a)), being controlled by intrinsic laryngeal muscles to open and close the glottis in movements of abduction and adduction of the VF. The rotation of the arytenoids can also be altered by the thyroarytenoid (TA), which represent the body mass of each vocal fold, and contributes to changing the tension and the length of the VF. The lateral cricoarytenoid (LCA) modify the base displacement of the arytenoids to open or close the glottis. Finally, the complete movement of the arytenoids is complemented by the transverse and oblique arytenoid muscles, allowing to control the proximity rotation of the arytenoids.[168]

In this study the arytenoid movement will be considered independent of the BCM muscle activation rules,[173] which means that their motion will not be tied to muscle activation. This approach is a simplification that allows convergence from the TBCM to the BCM, using the same activation rules for both models. The implementation of the arytenoid movement, fully-controlled with extended muscle activation, will be left for future studies.

Similarly as in the BCM, the TBCM consists of a series of masses interconnected with springs

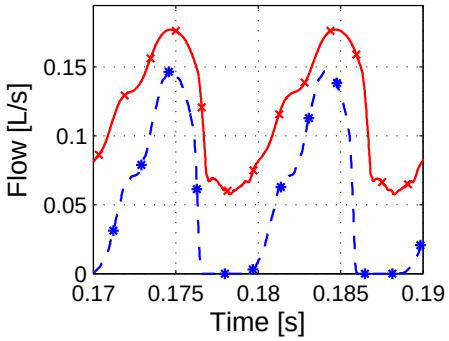


Figure 3.4. Resulting flow for the BCM and with and without PGO of $0.02[cm^2]$. [Legend: (*) BCM - (x) BCM + PGO]

and dampers (in Figure 3.7 represented as impedance boxes). However, the geometry of the cover masses is no longer parallel and is reshaped similar to the pre-phonatory shape of the VF. To describe for the different components of the TBCM, its mathematical conception is introduced in detail in this section. In addition, it is important to mention that this model is not restricted to symmetric cases, although only the symmetrical representation will be developed here for simplicity.

The equations of motions for the TBCM are

$$F_u = m_u\ddot{x}_u = F_{ku} + F_{kc} + F_{du} + F_{kuCol} + F_{duCol} + F_{eu}, \tag{3.2.1}$$

$$F_l = m_l\ddot{x}_l = F_{kl} - F_{kc} + F_{dl} + F_{klCol} + F_{dlCol} + F_{el}, \tag{3.2.2}$$

$$F_b = m_b\ddot{x}_b = F_{kb} + F_{db} - \left(F_{ku} + F_{du} + F_{duCol} + F_{kl} + F_{dl} + F_{dlCol}\right), \tag{3.2.3}$$

where $m_{\mathbf{x}}$ are the masses of the blocks with $\mathbf{x} \in \{u, l, b\}$ representing the upper, lower and body blocks, respectively; $x_{\mathbf{x}}$ represent the displacement of each mass over time, $F_{\mathbf{x}}$ stands for the force applied to each mass with k, d, c, and Col representing the springs, dampers, coupling spring, and collision elements, respectively. In this representation of the equations of motion, the time index was omitted to make the notation simpler.

Given that the displacement of each mass is usually considered in relation to its rest position, the following differentials displacement are defined

$$\Delta x_u = x_u - x_u^0, \tag{3.2.4}$$

$$\Delta x_l = x_l - x_l^0, \tag{3.2.5}$$

$$\Delta x_b = x_b - x_b^0, \tag{3.2.6}$$

where x_u^0, x_l^0 and x_b^0 are the rest positions of the upper, lower and body masses, respectively.

The spring forces can be defined as a combination of linear and non-linear components, obtaining

$$F_{ku} = -k_u\left[\left(\Delta x_u - \Delta x_b\right) + \eta_u\left(\Delta x_u - \Delta x_b\right)^3\right], \tag{3.2.7}$$

$$F_{kl} = -k_l\left[\left(\Delta x_l - \Delta x_b\right) + \eta_l\left(\Delta x_l - \Delta x_b\right)^3\right], \tag{3.2.8}$$

$$F_{kb} = -k_b\left[\Delta x_b + \mu_b\Delta x_b^3\right], \tag{3.2.9}$$

$$F_{kc} = -k_c\left(\Delta x_u - \Delta x_l\right), \tag{3.2.10}$$

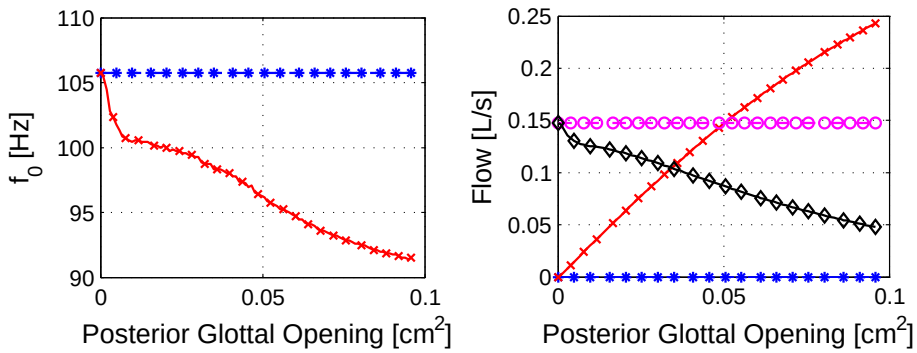
where $k_{\mathbf{x}}$ and $\eta_{\mathbf{x}}$ are the linear and non-linear spring coefficients, respectively.

The force produced by the dampers are

$$F_{du} = -d_u\left(\dot{x}_u - \dot{x}_b\right), \tag{3.2.11}$$

$$F_{dl} = -d_l\left(\dot{x}_l - \dot{x}_b\right), \tag{3.2.12}$$

$$F_{db} = -d_b\dot{x}_b, \tag{3.2.13}$$



(a) (b)

Figure 3.5. Simulation results for BCM and with and without varying PGO. (a) Fundamental frequency variation [Legend: (*) BCM - (x) BCM + PGO], (b) Minimum and AC flow variations. [Legend: (*) min Flow: BCM - (x) min Flow: BCM + PGO - (o) AC Flow: BCM - (◊) AC Flow: BCM + PGO]

where the damping coefficients $d_{\mathbf{x}}$ are defined as

$$d_u = 2\zeta_u\sqrt{m_u k_u}, \tag{3.2.14}$$

$$d_l = 2\zeta_l\sqrt{m_l k_l}, \tag{3.2.15}$$

$$d_b = 2\zeta_b\sqrt{m_b k_b}, \tag{3.2.16}$$

with $\zeta_{\mathbf{x}}$ being the damping ratio of each element.

Given the triangular shape of the glottis, the collision is no longer a step function but rather a progressive collision with a zipper-like closure of the glottal area. Therefore, the collision forces can be defined by its integral as

$$F_{kuCol} = -\frac{k_{uCol}}{\ell_g}\int_{\ell_g}\mathcal{H}_{x_u^c}^{(x_u(z))}\left[\left(x_u\left(z\right)-x_u^c\right)+\eta_{uCol}\left(x_u\left(z\right)-x_u^c\right)^3\right]dz, \tag{3.2.17}$$

$$F_{klCol} = -\frac{k_{lCol}}{\ell_g}\int_{\ell_g}\mathcal{H}_{x_l^c}^{(x_l(z))}\left[\left(x_l\left(z\right)-x_l^c\right)+\eta_{lCol}\left(x_l\left(z\right)-x_l^c\right)^3\right]dz, \tag{3.2.18}$$

where ℓ correspond to the membranous portion of the VF (with a total length ℓ_g), and x_u^c and x_l^c the collision plane of the upper and lower masses, which for the symmetric case can be assumed to be the mid-plane ($x_u^c = x_l^c = 0$), obtaining

$$F_{kuCol} = -\frac{k_{uCol}}{\ell_g}\int_{\ell_g}\mathcal{H}_{0}^{(x_u(z))}\left[x_u\left(z\right)+\eta_{uCol}\left(x_u\left(z\right)\right)^3\right]dz, \tag{3.2.19}$$

$$F_{klCol} = -\frac{k_{lCol}}{\ell_g}\int_{\ell_g}\mathcal{H}_{0}^{(x_l(z))}\left[x_l\left(z\right)+\eta_{lCol}\left(x_l\left(z\right)\right)^3\right]dz, \tag{3.2.20}$$

Assuming a homogeneous slope along the length of the glottis from the anterior commissure to the posterior vocal attachment (see Figure 3.8), the edges of the membranous area are defined by

$$x_u\left(z\right) = x_u + \frac{\delta_{x_u}}{\ell_g}z, \tag{3.2.21}$$

$$x_l\left(z\right) = x_l + \frac{\delta_{x_l}}{\ell_g}z, \tag{3.2.22}$$

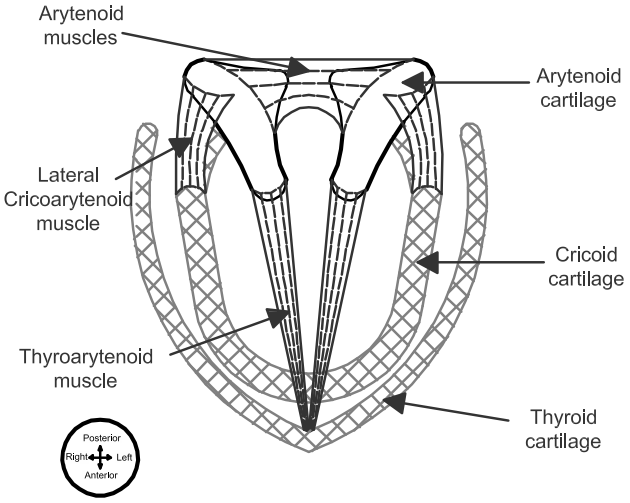
where δ_{x_u} and δ_{x_l} are the posterior rest displacements of the edges for the upper and lower masses, respectively. The collision forces can be restated as

$$F_{kuCol} = -\frac{k_{uCol}}{\ell_g}\int_0^{z_u^c}\left[\left(\frac{\delta_{x_u}}{\ell_g}z+x_u\right)+\eta_{uCol}\left(\frac{\delta_{x_u}}{\ell_g}z+x_u\right)^3\right]dz, \tag{3.2.23}$$

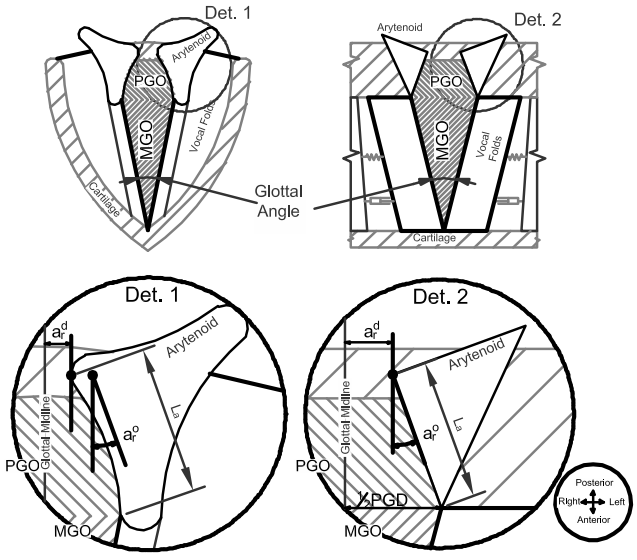
$$F_{klCol} = -\frac{k_{lCol}}{\ell_g}\int_0^{z_l^c}\left[\left(\frac{\delta_{x_l}}{\ell_g}z+x_l\right)+\eta_{lCol}\left(\frac{\delta_{x_l}}{\ell_g}z+x_l\right)^3\right]dz, \tag{3.2.24}$$

where z_u^c and z_l^c are the upper and lower collision detachment point in the z-axis. (see Figure 3.8)

Solving for the integrals and rearranging the terms, the following expression for the collision



(a)



(b)

Figure 3.6. Arytenoid cartilage posturing. (a) diagram with approximated muscular attachment, (b) Lumped mass elements modeling.

forces are obtained

$$F_{kuCol} = k_{uCol}\alpha_u \left\{ \left(x_u + \delta_{x_u} \frac{\alpha_u}{2} \right) + \eta_{uCol} \left[\left(x_u + \delta_{x_u} \frac{\alpha_u}{2} \right)^3 + \left(\delta_{x_u} \frac{\alpha_u}{2} \right)^2 \left(x_u + \delta_{x_u} \frac{\alpha_u}{2} \right) \right] \right\}, \tag{3.2.25}$$

$$F_{klCol} = k_{lCol}\alpha_l \left\{ \left(x_l + \delta_{x_l} \frac{\alpha_l}{2} \right) + \eta_{lCol} \left[\left(x_l + \delta_{x_l} \frac{\alpha_l}{2} \right)^3 + \left(\delta_{x_l} \frac{\alpha_l}{2} \right)^2 \left(x_l + \delta_{x_l} \frac{\alpha_l}{2} \right) \right] \right\}, \tag{3.2.26}$$

where $\alpha_{\mathbf{x}}$ are the normalized colliding portions of the upper and lower masses, which have the following definition

$$\alpha_u = \frac{z_u^c}{\ell_g}, \tag{3.2.27}$$

$$\alpha_l = \frac{z_l^c}{\ell_g}. \tag{3.2.28}$$

It is important to note that z_u^c and z_l^c can only exist between 0 and ℓ_g, therefore $\alpha_{\mathbf{x}}$ values are limited to the range $[0, 1]$. In addition, if the nonlinear term of the collision spring η_{Col} is set to zero, then the resulting force is equal to the collision used by Birkholz.[10] Note that, if no PGO is present, the solution is the same as in the original BCM.[154]

Similar to the collision spring forces, the damping forces will be also related to the ratio of collision for each mass, obtaining

$$F_{duCol} = -\frac{d_{uCol}}{\ell_g} \int_{\ell_g} \mathcal{H}_0^{(x_u(z))} \left(\dot{x}_u - \dot{x}_b \right) dz, \tag{3.2.29}$$

$$F_{dlCol} = -\frac{d_{lCol}}{\ell_g} \int_{\ell_g} \mathcal{H}_0^{(x_l(z))} \left(\dot{x}_l - \dot{x}_b \right) dz, \tag{3.2.30}$$

where d_{uCol} and d_{lCol} are the collision damping coefficients defined by

$$d_{uCol} = 2\zeta_u Col \sqrt{m_u k_u}, \tag{3.2.31}$$

$$d_{lCol} = 2\zeta_l Col \sqrt{m_l k_l}, \tag{3.2.32}$$

Solving for the integral and replacing the $\alpha_{\mathbf{x}}$ terms, the collision damping forces are

$$F_{duCol} = -d_{uCol}\alpha_u \left(\dot{x}_u - \dot{x}_b \right), \tag{3.2.33}$$

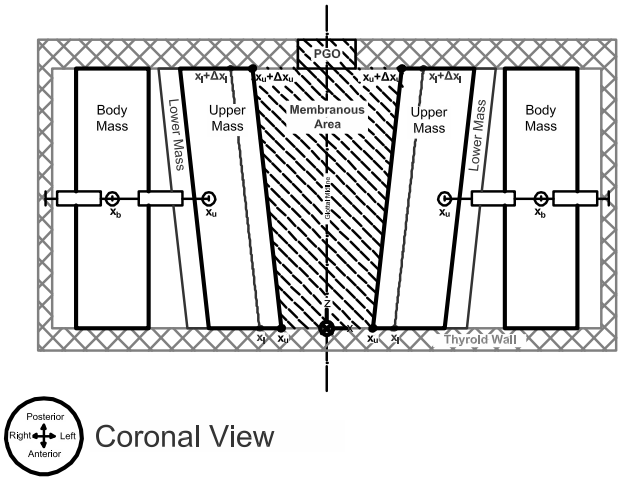
$$F_{dlCol} = -d_{lCol}\alpha_l \left(\dot{x}_l - \dot{x}_b \right). \tag{3.2.34}$$

The membranous glottal area produced by the triangularly shaped glottis can be defined as

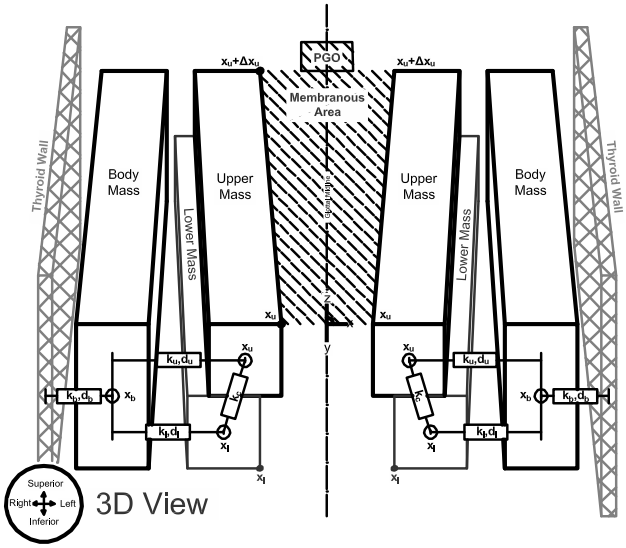
$$A_u = 2\ell_g \left(x_u + \delta_{x_u} \frac{(1 + \alpha_u)}{2} \right) (1 - \alpha_u), \tag{3.2.35}$$

$$A_l = 2\ell_g \left(x_l + \delta_{x_l} \frac{(1 + \alpha_l)}{2} \right) (1 - \alpha_l), \tag{3.2.36}$$

where A_u and A_l are the upper and lower open areas of the membranous portion. Therefore, the minimum membranous glottal area is defined by the minimum area between the upper and



(a)



(b)

Figure 3.7. Triangular Body Cover Model (TBCM). (a) Coronal View, (b) 3D View

lower masses, being calculated as

$$A_m = 2\int_{\ell_g} \mathcal{H}_0^{(x_m(z))} x_m(z)\, dz, \tag{3.2.37}$$

with x_m representing the minimum displacement along the z-axis, defined as

$$x_m(z) = \min\left(x_u(z), x_l(z)\right). \tag{3.2.38}$$

To relate the model parameters with the arytenoid posturing, a few terms must first be defined. For the posterior section of the glottis, the posterior glottal displacement (PGD) is the displacement of the posterior edges of the VF at the vocal process position, can be defined as

$$x_{PGD} = 2\left(a_r^d + \ell_a \sin\left(a_r^\circ\right)\right). \tag{3.2.39}$$

where a_r^d and a_r° are the displacement and rotation of the arytenoids cartilages, and ℓ_a is the length of the arytenoid face in the glottal channel.[86] In the symmetric case, PGD is related to the model by the following equations,

$$\delta_{x_u} = \frac{x_{PGD}}{2}, \tag{3.2.40}$$

$$\delta_{x_l} = \frac{x_{PGD}}{2}, \tag{3.2.41}$$

meaning that, in this particular case, the model is considered with equal posterior displacement for upper and lower masses. Nevertheless, this can be modified in future implementations in order to include the pyramidal shape of the arytenoids, which will imply a differentiated displacement for the upper and lower masses.

The posterior and membranous pre-phonatory areas (A_{PGO} and A_{MGO}), for arytenoid positive rotation and displacements, can be calculated as

$$A_{PGO} = \gamma_a \ell_a \cos\left(a_r^\circ\right)\left[x_{PGD} - \gamma_a \ell_a \sin\left(a_r^\circ\right)\right], \tag{3.2.42}$$

$$A_{MGO} = \frac{1}{2}\ell_g x_{PGD}, \tag{3.2.43}$$

where γ_a is the open portion of the arytenoid cartilage that interacts with the flow channel[62] (See Figure 3.6).

It is important to note that both A_{PGO} and A_{MGO} are related between each other by the arytenoid posturing, thus, no MGO can be observed without the presence of PGO. While the membranous area variates with time, the posterior area is a fixed onset condition ($A_p = A_{PGO}$).

To obtain the driving force due to the pressure on the glottal channel, a geometrical consideration must be made to allow for a simultaneous convergent and divergent sub-channels on the membranous glottis. According to the BCM,[154] the pressures on a parallel block representation

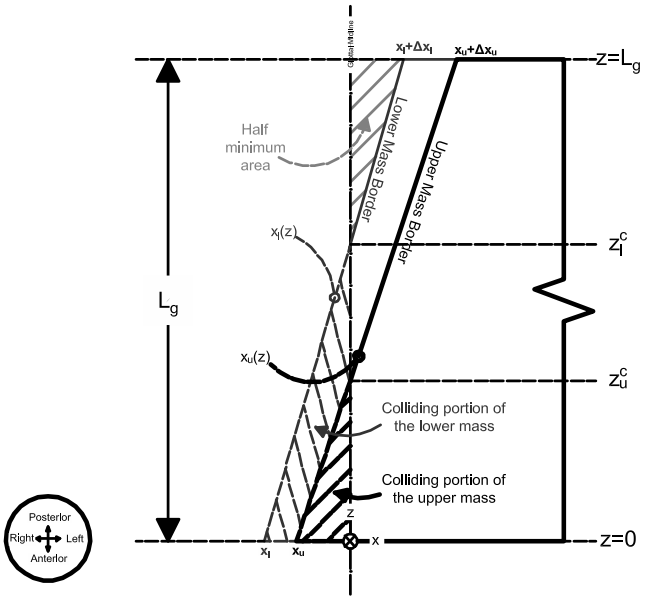


Figure 3.8. Collision detail for the upper and lower masses in the TBCM.

of the VF can be expressed as

$$P_u = \mathcal{H}_0^{(A_u)} P_e, \tag{3.2.44}$$

$$P_l = \mathcal{H}_0^{(A_l)} \left[P_s - (P_s - P_e) \left(\frac{A_m}{A_l} \right)^2 \right], \tag{3.2.45}$$

therefore obtaining the following forces

$$F_{eu} = T_u \int_{\ell_g} P_u(z)\, \mathrm{d}z, \tag{3.2.46}$$

$$F_{el} = T_l \int_{\ell_g} P_l(z)\, \mathrm{d}z, \tag{3.2.47}$$

For the TBCM, a similar approach will be taken. Assuming infinitesimal channels along the length of the glottis, the force due to pressure can be expressed as

$$F_{eu} = T_u P_e (1 - \alpha_u) \ell_g, \tag{3.2.48}$$

$$F_{el} = T_l \left[P_s (1 - \alpha_l) \ell_g - (P_s - P_e) \left(\ell_d + \int_{\ell_c} \frac{x_u^2(z)}{x_l^2(z)} \mathrm{d}z \right) \right], \tag{3.2.49}$$

where ℓ_d and ℓ_c are the portions of the glottis where the displacements, of the upper and lower masses produce a divergent or convergent channel, respectively. As before, if the PGD is set to zero, the Equations 3.2.48 converge to the BCM force equations.

In a similar way, the viscous losses can be calculated by separating the membranous glottal channel into a series of sub channels of opposing parallel plates with equal thickness.[96, 167] Therefore, the pressure differential for each i-channel is

$$\delta_{P_{m,i}} = \frac{12 \mu \ell_i^2 T_g}{A_{min,i}^3} Q_{m,i} = R_{m,i} Q_{m,i} \tag{3.2.50}$$

where for the membranous i-th channel, $\delta_{P_{m,i}}$ is the pressure drop due to viscous losses, ℓ_i is the length of the channel, $A_{min,i}$ is the channel minimum area, $symFlow_{m,i}$ is the volumetric flow, and $R_{m,i}$ is the flow resistance.

By circuit theory, the sum of several parallel independent channels have the following equivalent pressure drop

$$\delta_{P_m} = R_m Q_m = \left(\sum_{\forall i \in m} \mathcal{H}_0^{(a_{m,i})} R_{m,i}^{-1} \right)^{-1} Q_m, \tag{3.2.51}$$

where R_m is the oblique membranous resistance described in Equation 3.1.16. Therefore, the equivalent membranous flow resistance R_m for N parallel channels of equal width is

$$R_m^{-1} = \sum_{i=1}^{N} R_{m,i}^{-1}, \tag{3.2.52}$$

where $R_{m,i}$ is the i-th channel flow resistance defined by

$$R_{m,i} = \frac{3\mu T_g N}{2\ell_g} \frac{1}{\bar{x}^3_{min,i}}, \tag{3.2.53}$$

where $\bar{x}_{min,i}$ is the minimum midpoint edge displacement of channel i, obtaining an equivalent membranous flow resistance

$$R_m^{-1} = \frac{2\ell_g}{3\mu T_g N} \sum_{i=1}^{N} \mathcal{H}_0^{(\bar{x}_{min,i})} \bar{x}^3_{min,i}. \tag{3.2.54}$$

The minimum midpoint edge displacement of channel i ($x_{min,i}$) can be expressed as

$$\bar{x}_{min,i} = \min \left\{ x_u + \delta x_u \frac{(2i-1)}{N}, x_l + \delta x_l \frac{(2i-1)}{N} \right\}, \tag{3.2.55}$$

which in its differential form converges to Equation 3.2.38. Therefore, the equivalent flow resistance of the membranous area can be expressed as

$$R_m^{-1} = \frac{2}{3\mu T_g} \left[\int_{\ell_c} \left(x_u + \frac{\delta_{x_u}}{\ell_g} z \right)^3 dz + \int_{\ell_d} \left(x_l + \frac{\delta_{x_l}}{\ell_g} z \right)^3 dz \right], \tag{3.2.56}$$

which will be used in equation 3.1.18 and 3.1.17 to obtain the total flow resistance (R_g) due to viscous effect. In the TBCM formulation, it is assumed that the addition of an oblique glottal channel does not affect the analytic flow solution obtained for the BCM with a PGO. Therefore, the use of Equations 3.1.23, 3.1.24, and 3.1.28 remains unaltered.

To compare the behavior of the TBCM with the BCM with PGO a set of simulations has been performed to quantify the variations of common clinical measurements of glottal flow, as well as a comparison with statistical data of the glottal angle (See Figure 3.6).[50] The rotation of the arytenoid was varied from 0 to 5° with no displacement, or the equivalent PGO from 0 to 0.1 [cm^2], producing a PGD within the reported range up to 3 [mm][3, 8, 117, 132] producing a "glottal angle" up to 10 degrees.[24, 31, 74, 147] As in the BCM with PGO, the pressure was held constant at 800[Pa] with muscle activation parameters of 0.1, 0.8 and 0.5 for the cricothyroid, thyroarytenoid and lateral cricoarytenoid, respectively.[173]

Figure 3.9 shows resulting simulations withflow signals for the TBCM and the BCM with PGO with an opening of 0.02[cm^2] (Equivalent to a rotation of $a_r^\circ = 1°$). In this figure, the effect that the MGO has on the resulting flow when comparing it with the PGO only can be appreciated.

On Figure 3.10(a), the variation of the fundamental frequency is obtained as a function of the PGO. It can be appreciated here that for the TBCM a larger drop in frequency is observed, which can be attributed to the increased leakage of the TBCM. Figure 3.10(b) shows a similar behavior, where the TBCM has a more pronounced decay in flow measures than the BCM + PGO. It is interesting to highlight the presence of a local maximum in the AC Flow for small incomplete closures, which suggests a lower phonation threshold when MGO is considered, which is in agreement with previous experimental results in excised larynx.[5, 71, 107, 114, 128, 143, 144]

Additionally for lager PGD, the self-sustained phonation in the TBCM is hampered and ceases to occurs, thus suggesting that a compensatory mechanisms must be applied to keep sustained phonation in such scenarios. To illustrate this phenomena, a compensatory mechanism was proposed to evaluate hyperfunctional behaviors and its consequences on clinical measurements, allowing to evaluate and compare the outputs of the TBCM with results observed in clinical setups.

3.3 Discrete solver for differential equations

The use of lumped element models of voice production has been shown useful for understanding the underlying mechanisms of phonation.[44] Simplified systems, such as the lumped element models, allow accurate simulations of mechanical systems using only a fraction of the computational cost required for other methods (e.g., finite differences system[1,81,159]).[44] In addition, the low order nature of the lumped element models allows for a more direct representation of model parameters[173]

In this type of models, one of the most computationally demanding operations is solving the set of differential equations for each time step. This differential system of equations usually does not have an analytical solution, witch makes the use of ordinary differential equation (ODE) solvers such as Runge-kutta, Adams, or Rosenbrock, among others, compelling.[51] However, several ODEs solvers have a high computational demands and are impractical when multiple simulations are required (e.g., particle filtering, particle swarm optimization, etc.). Therefore, the implementation of a computationally efficient solver with small approximation error, for this specific problem, becomes a priority since several simulations will be required in subsequent chapters.

By definition all numerical models have an intrinsic representation error from the underlying assumptions. In addition, numerical solvers introduce approximation errors, thus incrementing the misrepresentation of the real system. In this section, a discrete solver for differential equations is applied to the proposed lumped element models of the VF, with the aim to reduce the computational time without significantly increasing the approximation error. This cost-efficient solver will improve our capacity to perform multiple simultaneous simulations.

The model used for this purpose is the TBCM discussed in the previous section, which will be simulated using a truncated Taylor series (TTS) and contrasted with a more commonly used solver Runge-Kutta of order 4. The TTS approximation is a discretization method based in the variable Taylor series[190] which provides a faster solution due to its explicit nature. The implementation of the TTS solver for the TBCM models is described in this section as part of the preliminary work for future chapters.

We start by defining a generic non-linear system as:

$$\dot{x} = f(x, u), \tag{3.3.1}$$

$$y = h(x), \tag{3.3.2}$$

where x is the evolving state of the system, y are the observations, u are the inputs, and f and g are the analytical non-linear state and observation functions.

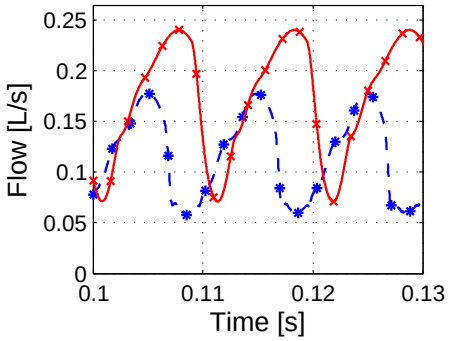


Figure 3.9. Resulting flow for the BCM with a PGO of $0.02[cm^2]$ and the TBCM with the equivalent PGO ($A_r^\circ = 1°$). [Legend: (∗) BCM + PGO - (x) TBCM]

A normalization of the system can be made such that,

$$\dot{x}(t) = f(x(t)) + g(x(t))\,u(t), \tag{3.3.3}$$

$$y(t) = h(x(t)), \tag{3.3.4}$$

where the general non-linear system is represented as an affine system, which can be redefined as a Taylor series representation, and where the same assumptions made for the linearly combined system can be used.

The relative degree r of the nonlinear system can be expressed as the numbers of times an output $y(t)$ has to be differentiated to obtain the input $u(t)$.[190] Therefore, the TTS for an affine system of degree r can be approximated as

$$x_{k+1} = x_k + \Delta_T x_k^{(1)} + \frac{\Delta^2}{2!}x_k^{(2)} + \frac{\Delta^3}{3!}x_k^{(3)} + \cdots + \frac{\Delta^r}{r!}x_k^{(r)}, \tag{3.3.5}$$

$$x_{k+1}^{(1)} = x_k^{(1)} + \Delta_T x_k^{(2)} + \frac{\Delta^2}{2!}x_k^{(3)} + \cdots + \frac{\Delta^{r-1}}{(r-1)!}x_k^{(r)}, \tag{3.3.6}$$

$$\vdots \tag{3.3.7}$$

$$x_{k+1}^{(r)} = x_k^{(r)}, \tag{3.3.8}$$

where Δ is the sampling period, $x_k = x(k\Delta)$ is the state sampled at time $k\Delta$, and $x^{(\ell)}$ is a short notation for the partial derivative $\frac{\partial^{(\ell)}}{\partial t^\ell}x(t)$

The fixed-time truncation error for this approach is a global vector of order Δ,[190] which provides better approximations when smaller time step are chosen. This characteristic is important since the level 2 interaction of the VF lumped mass models[166] conjointly with the WRA[90] and a high spatial resolution of the vocal tract,[157] demands a high sampling frequency due to the small amount of time that the sound wave requires to propagate backward and forward within a section of the tract.[90, 167]

When analyzing the structure of the kinematic equations of the TBCM, the following model can be obtained:

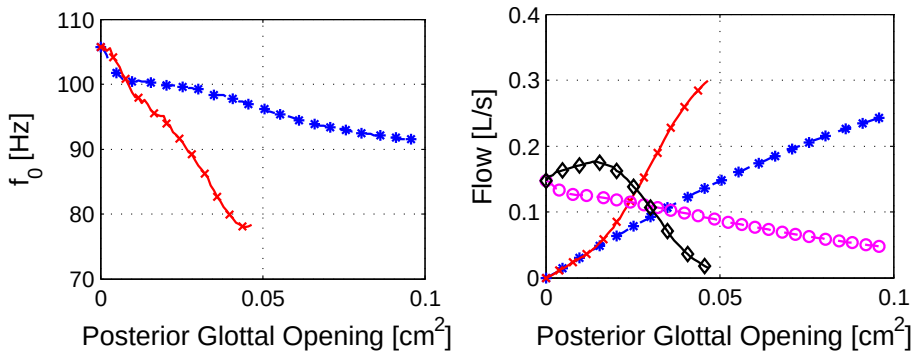
$$\ddot{x}(t) = m^{-1}\left[f(x(t)) + g(x(t))\,u(t)\right], \tag{3.3.9}$$

$$y(t) = h(x(t)), \tag{3.3.10}$$

where $f(x(t))$ is the combination of different nonlinear forces applied to the masses due to the springs, dampers, and collision forces; $u(t)$ are the current pressures on the sub and supra-glottal channels, which for the effects of the TTS approximation can be considered independent from the driving force; and $h(x(t))$ is the observation function, which provides the interaction for the acoustical and aerodynamic inputs.[166]

The state model for a lumped mass model can be rearranged to the following form

$$\dot{x}(t) = \begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}_{(t)} = \begin{bmatrix}x_2\\ m^{-1}\left[f(x) + g(x)\,u\right]\end{bmatrix}_{(t)}. \tag{3.3.11}$$



(a) (b)

Figure 3.10. Simulation results for the BCM and the TBCM with varying PGO. (a) Fundamental frequency variation [Legend: (∗) BCM + PGO - (x) TBCM], (b) Minimum and AC flow variations. [Legend: (∗) min Flow: BCM + PGO - (x) min Flow: TBCM - (o) AC Flow: BCM + PGO - (◊) AC Flow: TBCM]

Therefore, the corresponding TTS becomes

$$x_{k+1} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_{k+1} = \begin{bmatrix} x_1 + \Delta x_2 + \frac{\Delta^2}{2m}\left[f(x) + g(x)\,u\right] \\ x_2 + \frac{\Delta}{m}\left[f(x) + g(x)\,u\right] \end{bmatrix}_k \tag{3.3.12}$$

Using the equations of motion for the TBCM (Equation 3.2.1), the state space model using TTS becomes

$$x_{k+1} = \begin{bmatrix} x_u \\ v_u \\ x_l \\ v_l \\ x_b \\ v_b \end{bmatrix}_{k+1} = \begin{bmatrix} x_u + \Delta v_u + \frac{\Delta^2}{2m_u}F_u(x) \\ v_u + \frac{\Delta}{m_u}F_u(x) \\ x_l + \Delta v_l + \frac{\Delta^2}{2m_l}F_l(x) \\ v_l + \frac{\Delta}{m_l}F_l(x) \\ x_b + \Delta v_b + \frac{\Delta^2}{2m_b}F_b(x) \\ v_b + \frac{\Delta}{m_b}F_b(x) \end{bmatrix}_k \tag{3.3.13}$$

which can be further expanded using the expressions for the different forces given in Equation 3.2.1.

The TBCM was used to compare the results obtained between a TTS approximation and an ODE solver (Fourth order Runge-Kutta), which is considered, for the purpose of this study, the gold standard in VF model solvers.[192] For simplicity, the TBCM was simulated without vocal tract, turbulence, or glottal flow interaction. The input (u_t) of the system was thus taken as a random Gaussian pressure, with a mean of 85[Pa] and standard deviation of 10[Pa], filtered with a third order lowpass Butterworth filter with cutoff frequency at 500[Hz] (see Figure 3.11). The activation rules for lumped element models were used[173] with muscle activation constant of 0.1, 0.7, and 0.5 for the cricothyroid (CT), TA and LCA muscles, respectively. A PGO of $0.1[cm^2]$ was set for a male speaker, which implies an arytenoid rotation (a_r°) of $5^{\circ 62,86}$ with no displacement (a_r^d), thus producing a PGD of approximately 3[mm].

Three different sampling frequencies were simulated to quantify the effect of sampling period, and the same input signal was applied to the three scenarios while measuring processing time and accuracy. Both solvers were simulated independently using the same inputs and the same sampling frequency. The root mean square (RMS) error was obtained between ODE and TTS solutions. The summary of the resulting simulations are presented on Table 3.1

The minimum glottal area for the highest sampling frequency is presented in Figure 3.12(a), where a window of 100[ms] is shown. The great similitude between the results form ODE and TTS can be appreciated, which is also noted from Table 3.1, where it is shown that the error for that case is less than 0.01% with a computational improvement close to the 94%. The

Table 3.1. Comparative table with the results of the ODE4 solver and the TTS for different sampling frequencies.

Sampling Frequency	Time ODE	Time TTS	Improvement (%)	RMS error (%)
2000 [Hz]	13.03 [s]	0.74 [s]	12.29 [s] (94.3%)	63.7 $[mm^2]$ (0.318%)
4375 [Hz]	26.79 [s]	1.52 [s]	25.28 [s] (94.3%)	13.9 $[mm^2]$ (0.069%)
70000 [Hz]	373.22 [s]	23.99 [s]	349.23 [s] (93.6%)	0.7 $[mm^2]$ (0.003%)

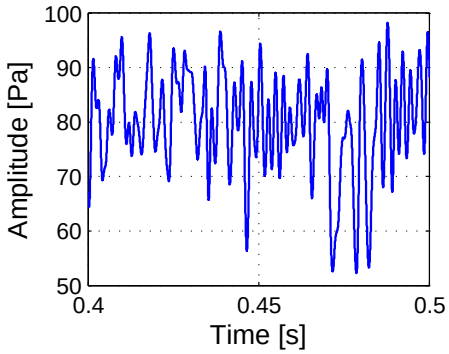


Figure 3.11. Simulated input incident pressure for the TBCM

accumulated RMS error is presented in Figure 3.12(b), where it becomes clear that a higher sampling frequency has lower sampling error.

The results from this simulation, combined with the high sampling frequency required for acoustic propagation, provides enough information to illustrate that the TTS approximation allows for a significant reduction of the computational requirements of the lumped element models of VF.

3.4 State space representation of vocal tract propagation

Several different systems (or components) are involved in the process of phonation, with an important part of the energy being shared between the interconnection of these different systems.[166] The propagation of acoustical waves in the vocal tract acts as a filter, introducing formants information that is key for voiced sounds and speech production,[164] paying a critical role in phonation.

The type 2 interaction introduced by Titze[166] relates the three main components required for voicing sounds: kinematic movement, airflow through the glottal area, and sound propagation in the sub and supra-glottal tract (see Figure 3.13). The first 2 elements (kinematics and flow) were discussed in previous sections, and we now discuss and propose a strategy to efficiently account for sound propagation.

There are two main schemes for sound propagation on concatenated tubes: A transmission line scheme[97] that uses an electrical representation to relate acoustical properties to electrical components, and a wave reflection analog (WRA) scheme[90,154] that uses a mathematical representation of traveling and reflecting sound waves. The selected propagation method in this thesis is a WRA scheme, since it has been broadly used in similar efforts with lumped mass element models.[44,166,192] It is important to clarify that, no statement is made to compare the two propagation schemes, thus leaving the option for future research in the efficient implementation of a transmission line scheme.

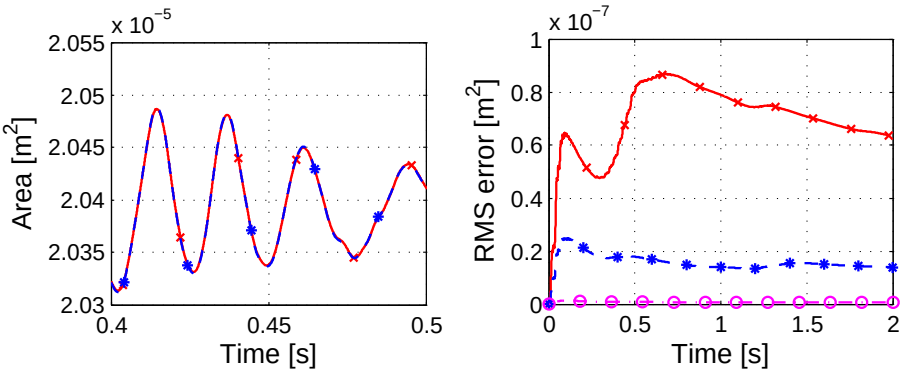
In spite of the simplicity of the WRA implementation, the computational time involved in its resolution could be considerable. However, given the linearity of the WRA system, and the incremental capability of modern computers to manipulate arrays of data (instead of single value registers), an important improvement can be obtained by rearranging the terms of the traditional formulation of WRA. Under this approach, the whole WRA computation can be reduced to a single step calculation that includes all tube sections, and all the inputs and outputs, even allowing the inclusion of inputs in no terminal tubes.

The WRA is based in the development of juncture analysis of the system shown in Figure 3.14.[90] The basic equations of the WRA for the departing pressures in juncture i can be expressed as

$$p^-_{i,k} = (1 - r_i)\,\alpha^+_i p^-_{i+1,k-1} + r_i \alpha^-_i p^+_{i-1,k-1}, \tag{3.4.1}$$

$$p^+_{i,k} = (1 + r_i)\,\alpha^-_i p^+_{i-1,k-1} - r_i \alpha^+_i p^-_{i+1,k-1}, \tag{3.4.2}$$

which is an adaptation of the same equations presented by Titze,[167] where $p^+_{i,k}$ and $p^-_{i,k}$ are the forward and backward departing pressures at time k, $p^+_{i-1,k-1}$ and $p^-_{i+1,k-1}$ are the incident



(a) (b)

Figure 3.12. Simulation results for the TTS solver. (a) Minimum glottal area observed [Legend: (x) ODE4 - (∗) TTS], (b) Accumulated RMS error for different sampling frequencies [Legend: (x) $f_s = 2000$[Hz] - (∗) $f_s = 4357$[Hz] - (o) $f_s = 70000$[Hz]]

pressures of the adjacent junctures at the previous time step, α_i^+ and α_i^- are the loss coefficients for the traveling wave in the forward and backward tube sections, and r_i is the juncture reflection coefficient obtained by

$$r_i = \frac{A_i^- - A_i^+}{A_i^- + A_i^+}, \tag{3.4.3}$$

where A_i^+ and A_i^- are the areas of the forward and backward tube sections adjacent to the juncture i. note that the areas connecting a juncture i and $i+1$ (A_i^+, A_{i+1}^-) are equal to the area of the i tube A_i. However, in order to keep the reference and notation relative to junctures i (and not to the tube section), the juncture notation A^+ and A^- will be used.

The loss coefficients $\alpha_i^\pm$ have diverse definitions according to different implementations of the WRA method.[167, 191] However, due to the fact that the description given in this thesis is not affected by the implementation of the attenuation factors, the definition used here will be the one adopted by Zañartu,[191] which is

$$\alpha_i^\pm \approx \begin{cases} 1 - \Delta_z \frac{3.8\times10^{-3}}{\sqrt{A_i^\pm}} & \text{supra-glottal tract} \\ 1 - \Delta_z \frac{11.2\times10^{-3}}{\sqrt{A_i^\pm}} & \text{sub-glottal tract} \end{cases}, \tag{3.4.4}$$

with Δ_z the length of each tube section in the tract.

Expanding the Equations 3.4.1 and 3.4.2 to a series of N concatenated tubes, and rearranging to fit a matrix state representation, we obtain:

$$\underbrace{\begin{bmatrix}\mathbf{p}^- \\ \mathbf{p}^+\end{bmatrix}_{k+1}}_{\mathbf{p}_{k+1}} = \tilde{\mathbf{A}} \underbrace{\begin{bmatrix}\mathbf{p}^- \\ \mathbf{p}^+\end{bmatrix}_{k}}_{\mathbf{p}_k} + \tilde{\mathbf{B}} \underbrace{\begin{bmatrix}\mathbf{p}_N^- \\ \mathbf{p}_0^+\end{bmatrix}_{k}}_{\mathbf{p}_{in,k}}, \tag{3.4.5}$$

where $\mathbf{p}^-$ is the arrangement of the backward pressures, $\mathbf{p}^+$ is the vector arrangement of forward acoustic pressures, and $\mathbf{p}_{in,k}$ is the input pressures of the system at time step k. The vectorial forms of the acoustic pressures are defined as

$$\mathbf{p}^- = \left[p_1^-, p_2^-, \ldots, p_i^-, \ldots, p_{N-2}^-, p_{N-1}^-\right]^T, \tag{3.4.6}$$

$$\mathbf{p}^+ = \left[p_1^+, p_2^+, \ldots, p_i^+, \ldots, p_{N-2}^+, p_{N-1}^+\right]^T, \tag{3.4.7}$$

while the state matrices are composed as

$$\tilde{\mathbf{A}} = \begin{bmatrix} (\mathbf{I}_{N-1} - \mathbf{R})\,\mathbf{A}^+\mathbf{S}_U & \mathbf{R}\mathbf{A}^-\mathbf{S}_L \\ -\mathbf{R}\mathbf{A}^+\mathbf{S}_U & (\mathbf{I}_{N-1} + \mathbf{R})\,\mathbf{A}^-\mathbf{S}_L \end{bmatrix}, \tag{3.4.8}$$

$$\tilde{\mathbf{B}} = \begin{bmatrix} (\mathbf{I}_{N-1} - \mathbf{R})\,\mathbf{A}^+\mathbf{M}_L & \mathbf{R}\mathbf{A}^-\mathbf{M}_U \\ -\mathbf{R}\mathbf{A}^+\mathbf{M}_L & (\mathbf{I}_{N-1} + \mathbf{R})\,\mathbf{A}^-\mathbf{M}_U \end{bmatrix}, \tag{3.4.9}$$

where $\mathbf{I}_{N-1}$ is the identity matrix of size $N-1$, $\mathbf{R}$ is the reflection coefficient matrix, $\mathbf{A}^\pm$ are the attenuation matrices, $\mathbf{S}_U$ and $\mathbf{S}_L$ are super and sub-diagonal matrices, and $\mathbf{S}_U$ and $\mathbf{S}_L$ are

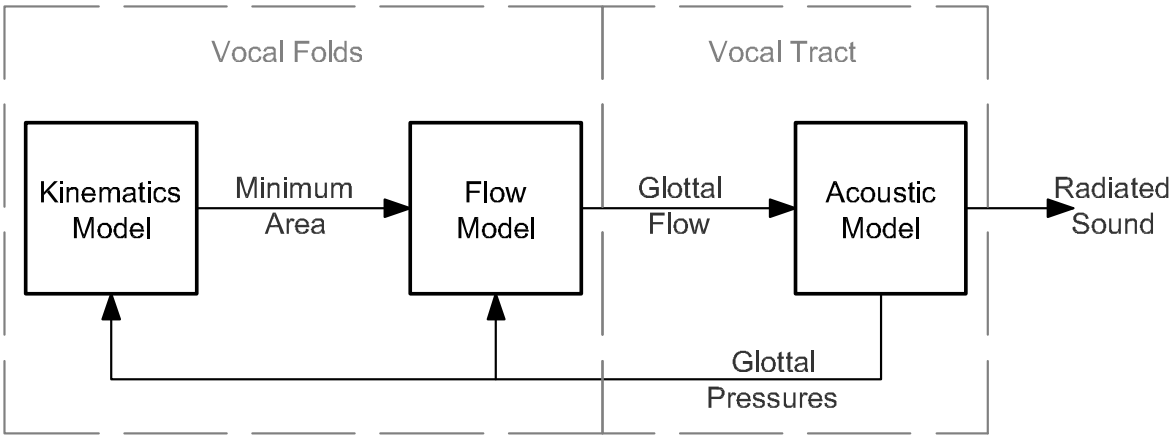


Figure 3.13. Three system interaction of the vocal folds. Kinematics model, flow model, and acoustic model

vectors used to adapt the position of the inputs. All these matrices are defined as

$$\left.\begin{aligned} \mathbf{R} &= \operatorname{diag}(r_1, r_2, \ldots, r_i, \ldots, r_{N-2}, r_{N-1}) \\ \mathbf{A}^+ &= \operatorname{diag}\left(\alpha_1^+, \alpha_2^+, \ldots, \alpha_i^+, \ldots, \alpha_{N-2}^+, \alpha_{N-1}^+\right) \\ \mathbf{A}^- &= \operatorname{diag}\left(\alpha_1^-, \alpha_2^-, \ldots, \alpha_i^-, \ldots, \alpha_{N-2}^-, \alpha_{N-1}^-\right) \end{aligned}\right\} \in \mathbb{R}^{N-1\times N-1}, \tag{3.4.10}$$

$$\left.\begin{aligned} \mathbf{I}_{N-1} &= (I)_{ij} = \delta_{i,j} \\ \mathbf{S}_U &= (S_U)_{ij} = \delta_{i,j-1} \\ \mathbf{S}_L &= (S_L)_{ij} = \delta_{i-1,j} \end{aligned}\right\} \in \mathbb{R}^{N-1\times N-1}, \tag{3.4.11}$$

$$\left.\begin{aligned} \mathbf{M}_U &= (M_U)_{ij} = \delta_{i,1} \\ \mathbf{M}_L &= (M_L)_{ij} = \delta_{i,N-1} \end{aligned}\right\} \in \mathbb{R}^{N-1\times 1}, \tag{3.4.12}$$

where$\delta_{i,j}$ is the Kronecker delta function, and $i, j = \{1, 2, \ldots, N-1\}$. Given that the simulation of WRA requires a double step process (due to the propagation of wave pressures[90]), the final state space model for the WRA is:

$$\mathbf{p}_{k+1} = \mathbf{A}\mathbf{p}_k + \mathbf{B}\mathbf{p}_{in}, \tag{3.4.13}$$

where

$$\mathbf{A} = \tilde{\mathbf{A}}^2, \tag{3.4.14}$$

$$\mathbf{B} = \left(\tilde{\mathbf{A}} + \mathbf{I}_{2(N-1)}\right)\tilde{\mathbf{B}}. \tag{3.4.15}$$

It is important to note that the input of the system is applied in the end junctures $i = 0$ and $i = N$ meaning that the inputs are reflected in the terms of p_0^+ and p_N^-, which are the forward and backwards incident pressures in the first and last tube sections, respectively. Therefore, since neither the radiation impedance, nor the nasal coupling are affected, the state space model (SSM) of the WRA can be implemented in sub-glottal, supra-glottal, nasal, and mouth tracts independently.

The nasal coupling can be modeled as three different tracts connected in a single juncture (see Figure 3.15). Therefore, following the same principles of conservation used to obtain the original WRA scattering equations, we can obtain[167]

$$p_{\mathcal{P},k}^- = \mathcal{F} - \alpha_{\mathcal{P}} p_{\mathcal{P},k-1}^+, \tag{3.4.16}$$

$$p_{\mathcal{M},k}^+ = \mathcal{F} - \alpha_{\mathcal{M}} p_{\mathcal{M},k-1}^-, \tag{3.4.17}$$

$$p_{\mathcal{N},k}^+ = \mathcal{F} - \alpha_{\mathcal{N}} p_{\mathcal{N},k-1}^-, \tag{3.4.18}$$

$$\tag{3.4.19}$$

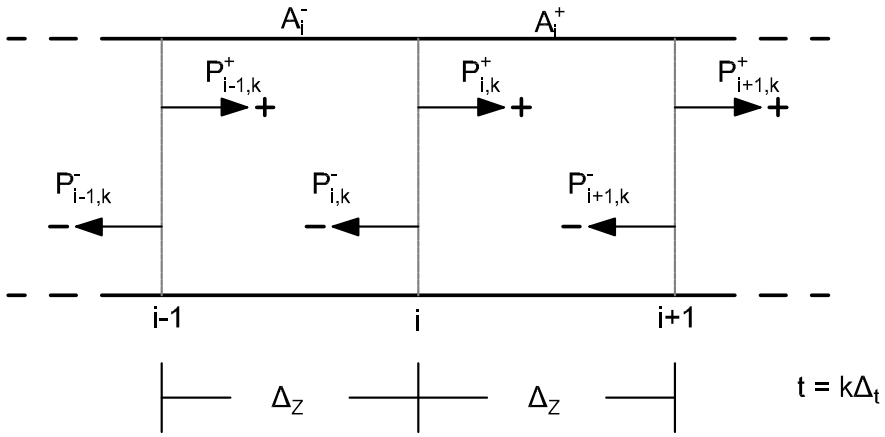


Figure 3.14. Juncture of two concatenated tube sections

where the common function $\mathcal{F}$ is defined by

$$\mathcal{F} = -\left(r_{\mathcal{M}} + r_{\mathcal{N}}\right)\alpha_{\mathcal{P}} p^{+}_{\mathcal{P},k-1} - \left(r_{\mathcal{N}} + r_{\mathcal{P}}\right)\alpha_{\mathcal{M}} p^{+}_{\mathcal{M},k-1} - \left(r_{\mathcal{P}} + r_{\mathcal{M}}\right)\alpha_{\mathcal{N}} p^{+}_{\mathcal{N},k-1}, \tag{3.4.20}$$

where the reflection coefficients are:

$$r_{\mathcal{P}} = \frac{\left(A_{\mathcal{P}} - \left(A_{\mathcal{M}} + A_{\mathcal{N}}\right)\right)}{\left(A_{\mathcal{P}} + A_{\mathcal{M}} + A_{\mathcal{N}}\right)}, \tag{3.4.21}$$

$$r_{\mathcal{M}} = \frac{\left(A_{\mathcal{M}} - \left(A_{\mathcal{N}} + A_{\mathcal{P}}\right)\right)}{\left(A_{\mathcal{P}} + A_{\mathcal{M}} + A_{\mathcal{N}}\right)}, \tag{3.4.22}$$

$$r_{\mathcal{N}} = \frac{\left(A_{\mathcal{N}} - \left(A_{\mathcal{P}} + A_{\mathcal{M}}\right)\right)}{\left(A_{\mathcal{P}} + A_{\mathcal{M}} + A_{\mathcal{N}}\right)}, \tag{3.4.23}$$

with the sub-index $\mathcal{P}$, $\mathcal{M}$ and $\mathcal{N}$ indicating the pharynx, mouth, and nose sections, respectively. Therefore, the following relationships with Equation 3.4.5 can be established

$$\left.\begin{aligned} p^{-}_{\mathcal{P}} &= p^{-}_{N} \\ p^{+}_{\mathcal{P}} &= p^{+}_{N-1} \\ \alpha_{\mathcal{P}} &= \alpha^{-}_{N} \\ A_{\mathcal{P}} &= A^{-}_{N} \end{aligned}\right\} \text{Pharynx Tract,} \tag{3.4.24}$$

$$\left.\begin{aligned} p^{+}_{\mathcal{M}} &= p^{+}_{0} \\ p^{-}_{\mathcal{M}} &= p^{-}_{1} \\ \alpha_{\mathcal{M}} &= \alpha^{+}_{0} \\ A_{\mathcal{M}} &= A^{+}_{0} \end{aligned}\right\} \text{Mouth Tract,} \tag{3.4.25}$$

$$\left.\begin{aligned} p^{+}_{\mathcal{N}} &= p^{+}_{0} \\ p^{-}_{\mathcal{N}} &= p^{-}_{1} \\ \alpha_{\mathcal{N}} &= \alpha^{+}_{0} \\ A_{\mathcal{N}} &= A^{+}_{0} \end{aligned}\right\} \text{Nose Tract,} \tag{3.4.26}$$

The simulation of radiation of the mouth and nose can be achieved by redefining the reflection coefficient of the end tubes of each tract, which according to Titze[167] produces the following radiation pressures

$$p^{-}_{N,k} = \frac{1}{b_2}\left[c_1\alpha^{-}_{N} p^{+}_{N-1,k-2} + c_2\alpha^{-}_{n} p^{+}_{N-1,k-1} + b_1 p^{-}_{N,k-1}\right], \tag{3.4.27}$$

$$p^{+}_{N,k} = \frac{1}{b_2}\left[\left(c_1 - b_1\right)\alpha^{-}_{N} p^{+}_{N-1,k-2} + \left(c_2 + b_2\right)\alpha^{-}_{N} p^{+}_{N-1,k-1} + b_1 p^{+}_{N,k-1}\right], \tag{3.4.28}$$

with the coefficients b_1, b_2, c_1, and c_2 defined by

$$b_1 = +I - R + RI, \tag{3.4.29}$$

$$b_2 = +I + R + RI, \tag{3.4.30}$$

$$c_1 = +I - R - RI, \tag{3.4.31}$$

$$c_2 = -I - R + RI, \tag{3.4.32}$$

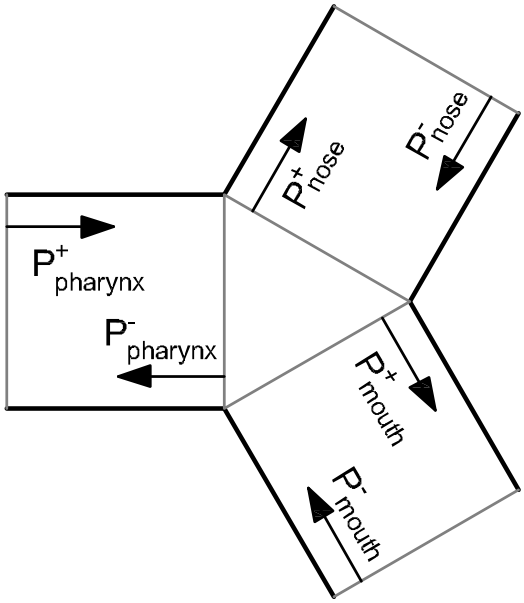


Figure 3.15. Three tract coupling juncture: Pharyngeal, nasal, and vocal tract.

where the inductance I and resistance R are the following[167]

$$R = \frac{128}{9\pi}, \tag{3.4.33}$$

$$I = \frac{2f_s}{c}\frac{8}{3}\sqrt{\frac{A_n}{\pi^3}}, \tag{3.4.34}$$

with f_s the sampling frequency which for the WRA must be equal to

$$f_s = \frac{C}{2\Delta_z}. \tag{3.4.35}$$

A toy example is used to illustrate the behavior of the SSM representation of the WRA, this model uses a vocal tract obtained from an MRI[157] for a male speaker with the sustained vowel /e/ (see Figure 3.16), and the input of the system as an impulse of 1[Pa]. The time series and the frequency response of both, the traditional WRA and the SSM implementation, are presented in Figure 3.17. An error below $1e^{-16}$ with a reduction up to 90% of the computational time was accomplished. Variations in the efficiency is, so far, attributed to the computational architecture, where the best improvements were observed in computers with higher bus-speed.

3.5 Selected parameter configuration

Several modifications to the original BCM were made to improve either physiological representativity or computational cost. The work introduced in this chapter produced several modifications in the configuration, which will constitute our "standard configuration" in the following chapters. The inclusion of a PGO is considered a crucial characteristic in both normal and pathological phonation.[192] Therefore, in the following analysis of the model, a PGO will be assumed to be always present unless otherwise stated. In the same way, the linkage of the PGO with the arytenoid posturing and the MGO has been proven to be of high relevance. Therefore, the TBCM will be the model of choice to link PGO with MGO, using all the considerations presented earlier in this chapter. To solve for the differential equations of the TBCM, the TTS approach will be used. The justification of this step is related to the high computational cost associated with solving differential equations with an ODE solver. The TTS provides an adequate level of precision with a low computational cost in the conditions of interest. However, if the sampling rate is lowered, then its error will be incremented and the TTS approach will no longer be suitable as a solver for the TBCM.

The vocal tract transformation is a compact and comprehensive way to represent the WRA behavior, eliminating the necessity of performing a two-step processing, and improving dramatically the computational cost of the simulations. Thus, its usage will also be considered part of the standard configuration in subsequent chapters.

The basic configuration used in the remaining chapters is summarized in Table 3.2, and special structural and configuration changes will be specified in a case-by-case basis. Additionally, when a stochastic approach is required, normally distributed parameters will be assumed, and the parameters will have mean and standard deviation as presented on Table 3.2.

To illustrate the validity of the complete model, a comparison with clinical measurements[116]

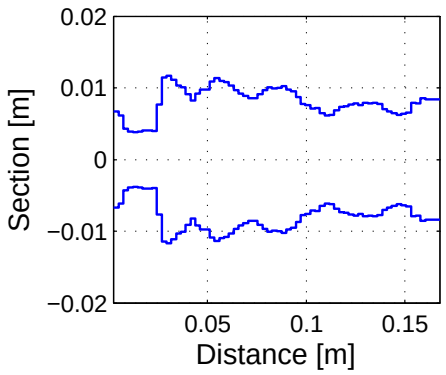


Figure 3.16. Vocal tract used for WRA simulation. The distance is measured from the glottal edge to the mouth for a male subject in a sustained vowel /e/[157]

was made. For this reason, normally distributed parameters were assumed according to Table 3.2, and the results of such simulation are presented in Table 3.3, where it can be appreciated that the fundamental frequency (f_0), the sound pressure level (SPL), and minimum glottal flow (Min-Flow) are within the expected values of the clinical measurements, while the MFDR and the amplitude of unsteady glottal flow (AC-Flow) are considerably low in the expected values. This behavior, while can still be considered as normal phonation, can be explained by the incongruency between muscle activation and onset posturing.

In addition, a sensitivity analysis was performed on each control variable listed in Table 3.2. This sensitivity was obtained varying a single parameter at the time, and moving it in a neighborhood of two standard deviations to avoid falling outside the operational point of the system, which is, the neighborhood where the system behave linearly in relation with the variation of the parameters. The results of the analysis are presented in Figure 3.18 and 3.19, where a comparison of the parameter and the most common measures used in clinical setups is made. In Figure 3.18(a), it can be appreciated that the most influencing parameters on the fundamental frequency is the activation of the CT muscle (a_{CT}). Nevertheless, other parameters such as arytenoid rotation, also have an important effects on the output. Figure 3.18(b) shows the sound pressure level variation over the selected parameters, where the thyroarytenoid muscle has almost no effect on the output, with the dominant parameters being the sub glottal pressure (P_s) and the rotation of the arytenoids (a_r°).

Figure 3.19 presents the aerodynamic response of the model for the variation of the different parameters. Figure 3.19(a) shows that the minimum flow obtained is dominated by the variation of the PGO through the rotation of the arytenoids. Figure 3.19(b) shows the dynamic amplitude of the flow signal, suggesting that the sub-glottal pressure and the PGO are key factors for this clinical measure. Figure 3.19(c) presents the maximum flow declination rate, reinforcing the idea of sub-glottal pressure and rotation being more sensitive parameters for the control of aerodynamic measures, with muscular activation influencing primarily spectral characteristics.

Table 3.2. Default parameter configuration for the remaining chapters, used when no other specification is given. (If no standard deviation is provided, the parameter is assumed a deterministic constant)

Parameter		Default (mean)	Standard deviation
Cricothyroid muscle activation (a_{ct})	[-]	0.04	0.03
Thyroarytenoid muscle activation (a_{ta})	[-]	0.18	0.22
Sub-glottal pressure (Ps)	[Pa]	726	99
Arytenoid Rotation (a_r°)	[°]	1	0.4
Lateral cricoarytenoid muscle activation (a_{lc})	[-]	0.5	
Supra-glottal pressure (Pe)	[Pa]	0	
Arytenoid Displacement (a_r^d)	[m]	0	
Arytenoid length ℓ_a	[m]	17.51×10^{-3}[86]	
Gender		male[173]	
Vocal tract		Takemoto /e/[157]	

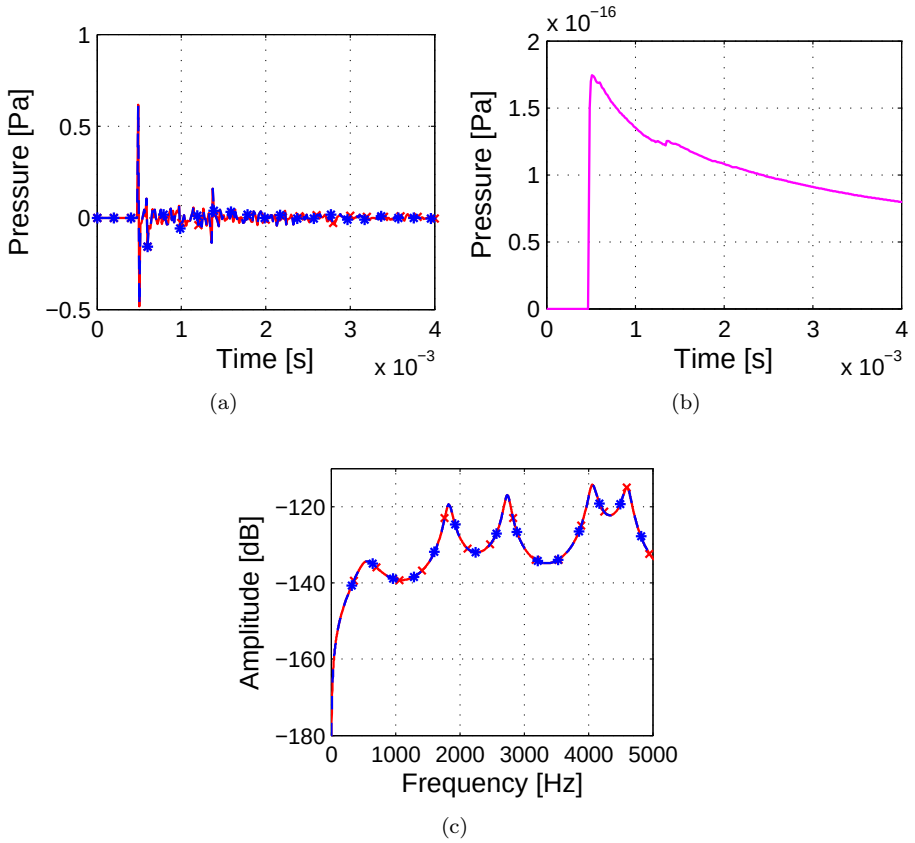


Figure 3.17. WRA results. (a) Radiated pressure of the time impulse response, (b) Cumulative RMS error, (c) Frequency components of the radiated pressure of the impulse response. Legend: (x) Sequential WRA, (*) State space WRA

3.6 Assessing the clinical relevance of the proposed model*

To conclude this chapter, and to validate the clinical relevance of the proposed model, a case of study is presented. Herein, the proposed TBCM is utilized in the study of compensatory mechanisms involved in phonotraumatic VH, namely VF nodules and polyps.

One of the most common applications of numerical models of voice production has been the study of underlying behaviors in the phonatory process. Including the analysis of VF modal response,[6] non-linear fluid-tissue-sound interaction,[43,151,166,195] asymmetric glottal-flow behavior,[40] VF collisions,[55,70,78,159,160] and normal and pathological voice production.[44,89,129,197] The current section shows the applicability of the TBCM in the analysis of vocal hyperfunction (VH).

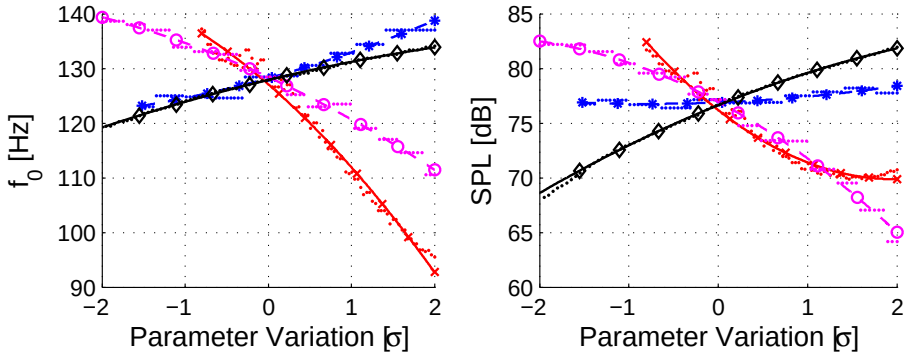
Clinical observations suggest a persistent incomplete glottal closure that involve both PGO and MGO can be precursor of phonotraumatic VH.[106] This suggestion was later supported by the prevalence of a large PGO after the surgical removal of phonotraumatic VF lesions.[73] Based upon that, we assume that VH is initiated by an onset condition that is associated to incorrect posturing that leads to incomplete glottal closure. Such posturing will produce a reduced efficiency in phonatory process that we assume will be self-perceived as a poor voice. In a attempt to compensate for such effect, we assume subjects display compensatory mechanisms that do not correct for the initial posturing problem. We will explore the use of common compensatory mechanism that have been observed in the clinical practice. The result of this compensation is expected to affect the tissue dynamics and exacerbate the initial incomplete glottal closure that acted as onset condition. This behavior is known as a vicious cycle of VH,[61] and is believed to be the main factor on the etiology of phonotraumatic VF lesions. Even when the vicious cycle theory was proposed more than 30 years ago, the physical mechanisms underlying this process have not been investigated, mainly because of the difficulties involved in performing *in-vivo* measurements of collision forces, energy transference, and other relevant parameters.[192]

In the proposed modeling framework, compensatory VH is assumed to be related with one, or a combination of, the following parameters: (1) Increased contraction of intrinsic TA and cricoarytenoid (CA) muscles, (2) Increased sub-glottal pressure, (3) constriction of the supra-glottal tract, and (4) altered auditory feedback. The contraction of intrinsic muscles is associated with an effort to adduct and maintain the closure of the membranous portion of the glottis. The increased sub-glottal pressure inserts more aerodynamic energy into the VF system, thus compensating for the energy losses due to the incomplete glottal closure.[60,61] The constriction of

*The content of this section has been accepted for publication in the Journal of Speech Language and Hearing Research.[50]

Table 3.3. Default model measures compared with clinically measured results[116]

Parameter	Perkell 94'[116] mean (std)	Default TBCM mean (std)
Sound Pressure Level (SPL) [dB]	77.8 (4)	72.8 (7.4)
Fundamental Frequency (f_0) [Hz]	112.4 (11.8)	118.0 (14.5)
Maximum flow declination rate (MFDR) [$^L/_{s^2}$]	337.2 (127.2)	150.1 (110.3)
AC Flow [$^L/_s$]	0.33 (0.07)	0.17 (0.07)
Min Flow [$^L/_s$]	0.08 (0.05)	0.08 (0.06)



(a) (b)

Figure 3.18. Sensitivity response of the model with adaptive polynomial fitting for (a) fundamental frequency, and (b) sound pressure level. [Legend: (x) a_{ct} - (∗) a_{ta} - (o) a_r° - (♢) p_s]

the supra-glottal tract, particularly from the epilaryngeal section, has been observed in normal and hyperfunctional voices,[5, 114, 128, 143, 144] and is believed to be linked with an increased acoustic coupling,[107] yielding an incremented efficiency in the transmission of acoustic energy. Finally, several studies have related the auditory feedback to phonotraumatic and non-phonotraumatic VH.[53, 109, 145] However, the role of auditory feedback in the context of vocal hyperfunction remains unclear.

A previous effort compared the effect of VH in presence of PGO,[192] introducing the concept of compensatory biomechanics in modeling hyperfunctional behaviors. The model used for that purpose was developed by our group and was described in the first section of this chapter, using a separate and independent channel to describe the aerodynamic and acoustic effects of PGO, and by compensating via sub-glottal pressure as a control variable. The goal was to maintain an SPL target between the normal scenario (no PGO) and the pathological scenarios with increasing PGO. However, the lack of an MGO limits the ability to simulate realistic scenarios when overly large PGO is present. In addition, the use of only one compensatory mechanism(sub-glottal pressure) and one target measure (SPL) introduces a bias in the analysis. Nevertheless, that study provided important insights on the possible pathogenesis of VH, by showing significant increase in MFDR and AC-Flow, which mimics clinical observations.[60, 61]

In the current section, the TBCM was used to extend our previous efforts and test prevailing assumptions about the physical mechanisms that underlie the role of compensation in phonotraumatic VH. The proposed assessment mechanism include the physiological enhancement of the TBCM, a more refined auditory feedback, and more comprehensive compensatory mechanisms such as increased sub-glottal pressure, selective activation of intrinsic laryngeal muscles, and supra-glottal compression. The main idea is to test the assumptions about the vicious cycle by controlling the posturing of normal vocal folds, creating a PGO linked to a MGO and mimicking clinical studies that propose these configurations as preceding onset of phonotrauma.[106] As before, the compensatory mechanisms are activated in the context of an optimization problem that aims to maintain a given auditory target without changing the posturing configuration. .

The measures used to quantify the characteristics of the determined voice configuration will be a combination of the unsteady component of the glottal flow AC-Flow, MFDR, fundamental frequency (f_0), SPL, harmonic to noise ratio (HNR), and signal to noise ratio (SNR). In addition, the net energy transferred (NET) to the vocal folds will be measured,[162] and the maximum contact pressure (MCP) of VF, which is calculated as the maximum of the ratio between the VF contact force and the VF contact area.

3.6.1 Modeling compensation in vocal hyperfunction

An optimization framework, where we search for combinations of compensatory mechanisms that can reach a given acoustic target for various incomplete closure scenarios, is proposed. Our prior approach adjusted lung pressure to reach a target SPL.[192] Herein, we extend this idea to include more control parameters and acoustic measures. Provided that acoustic measures of vocal function are used in the cost function, this optimization scheme can be viewed as mimicking a basic auditory feedback loop. It is acknowledged that this approach is an over-simplification of the complex mechanisms involved in auditory feedback, and does not account for psychoacoustic,

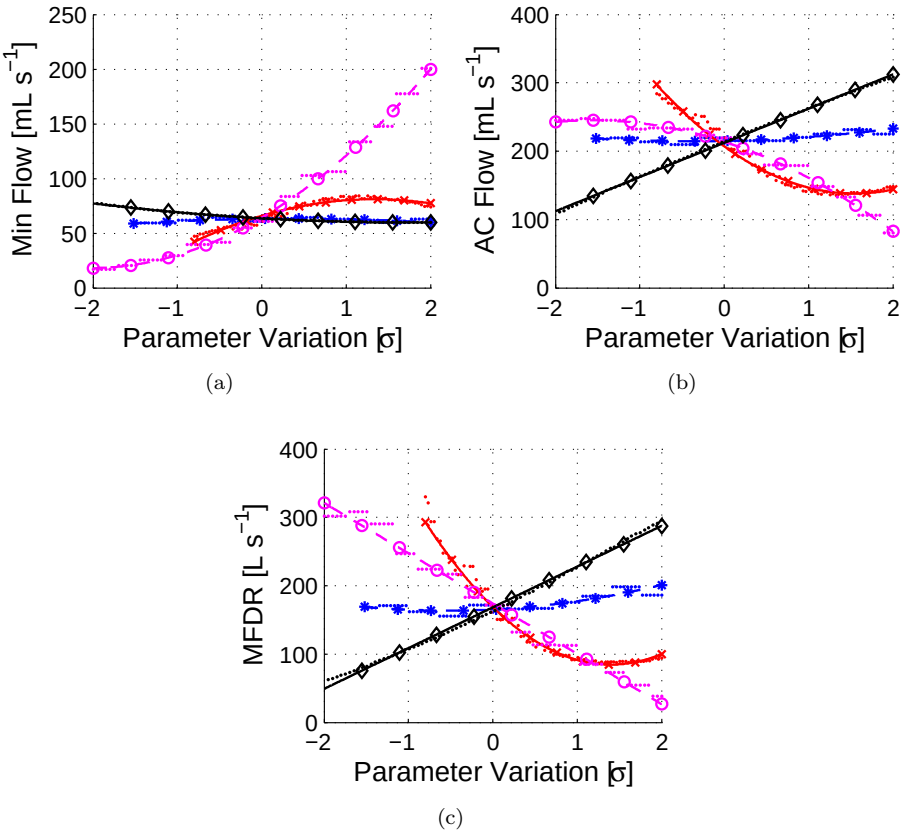


Figure 3.19. Sensitivity response of the model with adaptive polynomial fitting for (a) minimum flow, (b) AC flow, and (c) maximum flow declination range. [Legend: (x) a_{ct} - (*) a_{ta} - (o) a_r° - ($\Diamond$) P_s]

sensorimotor, and cognitive components that play important roles in feedback control.[54] Nevertheless, the proposed optimization framework is a reasonable first approximation to represent a basic auditory feedback mechanism in our self-sustained model of phonation.

To model hyperfunction, we create an onset VH condition by fixing a pre-phonatory configuration that is created by rotating the arytenoids, thus affecting PGO. While maintaining a given PGO configuration, we explore the role of compensatory mechanisms to restore a target output. Three types of compensatory mechanisms are considered, based on clinical and research observations: Increased lung pressure, VF tension (increased muscle activation), and supra-glottal compression (epilarynx tube narrowing).[130,131,142] We acknowledge that supra-glottal compression may not play a role in compensation (to maintain an acoustic output) and might rather just reflect a general increase in laryngeal muscle tension, particularly given the fact that it is also sometimes observed in normal subjects.[5] Nevertheless, it is frequently associated with VH and was included in the study in an attempt to directly assess its potential contribution/role in compensation. We assume that the compensatory mechanisms do not change the initial posturing configuration, which remains fixed for each of the conditions under evaluation.

In terms of the target output, we aim to maintain "voice quality", which is loosely referred to as a selected group of objective measures of vocal function, for simplicity. Four acoustic measures are used to quantify the vocal output: Namely, (1) SPL, (2) harmonic richness factor (HRF), (3) f_0, and (4) HNR. All of these acoustic parameters consider the spectral components up to 5 kHz, and are assumed to be sustained characteristics of phonation, meaning that neither intermittent cases nor variant cases are analyzed.

To represent the acoustic target in the optimization framework, a numerical space defined by

$$\delta(\omega) = \left(\sum_{i}^{\#\omega} \beta_i e^{\left| \frac{\omega_i}{\tau_i} \right|} \right) - 1, \tag{3.6.1}$$

is used to quantify a parametric distance, where ω is a characteristic voice vector of cardinality $\#\omega = 4$ with elements $i \in \{\text{SPL}, f_0, \text{HRF}, \text{HNR}\}$, β is a vector of scaling factors related to the importance of each parameter in the cost function, and τ is a dimensional tolerance factor used to regulate convexity across different parameters. In addition, a voice quality distance (QD) is defined as the parametric distance between a given voice vector ω, and a target characteristic voice vector ω_T, as

$$\text{QD}(\omega) = \delta(\omega - \omega_T). \tag{3.6.2}$$

This means that a given voice is closer to the desired target quality when the distance that separates them is shorter. Given that the parametric distance function is composed solely of objective measures, the associated QD can be used as a cost function for the optimization process, seeking to minimize the distance between a voice vector ω and the target voice vector ω_T. The parameter sets used for QD in this study are presented in Table 3.4. Note that other parameters, distance functions, and cost functions, may yield different solutions. Thus, the results obtained in this study can only be guaranteed for the numerical space defined by equation 3.6.1 and the parameters shown in Table 3.4.

To ensure physiological relevance of the compensatory variables, sub-glottal pressure is bounded between 0 and 2500[Pa], muscle activation must be in the range of [0,1] (from no-activation to fully-activated) for the TA and CT muscles, and supra-glottal compression alters the epilarynx tube section (A_e^c) with a factor of 0.2 - 1.8, meaning it varies between 20% - 180% of the default area.

Arytenoid positioning can be measured in various ways, and it is often reported using the vocal processes distance and the "glottal angle" measured between glottal edges at the anterior commissure (see Figure 3.6(b)). For pre-phonatory conditions, the vocal processes distance (referred to as PGD in this study) has been reported to span up to 3 [mm],[3,8,117,132] and the "glottal angle" has been reported to span up to 10 degrees.[24,31,74,147] To match these ranges, we consider a positive rotation of the arytenoids (a_r°) from 0 to 5 degrees, in approximately 0.08 degrees increments, with no separation of its base ($a_r^d = 0$), which leads to a "glottal angle" up to 10 degrees, with a PGD up to 3 [mm]. Note that maximum abduction arytenoid excursion can be much larger than the pre-phonatory values that we are considering in this study.

A Nelder-Mead optimization algorithm[34] is used to find the minimum QD. This algorithm has shown good results in similar applications,[37] and does not require a differentiable cost function to find local minima. The main disadvantage of this method is the inability to assure a global minimum, therefore, a random seeding approach is applied to minimize the error obtained in the prediction. For every compensated simulation, 8 random seeds are taken with each parameter distributed uniformly over the valid range. Each seed is validated to ensure self-sustained oscillation. The best result is identified by the error value, which is obtained from the QD previously described.

In summary, the method used to model compensation is based on the modification of the arytenoid rotation, which yields increasing PGD and affects both the PGO and the MGO. This change alters both the dynamics of the system and the voice quality-related measures used in the cost function, thus increasing the quality distance QD, which is interpreted as "non-optimal quality". To restore (or minimize) QD, the control parameters (sub-glottal pressure, muscle activation, and supra-glottal constriction) are conjointly-modified according to the Nelder-Mead optimization algorithm.

3.6.2 Results

Simulations and results are presented to asses the impact of compensation. The parameter sensitivity shows the variation of the model output given the variation of each control parameter independently (subglottal pressure, muscle activation, supraglottal constriction, and arytenoid rotation). Compensation was assessed by varying the rotation of the arytenoids to produce

Table 3.4. Proposed quality factors

Measure	Scaling (β)	Target (ω_T)	Tolerance (τ)
Sound Pressure Level (SPL)	0.4	81 [dB]	4 [dB]
Fundamental frequency (f_0)	0.3	118 [Hz]	12 [Hz]
Harmonic Richness Factor (HRF)	0.2	-10 [dB]	4 [dB]
Harmonic to Noise Ratio (HNR)	0.1	43 [dB]	10 [dB]

incomplete glottal closure while adjusting model parameters to achieve a target acoustic output.

Parameter sensitivity

The parameter sensitivity provides a description of the effects of selected input parameters (P_s, PGD, a_{CT}, A_e^c) on the output of the model, based on some of the vocal measures of interest (SPL, f_0, HNR). The input parameters are varied individually and its measures are compared using a normalized variation; this variation is a measure that scales each absolute value with the one obtained in the operational point (by default 1 means no variation), thus allowing for direct comparisons of the various measure sensitivities. The actual ranges for the model parameters are shown in Table 3.5, while the results of the parameter variation, in 64 steps, are presented in Figure 3.20. Only the parameter ranges where self-sustained oscillation was noted are reported.

While PGD was increased, SPL, f_0, and HNR, decreased reflecting what would be perceived as a softer, lower pitched, and breathier voice. Although better appreciated in our subsequent analysis, these parameters have a local maximum at a very small PGD, in agreement with previous studies that argue that a small leak is beneficial in phonation.[200]

Increases in P_s produced an increase in SPL, which is consistent with previous work;[192] an increase in f_0, which is well known in the literature;[167] and a higher HNR, most likely due to the stronger driving pressure. The variation of muscular activations for the CT muscles produced diverse effects on spectral parameters (f_0 and HNR). These results are consistent with the idea that CT activation and TA activation have antagonistic effects on vocal fold stress and strain[19] and that the operational point, chosen by the optimization of the QD with Perkell data,[116] weights SPL more than HNR. Increasing the supra-glottal constriction within the range of interest for this study (see Table 3.5) had only minor effects on all output measures.

To assess the contribution of each control parameter to the combined effect of the compensatory mechanism, a sensitivity analysis around the moving operational point was performed. The compensation algorithm was performed for different PGDs over the entire range of the study, and a series of operational points were obtained for each PGD. Herein, the variation of each parameter was set to 10% of its operational value. A normalization was performed to scale the effects of different isolated compensatory mechanisms, such that the sum of all sensitivities for a given PGD is set to be 1. Figure 3.21 illustrates the combined effect of all compensatory mechanisms across PGDs, where we note that sub-glottal pressure is the most dominant input parameter for most PGDs, followed by CT muscle activation. As expected based on their individual behavior, TA muscle activation and supra-glottal compression had minor effects in the

Table 3.5. Default and ranges for the input parameters in the triangular body cover model (TBCM).

Parameter	Default value	Range
Posterior Glottal Displacement (PGD)	$0.480[mm]$	$[0, 2.989]$
Sub-glottal pressure (P_s)	$900[Pa]$	$[0, 2500]$
CT muscle activation (a_{CT})	$0.105[-]$	$[0, 1]$
TA muscle activation (a_{TA})	$0.714[-]$	$[0, 1]$
Supra-glottal constriction (A_e^c)	$0.966[-]$	$[0.2, 1.8]$

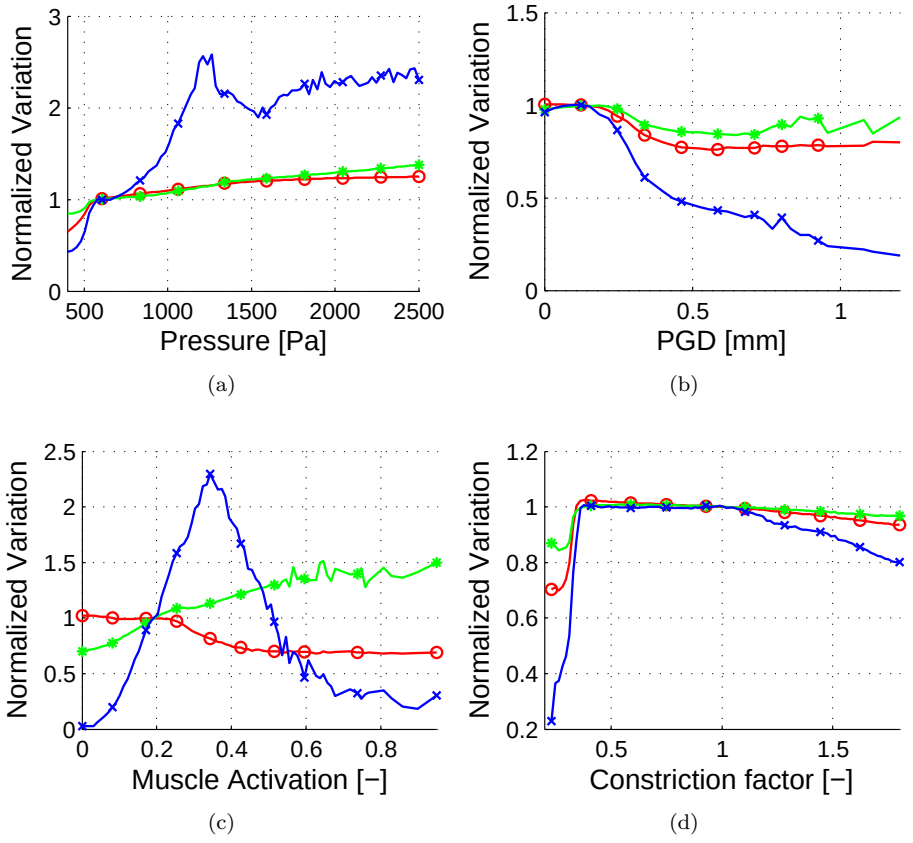


Figure 3.20. Effect of the model inputs on selected normalized vocal measures. Model inputs: (a) Sub-glottal pressure (Ps), (b) Posterior Glottal Displacement (PGD), (c) Cricothyroid muscle activation (a_{CT}), (d) Supra-glottal constriction (A_e^c). Model outputs: blue "x": Harmonic to noise Ratio (HNR); Red "o": Sound Pressure Level (SPL); Green "*": Fundamental frequency (f_0).∗

combined compensation effort.

Assessing the impact of compensation

To assess the impact of compensation, the rotation of the arytenoids was varied from 0 to 5 degrees in 64 steps, in order to produce incomplete glottal closure (including both the cartilaginous and membranous glottis) that is measured in terms of the PGD. For each step, we fixed the glottal configuration and introduced compensatory mechanisms to restore a target output through the feedback process previously described. Selected measures of vocal function that have been linked to phonotraumatic VH[60,61,192] are used to assess the effect of the compensatory action. Figure 3.22 shows the resulting MFDR, AC-Flow, MCP, and NET as functions of the PGD. Results are shown when both compensation action is present (compensated scenario) and absent (non-compensated). The higher variability of the compensated scenario data may be explained by the optimization algorithm, which may become entrapped in local minima rather than the global minimum, thus producing outlier measures. Thus, a trend line using a linear fit is also shown to facilitate the interpretation of the results of this case.

Figure 3.22(a) shows that when the gap becomes larger, AC-Flow becomes much smaller, if no attempts to compensate are present. However, when we compensate for the reduction in QD (given by a combination of SPL, f_0, HRF, HNR), the lack of closure and the compensatory action lead to a significantly increased AC-Flow and MFDR, both being more than doubled relative to normal for larger PGD. Higher than normal AC-Flow and MFDR are key features of patients with phonotraumatic VH and has been interpreted as reflecting the "vicious cycle" associated with these disorders.[60,61] Figure 3.22(c) shows that the impact on collision (MCP) as a function of PGD follows the same increasing trend as AC-Flow, MFDR, and NET; i.e., a marked increase in collision forces was observed when compensation was applied to offset an increase in PGD. This behavior is again repeated for the energy transferred to the vocal folds as the PGD increases. In the non-compensated case, NET decreases with incomplete glottal closure, while in the compensated condition it shows a marked increase that is consistent with those observed for MCP, MFDR, and AC-Flow

In addition, it can be seen in the non-compensated cases of Figure 3.22, AC-Flow, MFDR and NET have a local maximum at a PGD 0.2 [mm], which agrees with the idea that a small leak can improve efficiency in phonation.[200]

Figure 3.23 shows the impact of compensation on AC-Flow and MFDR as SPL is increased, based on a regressed Z-score analysis. This approach allows for comparing individual observations of aerodynamic parameters to normative sets of data (all measures converted to mean = 0 and standard deviation (SD) = 1), while at the same time adjusting/correcting the measures for the effects of sound pressure level and fundamental frequency (based on regression analyses of underlying correlations). This method has been used in previous work to identify measures for individual phonotraumatic patients that are abnormal (exceeding 2 standard deviations) after adjusting/correcting for the impact on the measures of SPL,[60,61] and further applied in subsequent studies and applications.[93,192] Using the same principles employed by Zañartu[192]

*The content of this image has been previously submitted for publication in the Journal of Speech Language and Hearing Research.[50]

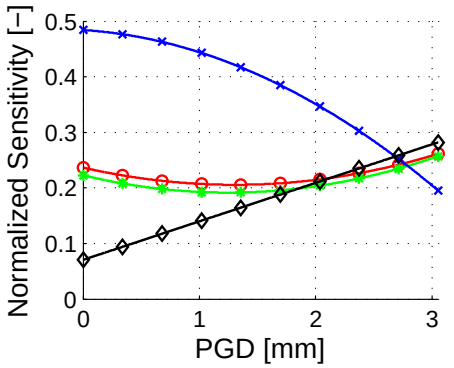


Figure 3.21. Combined effect of the compensatory mechanisms in terms of their normalized variation as a function of the Posterior Glottal Displacement (PGD). Model inputs: blue "x": Subglottal pressure (Ps); Red "o": Cricothyroid muscle activation (a_{CT}); Green "*": Thyroarytenoid muscle activation (a_{TA}); Black "□": Supra glottal constriction (A_e^c).∗

and Llico,[93] we linearly extended the normal bounds to provide a continuous function for the Z-score assessment. The results in Figure 3.23(b) are based on the normative data from Perkell[116] and show that the AC-Flow values for the non-compensated condition are within normal limits, but that the values exceed the normal range in the compensated condition as PGD increased. These findings are consistent with previous results for phonotraumatic patients[60,61]

The formation of benign VF lesions (e.g., nodules) is believed to result from chronic detrimental patterns of vocal behavior, which we now refer to as phonotraumatic VH.[103] It is believed that a key component of these disorders is the vicious cycle that develops as patients attempt to maintain the loudness and quality of their voices following the onset of vocal fold tissue trauma.[60,61] However, challenges associated with obtaining relevant in vivo measures have precluded the verification/quantification of these phenomena. We recently provided an initial demonstration that lumped element numerical models have the potential to provide new quantitatively-based insights into the underlying biomechanical and aero-acoustic mechanisms associated with the pathophysiology of these disorders, particularly with respect to the role of compensation.[192] The present study extends our previous work by expanding the modeling framework to be more physiologically and clinically relevant, including better differentiation of glottal closure (cartilaginous versus membranous), the addition of other mechanisms that may play a role in compensation, and a first approximation of how auditory feedback might impact compensation.

In line with our previous efforts,[192] we postulate that phonotraumatic VH is associated with biomechanical deficiencies that are exacerbated by abnormal compensations that can be quantitatively described through physics-based modeling. This study provides a numerical modeling framework that attempts to mimic the underlying physical mechanisms associated with phonotraumatic VH. Rather than modeling VF lesions,[81,89] we start by modeling a normal voice and alter it by introducing an incorrect glottal configuration and compensatory mechanisms to restore a target output through a feedback process. This scenario mimics the clinical descriptions from Morrison[106] and Hsiung,[73] proposing that a posterior glottal opening that extends into the membranous glottis can be a precursor to phonotrauma. The approach extends the efforts of Dejonckere[32] by expanding the analysis to include a closed loop framework that can mimic the vicious cycle of VH with further compensatory mechanisms and a first approximation to auditory feedback, as well as focusing on the correlation between clinical measures of interest for the assessment of vocal function.

When varying the PGD (due to the arytenoid rotation), we noted that various measures of interest were significantly reduced in magnitude. In order of significance, this affected SPL, HRF, HNR, and f_0. We argue that this reduction for increasing PGD is a result of the strong nonlinear source-filter interactions[166] that occur with increasing incomplete glottal closure due to the reduction of the source impedance. It can also be extracted from Figure 3.20 and Figure 3.22 (for the non-compensated scenario), that several parameters have a local maximum at a very small PGD, in agreement with previous studies that argue that a small leak is beneficial in phonation.[200] This feature is believed to be associated with the connection between the membranous and posterior gaps in the proposed voice production model in this study, as it was not seen in our prior efforts.[192] Figures 3.20 and 3.21 also illustrate that the effect of increasing posterior glottal gap can be counterbalanced, primarily, by increasing lung pressure, and muscle

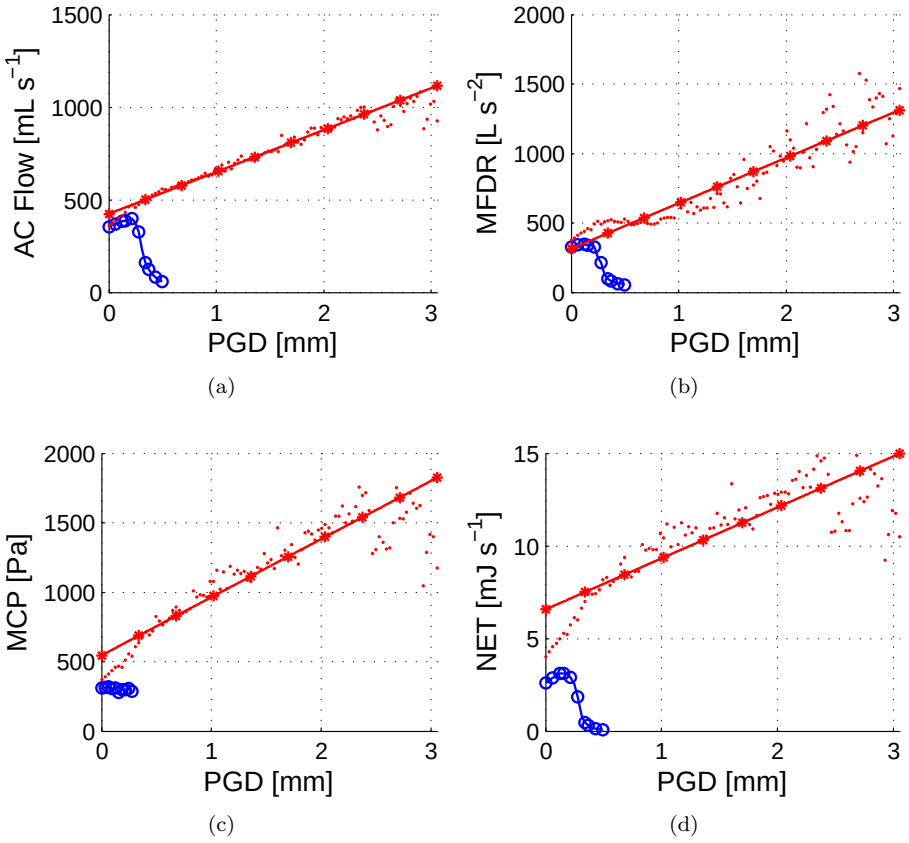


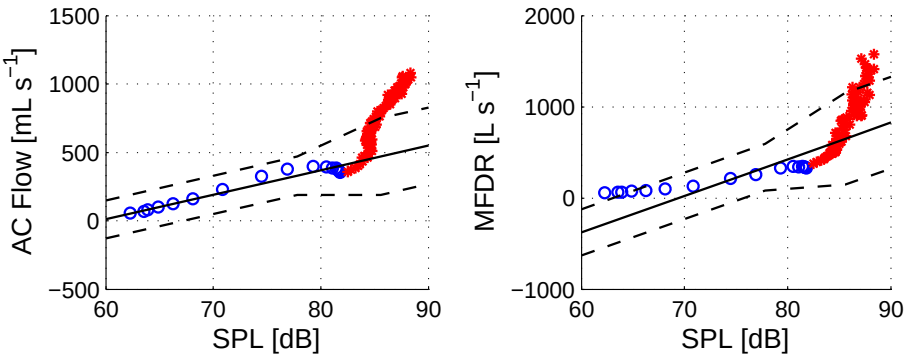
Figure 3.22. Selected output measures under increasing incomplete glottal closure as a function of the Posterior Glottal Displacement (PGD). (a) Flow range (AC-Flow), (b) Maximum flow declination rate (MFDR), (c) Maximum Contact Pressure (MCP), and (d) Net Energy Transfer (NET). [Legend. blue "o": non compensated case, red "*": compensated case, solid red: linear fit with R2 values of AC-Flow: 0.9808, MFDR: 0.8630, MCP: 0.8758, NET: 0.7226].∗

activation. The finding that supra-glottal constriction did not appear to play a significant role in compensation, combined with previous studies showing that it can sometimes also be observed in individuals with healthy voices,[5,114,128,143,144] casts some further doubt on the validity of using clinical (endoscopic) observations of supra-glottal compression as a metric for assessing VH.

Based on the clinical observations and associated hypotheses of Hsiung and Hsiao[73] and Morrison et al,[106] we created a static/fixed posterior glottal opening to mimic an onset or triggering condition for VH. A feedback mechanism was then implemented that attempted to reach a target for the four selected measures (SPL, HRF, HNR, and f_0), using lung pressure, supra-glottal constriction, and muscle activation as compensatory mechanisms. Maintaining these four measures within a given region is related to the idea of sustaining a desired loudness and quality; i.e., to compensate for the voice becoming too soft, breathy, and/or outside of a typical pitch range. This approach constituted an optimization problem, where we searched for the optimal combination of compensatory mechanisms that maintain the selected acoustic measures within a desired range (with a given tolerance, as noted from Table 3.4) for various posterior glottal openings. The feedback mechanism can be related to an auditory control loop, although it does not capture the psychoacoustic and sensorimotor complexities of the auditory and neural systems. We realize that the compensatory mechanisms could also affect the glottal configuration that we have assumed remains fixed for simplicity. Nevertheless, we argue that the proposed approach is sufficiently valid for a first approximation/proof-of-concept, given evidence that VH patients tend to maintain an incorrect glottal posturing[73,106] prior to vocal retraining (voice therapy), and that signs of some of the proposed compensatory mechanisms have been observed in such patients.[60,61,130,131,142]

When modeling phonotraumatic VH with the compensatory mechanisms in place, we observe that an increasing (membranous and posterior) gap due to incorrect posturing leads to increased AC-Flow, MFDR, and VF collision forces (see Figure 3.22). Increased AC-Flow and MFDR are key features of patients with VH and they have been associated, along with increased VF collision forces, with a vicious cycle in VH. Similar behavior is seen in the regressed Z-score analysis that contrasts our simulations with human recordings.[60,61] It is important to highlight that the acoustic output remains the same (within the given tolerance).

Our results point to a high correlation between AC-Flow and MFDR with VF collision forces. These correlations are expected to be physiologically relevant, but also possibly related to our modeling assumptions. We plan on further validating our numerical simulations with physical experiments using silicone models[108] to further understand the relation between collision forces and aerodynamic measures. Note that a stronger determinant of impact force may be given by the maximum area declination.[167] However, maximum area declination currently has less potential as a routine clinical measure than the aerodynamic measures used in this study because it requires calibrated laryngeal high-speed videoendoscopic imaging that is more difficult (and much more expensive) to implement in a clinical setting. Therefore, our findings support the use of aerodynamic measures for the clinical assessment of vocal function because of their potential to provide insights into critical underlying pathophysiological mechanisms that may not be adequately reflected in acoustic measures due to the types of compensation illustrated in this study. More attention is being placed on these measures in clinical studies,[101] and are also



(a) (b)

Figure 3.23. Regressed Z-Score for (a) Flow range (AC-Flow) vs Sound Pressure Level (SPL). (b) Maximum flow declination rate (MFDR) vs Sound Pressure Level (SPL) [Legend. Blue "o": non compensated case, Red "*": compensated case, Black solid: Z-Score mean, Black "-": Z-Score double standard deviation].∗

gaining attention in ambulatory studies.[93, 103, 193] The observation that aerodynamic measures are potentially more sensitive to changes in underlying pathophysiology than acoustic measures has been previously suggested based on studying treatment-related (voice therapy) vocal function changes in patients with vocal nodules.[67]

Further research is needed to explore the correlations between vocal measures of interest in other scenarios, as the cases investigated in this study and the modeling assumptions are based on modal phonation in normal male voices. Extensions to other groups and conditions,[118] including voice production in adult woman and children, are pending. We anticipate that some of the correlations may be less significant in women and children since the expected higher source-filter interactions in those cases could create more variability in the relationships between aerodynamics measures and contact forces.[153] However, the application to adult female voice is of great interest due to the higher prevalence of hyperfunctional voice disorders in women.[125] In addition, the presence of vocal fold lesions (e.g., nodules) is expected to further reduce glottal closure (increase the glottal gap), and thus increase the difficulties, forces, and tissue trauma associated with phonation (an exacerbation of the vicious cycle).

Conclusions

In this section, we illustrated that numerical models of voice production can provide important insights into the pathophysiology of phonotraumatic VH by providing estimates of parameters that are not readily accessible and thereby enabling the investigation of potential key relationships between vocal dysfunction and compensation. Specifically, modeling was used to show that compensating (increased lung pressure and laryngeal muscle activation) for what has been described as a VH onset condition (posterior glottal gap extending into the membranous glottis) can restore the target acoustic vocal output (SPL, f_0, HNR, HRF), but leads to high VF collisions forces (which is reflected in aerodynamic measures), and thus significantly increases the risk of developing phonotrauma. The onset of phonotrauma is expected to further increase compensatory forces, thus eliciting the vicious cycle that has been associated with phonotraumatic voice disorders (e.g., vocal fold nodules). The results also point to the potential clinical value of using aerodynamic measures to help detect VH; and further suggest possible limitations/risks associated with relying solely on auditory perceptual or acoustic measures to make clinical decisions about vocal function (e.g., for voice assessment and during voice therapy) since these parameters may not reliably reflect underlying VH due to compensation.

Chapter 4

BAYESIAN ESTIMATION OF VOCAL FOLDS MODEL PARAMETERS

This chapter includes the proposed Bayesian system identification scheme to obtain vocal folds (VF) model parameters and construct subject-specific representations. Herein, the base of statistical inference are presented along with the general formulation of Bayesian estimator based in particle filtering methods. In addition, the stochastic version of the vocal fold model for the estimation process used in clinical setups is introduced.

Bayesian estimation is a well-known mathematical technique used to perform stochastic system identification.[13] Given that the derivations of Bayesian estimators are explained in detail by other authors,[13] only the basic concepts will be presented in this chapter.

To introduce the topic of stochastic inference, it is necessary to understand the probability theory behind it, and in particular, how uncertainty is applied to the vocal folds models, and the benefit of such approach. In general terms, the core of estimating stochastic systems is uncertainty. While in a deterministic system the value of a variable is fully determined by a single realization, in a stochastic system, at least one variable is fully determined by a probability density function (pdf). In the case of vocal-fold modeling, the numerical construction produces a deterministic model. However given that the observed (real) system is stochastic by nature, the observations will have associated errors due to incomplete modeling, unaccounted randomness, measurement errors, etc. To correct such inconsistency two different approaches are possible: (1) re-modeling the VF system in a stochastic way, and (2) assume that the randomness is produced by control parameter variation and measurement errors. This work will use the second approach to the problem, adding measurement noise to the observations, including process noise to account for the un-modeled aspects of phonation, and input noise for the parameter variation.

Traditionally, the way of estimating model parameters from clinical measurement is through direct observation, which means, a direct measurement of the variable of interest (e.g., length of vocal folds during self sustained phonation). However, the direct observation of some subject-specific parameters is often complicated or impossible to obtain on *in vivo* subjects (e.g., intrinsic muscle activity in the larynx). Consequently, a way to represent the parameter is to perform direct measurements on excised systems, and use common statistics (e.g., mean and standard

deviation) as inputs in the model. However, the main problem of using average parameters is that the model created in this way stands for everyone in general, but no one in particular.[23] To correct this problem, a subject-specific approach can be design, where the parameters can be estimated for one subject in particular rather than a population. To perform such estimation, a technique that quantifies the uncertainty must be applied.

Given that obtaining proper model parameters is a fundamental part of VF modeling, and that subject-specific modeling produces subject-specific parameter set. The estimation of subject-specific parameters will produces a model especially tuned to represent a subject in particular, enabling capabilities such as the study of populations based on a case by case scenario, the study of pathologies based on the distribution of subject model parameters, and the estimation of features that can enhance the diagnosis and treatment capabilities for certain pathologies.

In subject-specific modeling, some efforts have been made using clinical measurements,[127, 187, 197] with most of the studies focusing on deterministic model parameters,[37, 126, 136, 137, 161, 183–185] and time-invariant stochastic parameters,[17] which used a microphone signal to estimate the muscle activation on a body-cover model (BCM). Here we introduce Bayesian inference scheme, which was recently reported by Hadwing et al.[56] Note that this approach was not tested in a clinical setup, which will be performed in Chapter 5 in this thesis.

Voice production is a complex phenomenon. The non-linear interaction between fluid mechanics, kinetic motion, and acoustics produce a broad scale of possible dynamic behaviors.[166] Given the physical nature of the voice, each one of the different elements involved in the process have, at a certain time, a unique realization or value that characterize them. As a result, it may be logical to extend this assumption to the modeling domain, by assuming that the system parameters are deterministic. However, due to the impossibility of accurately measuring the elements involved in phonation, the variability of the VF structures, the limitations of the modeling assumptions, and the inherent random of biological systems, it is convenient to represent the model parameters as random variable (rv) instead.

One of the important problems in stochastic modeling is the result interpretation. Having a result that is not "as easy" to interpret as in a deterministic method, produces an important barrier in the adoption of stochastic models. Stochastic estimations provides a complete provability distribution of a rv instead of a single value, making the comparison between the traditional deterministic tools and the stochastic approaches difficult. However, this is overcome when it is assumed that the deterministic solution is only one possible solution of the system and it is included in the solution obtained by the stochastic method, by assigning it a probability of occurrence based on the likelihood of the solution. Under this approach the stochastic estimation is an improvement of the deterministic approach, since it not only contains previous results, but also complements them with new information.

4.1 Framework

Before delving into the different estimation theories used in this chapter, it is necessary to define some common terms and notation. In addition, the general structure of the proposed stochastic system must be outlined to clarify what we understand by "system" and by "measurements" in

this case. Therefore, in the present section a comprehensive notation will be presented.

We define the discrete rv $X : \mathbf{\Omega} \to \mathbb{R}^n$ as the function that maps the probability space $\mathbf{\Omega}$ to the measurable space $\mathbb{R}^n$. The realization of the random variable X is defined by the lower case x with probability $\mathbf{Pr}\left(X \in \mathbf{A}\right)$ defined by $\mathbf{Pr}\left(\{\omega : X(\omega) \in A\}\right)$ with a pdf $f_X\left(X\right)$. The ensemble history of the rv X is defined by its bold letter $\mathbf{X}$

The basics of system identification is the estimation of either the state of the system, or the parameters that rules its behavior. To understand this concept we define the following stochastic system

$$\mathbf{Pr}\left(X\right) \sim f_X\left(X, U, \Theta, \eta\right), \tag{4.1.1}$$

$$\mathbf{Pr}\left(Y\right) \sim f_Y\left(X, U, \Theta, \nu\right) \tag{4.1.2}$$

where the system is defined by the pdf $f_X\left(\cdot\right)$ and $f_Y\left(\cdot\right)$, with input signal U, state X, parameter Θ, observation Y, and random noises η and ν (all rv). Consequently, the objective of system identification is to properly identify $\mathbf{Pr}\left(X\right)$ and/or $\mathbf{Pr}\left(\Theta\right)$, which, in the case of Bayesian estimation, can be obtained by $\mathbf{Pr}\left(X|Y\right)$ or $\mathbf{Pr}\left(\Theta|Y\right)$, which is the likelihood of having a certain state (or parameter) given the observation Y.

Obtaining $\mathbf{Pr}\left(X|Y\right)$ is linked to relating the state X with the observation Y, which in discrete (time series) models can be achieved by the following equations

$$X_{k+1} \sim f_X\left(X_{k+1}|\mathbf{X}_K, \mathbf{U}_K, \mathbf{\Theta}_K, \eta_k\right), \tag{4.1.3}$$

$$Y_k \sim f_Y\left(Y_k|\mathbf{X}_K, \mathbf{U}_K, \mathbf{\Theta}_K, \nu_k\right), \tag{4.1.4}$$

where the subindex K is the collection of all previous time $0, 1, ..., k$, with $X_0 \sim \mathbf{Pr}\left(X_{\text{o}}\right)$.

It is not hard to see that this representation can be a generalization of the Markovian representation of the vocal fold models (presented in Chapter 2), where X_{k+1} can be completely determined by the evolution of the previous state X_k, the input U_k, the configuration Θ_k, and the perturbation η_k.

The stochastic assumption presented in Equations 4.1.3 and 4.1.4 can be extended to the stationary case, allowing the parameter Θ have a time-invariant distribution. With this assumption, the following simplification can be made

$$\mathbf{Pr}\left(X\right) \sim f_X\left(X|\mathbf{U}, \Theta, \eta\right), \tag{4.1.5}$$

$$\mathbf{Pr}\left(Y\right) \sim f_Y\left(Y|\mathbf{X}, \Theta, \nu\right), \tag{4.1.6}$$

this simplification allows for separating the problem into two different branches, stationary and non-stationary estimation. Stationary estimation aims to identify the parameters Θ given the observation $\mathbf{Y}$ using the complete time series representation. On the other hand, non-stationary estimation allows to identify the state X given the observation Y using a Markov chain processor, in particular, a hidden Markov model (HMM).

The model parameters (Θ) are defined as the set of variables that rules the underlying behavior of the VF movement, which includes, for instance, the rest length of the VF, muscle activation, pre-phonatory posturing, driving pressures, etc. In a similar way, the observations

(Y) and measurements of the system will depend on the available data and model complexity, which include for example: microphone recordings, acceleration measured in the neck, high speed video recordings, flow measurements, etc.

The following sections contain a more in depth explanation on how the different estimation methods work, and how they behave under different circumstances and configurations.

4.2 Stationary estimation

The stationary process of VF parameter estimation can be defined as a pdf with the following distribution

$$\mathbf{Pr}\left(X\right) \sim f_{X|\mathbf{U},\Theta,\boldsymbol{\eta}}\left(X|\mathbf{X},U,\Theta,\eta\right), \tag{4.2.1}$$

$$\mathbf{Pr}\left(Y\right) \sim f_{Y|\mathbf{X},\Theta,\boldsymbol{\nu}}\left(Y|\mathbf{X},\Theta,\nu\right), \tag{4.2.2}$$

where X is the state of the system, Y the observation, U the input, and Θ the underling parameters. The noise η and ν are called, respectively, process and measurement noise, and represent the unknown and non-modeled aspects of the system and the measurement uncertainty.

The basic assumption of stationary estimation is that the parameters Θ do not vary over time. Thus, the observation Y, after a transient period, should remain stationary, allowing the computation of related measures of vocal function (e.g., fundamental frequency, maximum flow declination rate, etc.). Therefore, assuming a stationary process, the following Bayesian estimator can be drawn

$$\mathbf{Pr}\left(\Theta|Y\right) = \frac{\mathbf{Pr}\left(Y|\Theta\right)}{\mathbf{Pr}\left(Y\right)}\mathbf{Pr}\left(\Theta\right) \propto \mathbf{Pr}\left(Y|\Theta\right)\mathbf{Pr}\left(\Theta\right), \tag{4.2.3}$$

where $\mathbf{Pr}\left(\Theta|Y\right)$ is the probability of the parameter set Θ given that the measure Y is observed, $\mathbf{Pr}\left(\Theta\right)$ is the prior knowledge of Θ, $\mathbf{Pr}\left(Y|\Theta\right)$ is the likelihood of obtaining the measure Y given Θ, and $\mathbf{Pr}\left(Y\right)$ is a normalization factor called "evidence" which is no other than the probability of obtaining a certain measure.

The parameter Θ with pdf $f_{\Theta}\left(\Theta\right)$ is used as a configuration variable in the following differential system

$$\dot{X} = f(X, \Theta, U, \eta), \tag{4.2.4}$$

$$Y = g(X, \Theta) \tag{4.2.5}$$

where $f(X, \Theta, U, \eta)$ is the state model of the system with input U and noise η, and $g(X, \Theta)$ is the observation model.

If the system parameters are considered stationary, and the state and measurement models are time-invariant, then, the observations of that system should also be stationary, therefore, the measures of vocal functions obtained from those observations should be stationary as well. With this in consideration, the likelihood of equation 4.2.1 can be obtained with the following

pdf arrangement

$$f_{Y|\Theta}\left(Y|\Theta\right) = f_{Y|Y}\left(Y|g\left(X,\Theta\right)\right) \tag{4.2.6}$$

where marginalizing over Θ we can obtain the evidence as

$$f_{Y}\left(Y\right) = \int_{\Theta} f_{Y|\Theta}\left(Y|\theta\right) f_{\Theta}\left(\theta\right) d\theta \tag{4.2.7}$$

This approach allows for the definition of a prior distribution of Θ implemented in a state model, obtaining the subsequent dependent distribution of X and Y. Later, the prior distribution of Y is compared with the measurements, obtaining the probability of having the measurement y given the distribution Y, thus obtaining the posterior distribution of the parameters Θ by the following Bayes equation

$$\mathbf{Pr}\left(\Theta|Y=y\right) = \frac{\mathbf{Pr}\left(Y=y|g\left(X,\Theta\right)\right)}{\int_{\Theta}\mathbf{Pr}\left(Y=y|g\left(X,\theta\right)\right)\mathbf{Pr}\left(\Theta=\theta\right)d\theta}\mathbf{Pr}\left(\Theta\right) \tag{4.2.8}$$

In practice, the calculation of posterior probabilities, such as $\mathbf{Pr}\left(\theta|Y\right)$, is usually a complex and arduous labor,[83] in many cases not having a closed analytical solution. This is particularly true for non-linear systems such as the voice production models. As an alternative, a sampled distribution can be obtained, where the Monte Carlo sampling technique is the most well-known sampling method.[92] Therefore, the previous pdf can be represented by probability mass function (pmf) obtained in a discrete form[150]

The state of the system is given by the arrangements of position and velocities of the vocal fold masses, obtained by the truncated Taylor series (TTS) approximation, which is defined by

$$X_{k+1} = \begin{bmatrix} x_u \\ v_u \\ x_l \\ v_l \\ x_b \\ v_b \end{bmatrix}_{k+1} = \begin{bmatrix} x_u + \Delta v_u + \frac{\Delta^2}{2m_u}F_u\left(X,U,\Theta\right) \\ v_u + \frac{\Delta}{m_u}F_u\left(x\right) \\ x_l + \Delta v_l + \frac{\Delta^2}{2m_l}F_l\left(X,U,\Theta\right) \\ v_l + \frac{\Delta}{m_l}F_l\left(x\right) \\ x_b + \Delta v_b + \frac{\Delta^2}{2m_b}F_b\left(X,U,\Theta\right) \\ v_b + \frac{\Delta}{m_b}F_b\left(x\right) \end{bmatrix}_{k} + \eta \tag{4.2.9}$$

where x_u, x_l, x_b are the positions of the upper, lower, and body masses of the TBCM of the vocal folds, and v_u, v_l, v_b are, respectively, the velocities of the upper, lower, and body masses. Further details of the model and TTS can be found on Section 3.2 and 3.3 of Chapter 3. The input U and the parameters Θ, for this simulation, are considered to be a time-invariant driving pressure and muscular activation, respectively. The noise η is the vector containing all the noise components of the state X, where η is an independent and identically distributed (iid) noise. It is important to note that even when the positions and velocities are represented by lowercase letters, they actually stand for random variables and not just realizations.

To illustrate the process, two estimations based on simulations using the triangular body-cover model (TBCM) were performed. In these simulations two different scenarios were proposed: (1)

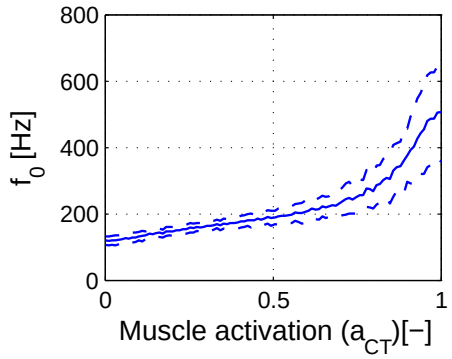


Figure 4.1. Prior fundamental frequency distribution as function of the CT muscle activation. Legend: (`Solid`) Mean value, (`--`) Standard Deviation

stationary configuration with cricothyroid (CT) muscle activation of 0.3, and (2) a non-stationary configuration of CT muscle activation varying between 0.2 and 0.5 with quasi-stationary periods, with each stage having equal time-length and equal transient-time. In both simulations, the system was configured equally, and used a time-frame of 6[s] with measures of vocal function taken every 50[ms]. The other parameters of the models were considered deterministic and known, with values established in the model configuration of the previous chapters (see Table 3.2 on Chapter 3). As an observation model, the fundamental frequency of the glottal movement is used, i.e. the fundamental frequency of the projected minimum glottal area (as being observed by an endoscope).

The *prior* information obtained from the model can be observed in Figure 4.1. In this figure, the distribution of fundamental frequency due to the activation of the CT muscle is presented. It can be appreciated here that the fundamental frequency increases with higher values of muscle activation, phenomena that is also observed in clinical studies[63] and simulated scenarios.[173] It is also clear from this figure that the standard deviation (SD) of the fundamental frequency, produced by the configuration variability of the model (see Table 3.2 on Chapter 3), also increases with the value of the fundamental frequency.

The simulated measurement of the fundamental frequency can be observed in Figure 4.2(a). In here, to avoid using the same data for simulation and identification, the measurement was obtained contaminating the model observation (output) with additive white Gaussian noise of zero of mean $\mu_\nu = 0$ and SD $\sigma_\nu = 10[Hz]$.

Figure 4.2(a) also presents the posterior estimation fundamental frequency, which is obtained by the execution of the Bayesian estimator, and posterior update of the observations. Figure 4.2(b) presents the simulated muscle activation along with the posterior estimated muscle activation. Both results presented in Figure 4.2(a) and (b) are obtained sequentially updating and increasing the number of samples used in the estimation. As a result, the estimation produces a narrowing distribution upon the target value (simulated input). As can be appreciated in the figure, when a stationary parameter is fed into the system, the estimation quickly converge to its target value, therefore validating the theoretical possibility of using this Bayesian framework in a VF model,[17] and particularly in the proposed TBCM.

However, if a stationary method is used to estimate time-varying parameters, system information will be inevitably lost due to misinterpretation of the measurements. This brings deterioration in the parameter estimation, eventually failing to describe the observed behavior. This phenomena is illustrated in figure 4.2(c) and (d), where an estimation of time-varying parameters is made. In here, it is appreciated how the estimation follows the first segment of the simulation. However, given that this process does not update the previous information, it fails to follow on the time-variation of the parameter. This error is produced because the stationary estimation does not correctly update the estimation history, therefore, assuming that the the new information is produced by the same underlying parameter values that produced previous observations.

An alternative to deal with time-varying conditions in a stationary framework is the common assumption of quasi-stationary behavior of speech signals.[113] This assumption allows the use of windowing techniques to estimate small time-frames, considering them as independent measure-

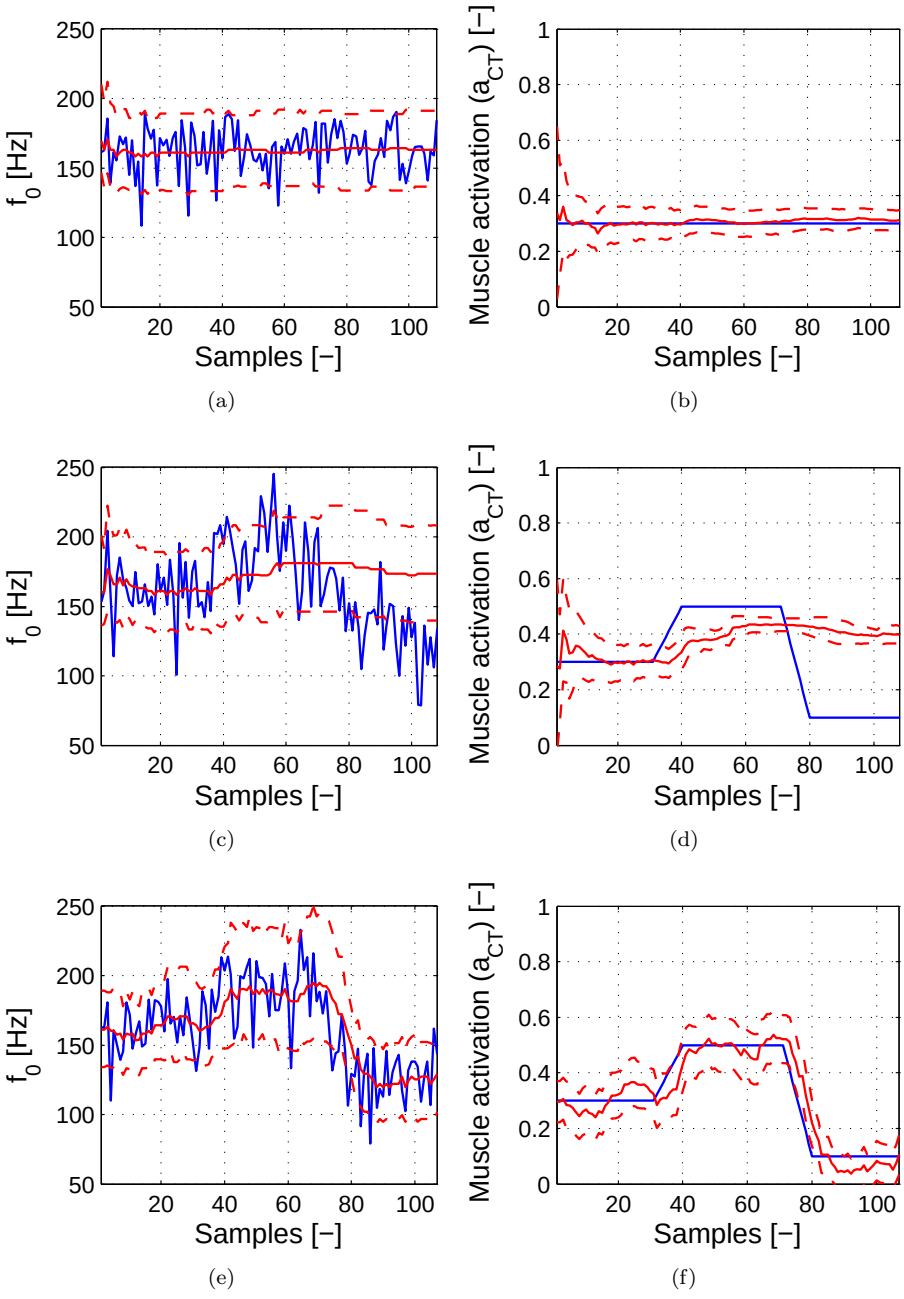


Figure 4.2. Fundamental frequency measurement and estimated observation (left) and CT muscle activation and its estimate (right). (a) and (b) Stationary parameters, (c) and (d) non-stationary parameters, (e) and (f) non-stationary parameters with windowing technique. [Legend: (Solid) measurement (left)/true value (right), (Solid) mean estimate, (--) 95% credibility]

ments, thus enabling the estimation of independent parameters on quasi-stationary intervals, as a result, liberating the history carriage produced by the stationary estimation. Figure 4.2(e) and (f) shows the application of such principle on a moving window of 10 samples with 9 samples overlapping. The improvement in the estimation when comparing the non-windowed result with the windowed result can be clearly appreciated (see Figure 4.2(d) and (e)). However, it is also noticeable that the transient period of the passage (using windows) is still not well estimated. The reason wy this process does not behave well on transient periods is because the underlying assumptions of the stationary estimator. In here, the method assume that any observation is produced by the system response to a time-invariant parameter. Thus, any changes in the observations is assumed to be a product of measurement noise, or to the incomplete modeling of the system (state noise). Therefore, the changes in the measurement are not attributed to variations in the parameter value, yielding a biased estimation of the system.

4.3 Non-stationary estimation

The main disadvantage of the stationary process is the inability to accurately estimate non-stationary scenarios, therefore, leaving aside many cases of analysis where the variation of the parameter critical (e.g., muscular variation in patients with muscle tension dysphonia (MTD) or Parkinson's disease).Biological systems, by nature, have stochastic fluctuation, and modeling of voice production, particularly during running speech, has time-varying model parameters (e.g., lung pressure, muscle activation, tract configuration, etc.). This type of behavior, as demonstrated in the previous section, is not well captured by stationary estimation methods. If a stationary estimator is used in a non-stationary problem, then, vital information could be lost, resulting in an identification error with misleading uncertainty.[83,84] This erroneous identification does not necessarily mean an undetermined pdf, but rather a pdf that does not truly represent the physical phenomena. Consequently, the importance of the estimation method is as important as the model used for the estimation, if any of them are not up to the physical behavior of the system, no useful estimation can be obtained.

The representation of the voice production system as a lumped element model with discrete time representation, allows for the application of sample-based methods used for system identification.[14,83,92] Hence, to produce a time-varying Bayesian estimator it is convenient to rewrite the system as a time-dependent unit. Consequently, we define the following system

$$\mathbf{Pr}\left(X_{k+1}\right) \sim f_{X|\mathbf{X},U,\Theta,\eta}\left(X_{k+1}|X_{1:k},U_k,\Theta_k,\eta_k\right), \tag{4.3.1}$$

$$\mathbf{Pr}\left(Y_k\right) \sim f_{Y|\mathbf{X},\Theta}\left(Y_k|X_{1:k},\Theta_k,\nu_k\right), \tag{4.3.2}$$

In this representation, the inclusion of the measurement noise ν_k is considered natural, since all practical measurements are affected by this type of error. The state error η_k is, as before, the representation of the non-modeled part of the system. All numerical models are a representation of real physical system. Hence, the inclusion of state noise to reproduce a true physical phenomena is a vital part of the numerical model.

In Bayesian inference, there are several techniques that merge the estimation of state X_k and observations Y_k together, thus allowing a thorough analysis of their evolution[13,14] on a sample

by sample basis. The combination of the state-space model with the Markov model properties, produce a HMM representation[13,14] that describes the evolution of a state-space model on a sample-by-sample basis and with dependency only to the last state. In the HMM framework, a series of stochastic filters can be applied to estimate the state X_k from the observations Y_k, particularly when the system parameters Θ_k are known (e.g., Kalman filter, particle filter, etc.). However, when the parameter set is unknown, the estimation of the state is not direct and generally requires additional system manipulation. One of the alternatives to estimate the state with unknown parameters, is the extension of the state model to include the distribution of the parameter Θ. In this way, an extended state $\tilde{X}$ is produced which is an ensemble of X and Θ. Herein, we employ a particle filter method[14] to estimate the ensemble $\tilde{X}$. This estimation scheme exploits the properties of sequential Monte Carlo techniques, by propagating a series of iid point masses of the state through the system.[14] Later on, these point masses (referred to as "particles") will be used to approximate the *prior* observation of the system. After comparing the *prior* observations with real measurement, the likelihood of measurement given augmented state is obtained. Therefore, using Bayes' theorem, it is possible to approximate the state *posterior* pdf using the *prior* information given by the state particle distribution. Thus, thanks to the expanded state distribution $\tilde{X}$, the particle filter will approximate the state X and parameters Θ simultaneously. To simplify our notation, future references to the extended state $\tilde{X}$ will be left as the normal state X since the estimation process is invariant for both normal and extended states.

The basic principle of the particle filtering is producing the swarm sequence of point-mass estimates of the state at time step k. As explained before, these point-mass estimations of the state are referred as particles, and are samples of the state distribution X_k. Loosely, each particle can be considered a realization of an *iid* random variable. One of the most common ways to ensure the correct distribution of the state particles is through the repeated application of the Bayes theorem using Monte Carlo sampling.[14,56] This procedure can be accomplished via sequential importance resampling algorithm,[52] which is also called SMC method.

Assuming that the prior state is sufficiently well defined to properly approximate the real state-distribution, the sequential Monte Carlo (SMC) algorithm allows to approximate the state pdf evolution, by comparing the observation prior pdf, with the current measurement, and assigning a weight to its likelihood. After the weight are calculated, a resampling algorithm can be used to reproduce the particles that are more likely to be correct and discarding the ones that are less likely. This procedure, including both sampling and resampling steps, allows for the approximation of the *posterior* state distribution on every time step k, using the prior knowledge of the state, the state-model evolution, and the sequence of measurements from the beginning up to the k-th time-step.[52,138]

Intuitively, at the k-th time step, the SMC method allows to predicts the state X_k from the evolution of N_p independent samples of the pdf of X_{k-1}. The corresponding observations Y_k of the sampled state X_k are compared with the current measurements, obtaining a likelihood weight distribution $\hat{W}_k$ that allows for the update of the state distribution $\hat{X}_k$ (state *posterior*). These values are then propagated to the next time step and the procedure is repeated. The particular implementation of the particle filter used in this research corresponds to the sequential Markov

chain Monte Carlo (SMCMC) technique,[188] which allows to sample the state distribution using a random walk process. As a result, this approach yields a better sampling of the most probable values of the state distribution. The complete algorithm for SMCMC is described below:[13]

1. Initialize:

$$x_0^i \sim f_{X_\mathrm{o}}\left(X_\mathrm{o}\right) \tag{4.3.3}$$

$$w_0^i = \frac{1}{N_p} \tag{4.3.4}$$

$$i = \{1, 2, \ldots, N_p\} \tag{4.3.5}$$

2. Importance sample:

$$x_k^i \sim f_{X_k|X_{k-1}}\left(x_k|x_{k-1}^i\right) \tag{4.3.6}$$

3. Likelihood:

$$w_k^i = f_{Y|X}\left(y_k|x_k^i\right) \tag{4.3.7}$$

4. Weight normalization:

$$\hat{w}_k^i = \frac{w_k^i}{\sum_{i=1}^{N_p} w_k^i} \tag{4.3.8}$$

5. Resampling:

$$\hat{N}_\mathrm{eff} = \frac{1}{\sum_{i=1}^{N_p}\left(w_k^i\right)^2} \tag{4.3.9}$$

if $\hat{N}_\mathrm{eff} \leq N_{thresh}$ resample $\hat{x}_k^i \overset{\hat{W}_k}{\Longrightarrow} x_k^i$ otherwise, accept particles $\hat{x}_k^i = x_k^i$,

6. Diversification acceptance probability

$$d_i = \min\left(\frac{f_{Y|X}\left(y_k|\tilde{x}_k^i\right)}{f_{Y|X}\left(y_k|\hat{x}_k^i\right)}, 1\right) \tag{4.3.10}$$

7. Diversify

$$\tilde{x}_k^i = \hat{x}_k^i + \epsilon_k^i \tag{4.3.11}$$

where $\epsilon_k^i \sim \mathcal{N}\left(0, R_{\epsilon\epsilon}\right)$

8. Diversification decision

$$x_k^i = \begin{cases} \tilde{x}_k^i & \text{if} u_k < d_i \hat{x}_k^i \\ & \text{otherwise} \end{cases} \tag{4.3.12}$$

where $u_k \rightarrow \mathcal{U}(0,1)$

The implementation of the SMCMC algorithm has an additional resampling decision, based on the dispersion of the particles (step 5). In addition, the diversification process allows to sample the unknown distribution of the state using a random walk procedure,[18] improving the estimation of the *posterior* due to the denser sampling of the most likely values of the state.

To illustrate the particle filter behavior, as before, two different simulations were performed with stationary and non-stationary muscle activation. In these simulations, the Bayesian estimation framework was employed to infer vocal fold model parameters from synthetic observed data. The measurements used for this purpose were the minimum glottal area[146] (minimum area projected from the membranous glottal area as viewed from an endoscope), and the glottal flow, both which generated by the TBCM described in Chapter 3. In practice, these signals can be obtained through the use of HSV[33] and flow inverse filtering estimation methods.[119, 127]

Given that the TBCM is implemented using a Truncated Taylor Series approximation,[51] special care must be placed to avoid the so-called "inverse crimes" in which the observed data is absent of noise and is generated with the same model.[84] To avoid such problems, two sources of uncertainty and unfitness were considered: (1) the vocal tract used to generate the observations signals was not included in the estimation model, and (2) the observed measures were contaminated with additive Gaussian noise of zero mean and SD of 5% of the synthetically generated observations. In practice, these modifications to the model and signal are expected to produce enough distortion and uncertainty to resemble clinically acquired measurements.

To avoid transient problems of the model stabilization, an onset time of 200[ms] to reach steady state was discarded in every simulation. In addition, to avoid the same problems in the transient of the estimation, a settling time of $10[ms]$ was set to propagate the particles. In this process only the initial state distribution was propagated recursively using no resampling or weighting in the algorithm. These two modifications dramatically improves the behavior of the particle filter by increasing accuracy, reducing uncertainties, and reducing the convergence time of the estimation process.

For the following estimation example, the CT muscle activation (a_{CT}) is treated as an unknown parameter.[173] In a similar way, the incident acoustic sub and supra-glottal wave pressures[154] are considered independent variables of the system state, acting as unknown inputs of the system. Given that CT muscle activation is naturally bounded between zero and one, the prior distribution of this parameter was chosen to be a uniform distribution over that entire range. On the other hand, the incident acoustic glottal pressures were normally distributed to fit all the possible pressure values achieved in model simulations simulations (See Table 3.2 on Chapter 3). This particular distribution choice was motivated by the Principle of Maximum Entropy,[26, 79, 80] which states that a *prior* should only reflect the state of testable information. The proper choice of *prior* distribution can often fill in some missing information. However, the efforts to reduce uncertainty on the *prior* distributions should not exceed the information that is certain. Otherwise, the *posterior* estimation could be biased by a subjective *prior* belief or expectation (also referred as to "self-fulfilling prophecy").

The two simulations conducted were: (1) a sustained phonation on a constant CT muscle activation of 0.3[-], and (2) CT muscle activation varying on equal segments of time, using values

of 0.3, 0.5 and 0.1, where each stage had equal time-length and transient-time. The remaining model parameters were considered deterministic and known. The values of such parameters are presented in Table 3.2 of Chapter 3. This simulated scenario, will help to understand the behavior of the estimation process, wherein two simultaneous observations are being used to estimate multiple unknown time-varying signals (state values, system parameters, and observations). The same estimation process will be used in Chapter 5 for a clinical setup, in an effort to assess the concept of subject-specific model estimation with clinical data.

Results for the stationary scenario

The stationary scenario is designed to test the estimation of CT muscle activation during a normal sustained vowel gesture. This means that the muscular activation should remain constant along the entire time span. Here, 200[ms] of stationary simulated data were obtained as subject-specific measurement from the TBCM using time-invariant control parameters. The first 30[ms] (after the transient period) of observed measurements are presented in Figure 4.3. The true observed glottal area in Figure 4.3(a) and the measured glottal flow in Figure 4.3(b) are presented in solid blue lines. These references are contrasted with the resulting estimates from the proposed procedure in solid red for the weighted mean and dotted red for the credibility intervals of 95%. It is important to note that the observed glottal area does include the posterior glottal opening (PGO) area and not only the membranous portion of the glottis, producing and offset in the observation and estimation.

To obtain these results one hundred particles were used in the estimation. Each particle consisted of the position and velocity for each vocal fold mass, the incident and departing pressures from the glottis, and the muscular activation of the larynx. The true value and posterior estimate of CT muscle activation is presented in figure 4.3(c), where the convergence of the estimator is clearly observed. The initial off-value estimation is explained by the transient behavior of the estimator, where there is not enough data to produce a valid estimation.

Given that the proposed estimation captures the behavior of the complete time-varying simulated system, additional hidden measurements can be computed, such as, the contact forces during the collision of the VF. Figure 4.3(d) illustrate this idea through the force applied to the vocal folds during the contact with the opposite vocal fold. To obtain this measure, the sum of the contact spring forces in both masses was calculated. However, it has been suggested that the relation of contact pressure is more relevant than the contact force[60, 160] since it is believed to be a better descriptor of potential lesions. It is important to note that this measure is extremely difficult to obtain in clinical setups,[82] since the measuring process is highly invasive and it will disrupt the system behavior, yielding an altered measurement.

Results for the non-stationary scenario

In contrast with the previous simulation, a variation in the configuration parameters was performed aiming to modify the output. Here, the CT muscle activation (a_{CT}) was varied between 0.3, 0.5 and 0.1 with equal transient periods. This variation in the muscle activation is meant to emulate pitch changes observed in different vocal gestures (e.g., pitch glides and glissando). This approach allows for the evaluation of the estimator during transient periods, which is a prelude

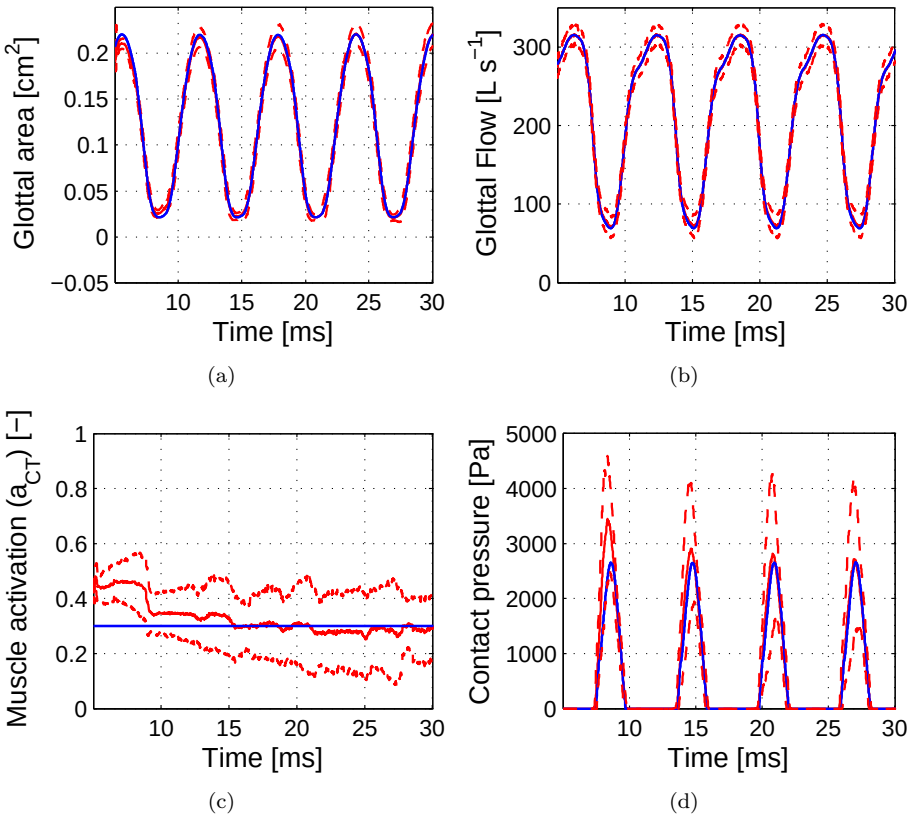


Figure 4.3. Simulated measurements and estimations for a stationary scenario using non-stationary method: (a) Minimum glottal area, (b) glottal flow, (c) Cricothyroid muscle activation, (d) Vocal folds contact pressure. [Legend: (Solid) measurement/target, (Solid) mean estimate, (--) 95% credibility]

to a more complex analysis of the subject-specific modeling.

As before, the measurements used in these cases, were the minimum glottal area projected from the membranous portion, and the glottal flow obtained at the glottal output, being both signals synthetically generated by the TBCM as described in Chapter 3.

In Figure 4.4(a) and (b), 200[ms] of glottal area and flow are presented. As before, the true values are shown in solid blue with its respective estimations in red (solid for mean and dotted for credibility intervals). However, given the large amount of data and the close estimates, it becomes difficult to separate the true value from the estimated one. In Figure 4.4(a) and (b), the variation of glottal area and flow is observed as the CT activation changes over time. The time variation of this activation and the resulting estimates are shown in Figure 4.4(c). For this figures, the advantages of using a non-stationary estimator on a non-stationary problem (in contrast with the results in the previous section) become evident. It can be observed in Figure 4.4(c) that this allows for the estimation of a small variation on the phonatory process, which is particularly important for clinical analysis.[12, 147]

Figure 4.4(d) shows the variation of the contact forces (solid blue) and the estimation obtained with the SMCMC method (solid red for mean and dotted red for credibility intervals). 4.4(f) show the last 30[ms] of estimation. It is appreciated, in Figure 4.4(d) and (f), how the estimation process follows the changes in the estimation of contact pressure, and how accurately estimate each cycle of contact pressure. In addition, comparing results in Figure 4.4(c) with previous results in Figure 4.3(c), it becomes clear that the SMCMC algorithm provides a better time resolution than the stationary estimation. It is important to note that mean estimates of contact pressure match within the ranges observed in clinical setups.[41]

Figure 4.4(e) presents the pdf of the mean estimate of CT muscle activation for all the simulated time. It is appreciated how the distribution presents multi-modal characteristics, which is produced by the three different values of activation used in the simulation. This is one of the main advantages of using stochastic estimation rather than deterministic methods, since the later would probably estimate only one of those three maximum of the posterior distribution.

To provide a complete comparison between the stationary and non-stationary methods, it is necessary to account for the computational cost associated with each estimator. The stationary estimation requires to generate a set of *prior* observations to be later used in the likelihood of the observed measurements. Each *prior* sample is synthetically generated by the model using a combination of model parameters, and storing the results for the Bayesian processor. This requirement force to have a conditional distribution of the observation for each combination of the stochastic parameters. To generate such set of data, a total of $N_{sam} = (N_{par})^{N_{Par}}$ simulations are required, where N_{par} is the number of samples of each parameter, and N_{par} is the number of parameters used. This number of simulations, however, can be can be conveniently adjusted using importance sampling techniques to concentrate the most probable values, thus reducing the number of samples required for each parameter. Over all, the total computational cost of the stationary method is the number of samples required for the method, times the cost associated to the computation of each simulation.

On the other hand, the particle filter method use the same amount of particles N_{par} for each iteration, but it does not requires to multiply those samples for each one of the stochastic

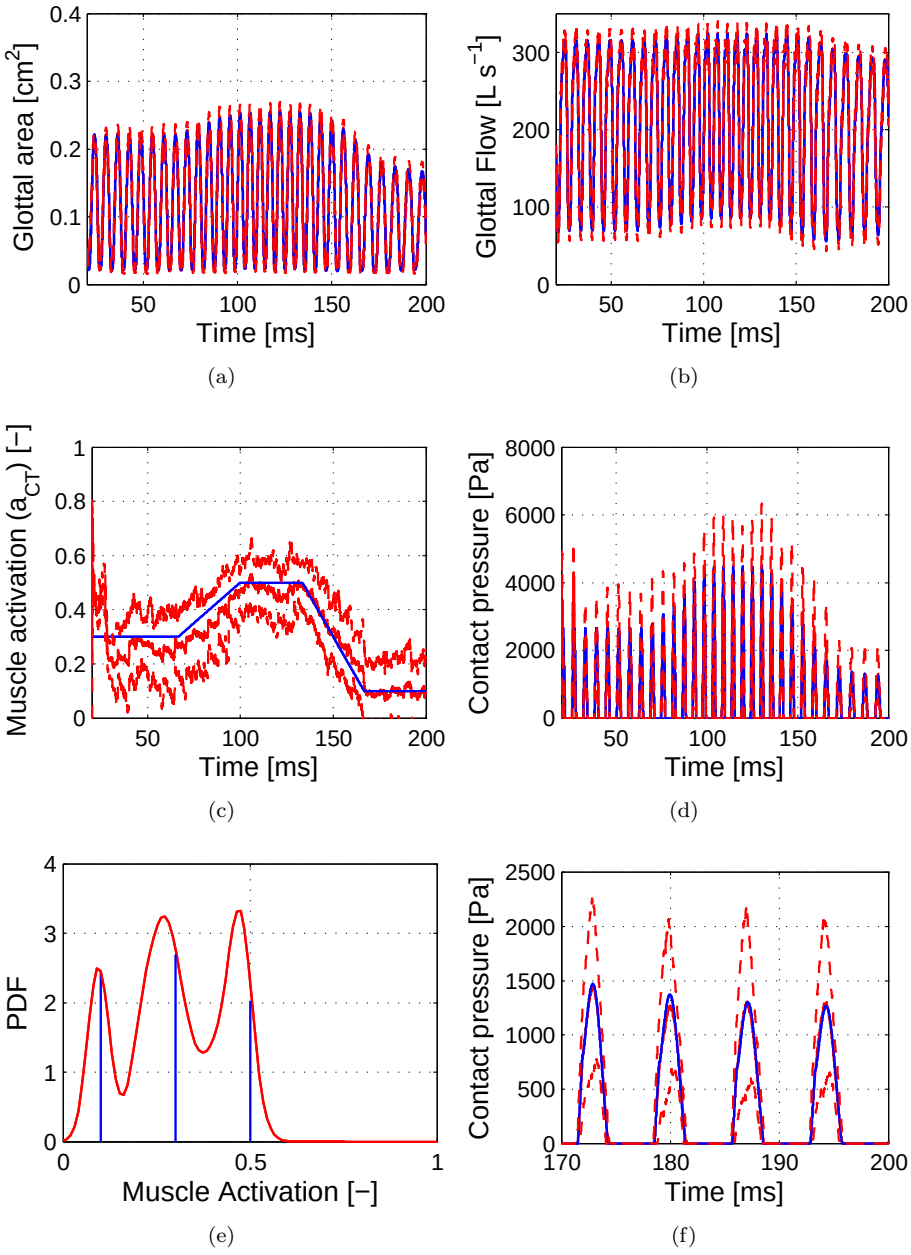


Figure 4.4. Simulated measurements and estimations for a non-stationary scenario using non-stationary method: (a) Minimum glottal area, (b) glottal flow, (c) Cricothyroid muscle activation, (d) Vocal folds contact pressure, (e) complete muscle activation estimation pdf, (f) Zoom in of 30[ms] of contact pressure. [Legend: (Solid) measurement/target, (Solid) mean estimate, (--) 95% credibility]

parameters. Therefore, the total computational cost of the particle filter is proportional to the number of particles used in the estimation, times the computational cost associated to a single simulation. In consequence, for a large number of unknown parameters, the computational cost required for the particle filter is considerably less than the required for the stationary estimation. However, the stationary estimation only requires to compute the *prior* once, independent of the number of cases to be estimated. While the particle filter requires to perform each estimation case independently.

The proposed non-stationary estimation provide an important advantage over other estimation methods, allowing to compute additional measurement with almost no increase of computational cost. Given that the SMCMC method allows to estimate the system evolution for each determined time step, it is possible to compute related measurements from the underlying estimated model (e.g., contact forces). This additional information can be treated as a virtual sensor signal, thus increasing the clinical application of the proposed method.

Chapter 5

SUBJECT SPECIFIC MODELING FROM CLINICAL DATA

In this thesis we aim to develop a subject-specific model using clinical recordings. For this purpose, the numerical model was designed to produce similar signals to the ones obtained in a clinical setups, namely glottal flow, observed vocal folds (VF) kinematic, and radiated acoustic pressure. In addition, the model allows for obtaining signals that are difficult or impossible to obtain on clinical setup (e.g., VF collision forces, muscle activation parameters, etc.). The ability of the VF models to produce these unobservable signals, has been referred in this thesis as the possibility to create "virtual sensors" for an enhanced clinical assessment of vocal function.

Chapter 2 presented the bases of phonation theory and system identification. In Chapter 3, the proposed numerical model of VF was defined, and later utilized in Chapter 4 with two different Bayesian estimators. In the current chapter, a proof-of-concept is performed using the proposed sequential Markov chain Monte Carlo (SMCMC) Bayesian estimator applied to clinical data. From the resulting subject-specific model, the cricothyroid (CT) muscle activation, the collision contact pressure, and the microphone signals are extracted. As a compared-analysis step, the predicted microphone signal through the model is compared to the clinical signal. A complete synthetic case is also analyzed to provide further insights on the uncertainty of the proposed scheme.

5.1 Methods

It has been shown in Chapter 3, we developed a triangular body-cover model (TBCM) that properly describes the physiology of normal VF, and that can also emulate the pathophysiology of vocal hyperfunction (VH). In this chapter, we utilize the TBCM to produce a subject-specific model since it will be applied to a case with no VF lesions. The TBCM (described in Chapter 3) consist of a lumped element model of the VF with a triangular glottal shape, and posterior glottal opening produced by the arytenoid posturing. To estimate the VF model parameters we use the SMCMC described in Chapter 4, which is a Bayesian estimator with non-stationary assumptions.

For the purpose of method assessment, two different scenarios are explored: (1) a case study from clinical data, and (2) a synthetic case with simulated data. The idea is to investigate the same scenario with the proposed identification process, to allow for an analogy between the

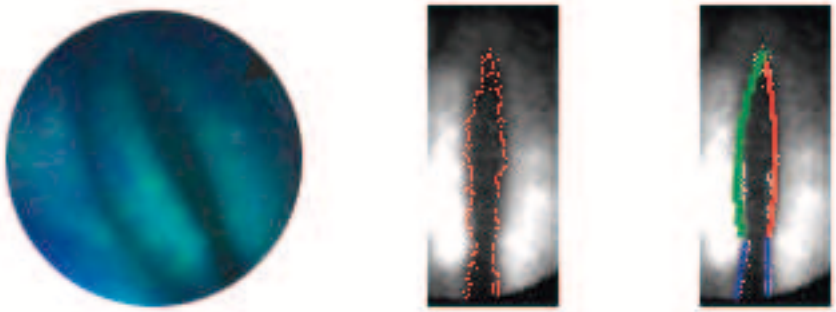


Figure 5.1. Glottal edge extraction, from right to left: original frame, edge detection, polynomial fitting

known synthetic system and the clinical results.

The clinical data used for the estimation process was obtained from a laryngeal flexible high speed videoendoscopy (HSV) recording at 4000 [FPS] at a resolution of 128×128 pixels on a 8bit RGB (Photron SAx2 camera). The HSV was specially calibrated using a green-dot laser-grid projected into the VF during the recordings. In addition, a set of aerodynamic and acoustic signals were simultaneously acquired in synchrony with the video signal. The glottal flow signal was obtained through an inverse filter[46] of oral flow obtained through a Rothenberg mask (model MA-1L, Glottal Enterprises), and an audio signal was acquired through a microphone (Sennheiser electronic GmbH & Co. KG, model MKE104) at a sampling rate of $100[kHz]$.

The data was obtained from a normal male subject with no history of hyperfunctional behavior or vocal pathology. In a pre-processing step, a portion of the video was manually selected according to the duration of a voiced gesture. Up to our knowlegde, this is the only HSV recording with simultaneous recording of flow, microphone and calibrated laser projection. Prior efforts have used separate recordings of these signals.[100, 194] Once the voiced segment is selected, an automated process of brightness and gamma correction was performed. Then, under the assumption that the larynx is stationary during the hole segment duration, an average image was obtained from the video-segment. Assuming that no considerable arytenoid movement was present, the averaging process allows to identify the arytenoids edges and the first estimate of the posterior glottal opening (PGO) area. During this averaging process, the cartilaginous area and the anterior commissure were manually marked by a trained voice therapist. This allows for a better definition of the glottis and the posterior glottal edge displacement.

Subsequently, each frame was rotated to align the glottal-midline with the vertical center of the frame. After the rotation, the resulting frame was cropped to fit the maximum glottal displacement, thus reducing the influence of visual artifacts, and improving the behavior of VF edge detection algorithms. After the frame and video adjustments, the RGB video was converted into a monochromatic video which is used to detect edges based on standard detection methods (e.g., canny, gradient operators, etc.). The edge information of each frame was fitted into a second order polynomial curve, and pinned between the anterior commissure and the vocal process. The process of fit a polynomial curve removes several outliers of the edge detection stage and allows for further analysis of the glottal behavior. From the polynomial curves, it was possible to obtain the membranous area, while the initial posterior glottal area was obtained from the cartilaginous edge information. It is important to note that even when the glottal area obtained from the video is inherently asymmetric, the current study assumes symmetric behavior. Therefore, the small asymmetries present in the video are forced to be symmetric by correcting the position of the edges while maintaining the corresponding area (see Figure 5.1). Even when the symmetry correction is small in healthy VF, the correction will introduce "measurement error", that affects the statistics of the Bayesian estimator.

From the video pin-points and the area signal, the position of the arytenoids was extracted and used to define the onset condition of the model. The rotation and displacement of the arytenoid cartilages were assumed to be constant along the whole video segment. However, given the possibility of incorrect measurement, and/or a PGO area partially hidden from the top view, the rotation and displacement of arytenoids were also considered stochastic parameters with a process

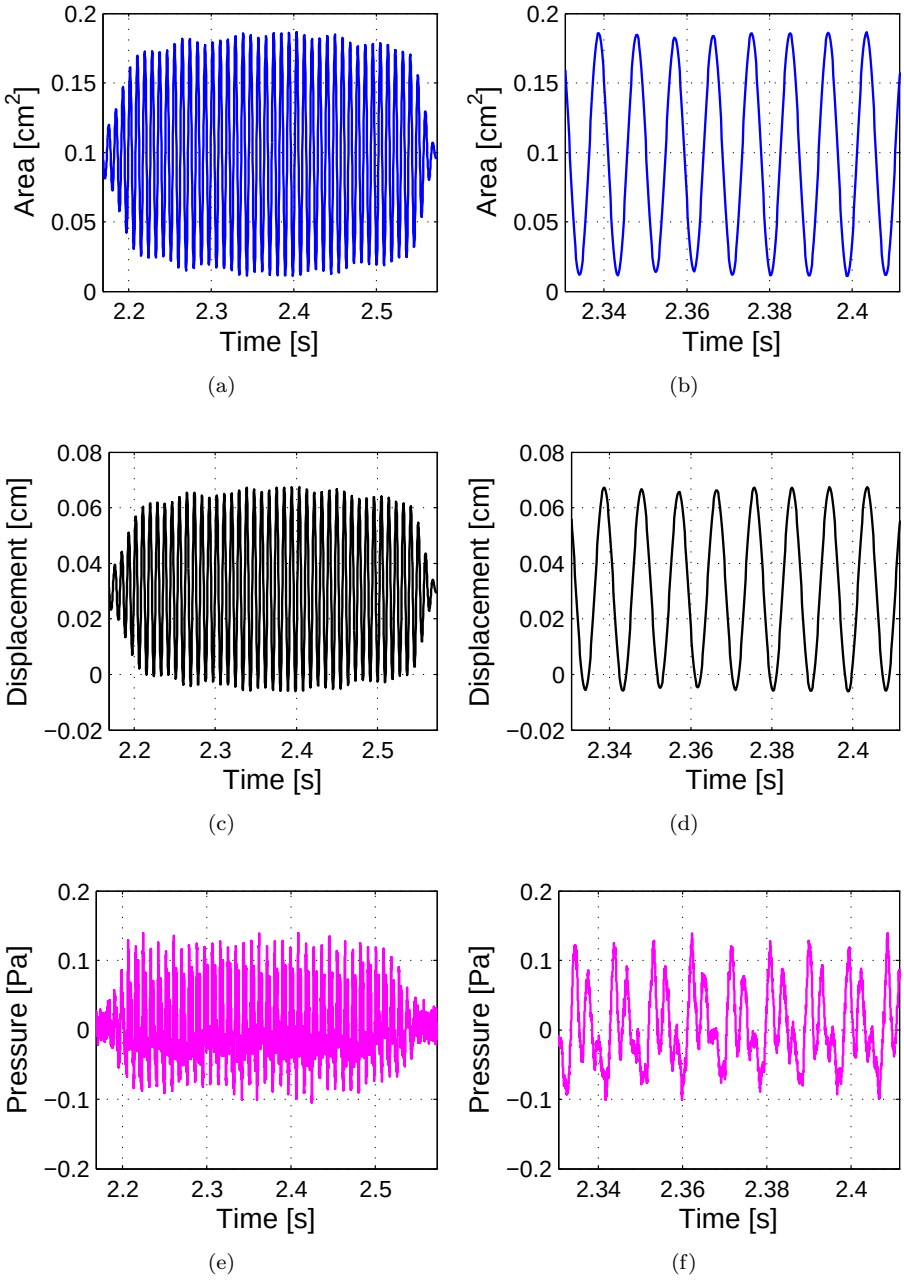


Figure 5.2. Time series obtained from calibrated clinical measurements. (a) Glottal area from video segment, (b) glottal area zoom in a 0.1[s] window, (c) left edge displacement from video segment, (d) left edge displacement zoom in a 0.1[s] window, (e) microphone signal from synchronized segment, (f)) microphone signal zoom in a 0.1[s] window

noise as presented in Table 5.1. This allows for a small correction of the parameters, yielding an auto-adjustment related to the observed measurements. Using the membranous glottal area, the measured posterior glottal displacement (PGD), and the glottal length, the kinematic model parameters can be estimated.

The resulting area, edge displacement, and microphone signal can be represented as the time series shown in Figure 5.2, where a */pi/* gesture was recorded. For the comparison of the two cases (simulated and real) only the stationary (steady state) part of the signal will be used.

The fixed and unknown parameters used in both cases are presented in Table 5.1. The underlying parameters to be estimated are: (1) sub-glottal pressure, (2) CT muscle activation, (3) arytenoid rotation, and (4) arytenoid displacement. An additional adjustment was made to the default model presented in Chapter 3: Using the rate between the default arytenoid length[86] and the visible length from the video, a proportion factor was obtained and used to scale the default model parameters, modifying the rest length, thickness, and depth of the VF.

The HSV, the aerodynamic and acoustic data, and the estimation process run at different sample rates, thus producing an alteration of the conditions simulated in Chapter 4, where all the data was sampled at the same time and at the same rate. To overcome such asynchrony, the estimation algorithm used a re-sampling approach to match the model sampling rate. To overcome physiological and electronic delays, all separate signal were later time-adjusted using a cross-correlation technique in a small time-delay neighborhood. Another possible approach to the sampling problem is to assume a lossy channel, updating the observation probability only when a valid data is acquired. However, the big difference between sample rates indirectly give more importance to observations with higher sampling frequency. Given that the HSV is the main source of information and has the lowest sample rate, the lossy channel approach gives the HSV the lowest importance, thus not pondering its real relevance.

5.2 Synthetic case

As previously mentioned, a synthetic case was obtained through simulation using the TBCM configured to hopefully mimic the conditions observed in the clinical setup. This procedure is meant to test the exactly same algorithm used in the clinical setup, but with a set of known parameters. To accomplish this goal, some manipulation of the data was necessary to match

Table 5.1. Clinical model parameters used in the estimation process

Parameters	**Real**	**Simulated**	**Noise Std**
Sub-glottal pressure (P_s)	Unknown	900 [Pa]	10[Pa]
Cricothyroid muscle activation (a_{CT})	Unknown	0.15 [-]	0.02[-]
Arytenoid Rotation (a_r°)	Unknown	-6.38 [°]	0.02 [°]
Arytenoid Displacement (a_r^d)	Unknown	1.8 [mm]	0.02 [mm]
Thyroarytenoid muscle activation (a_{TA})		0.18 [-]	
Lateral cricoarytenoid muscle activation (a_{LC})		0.5 [-]	
Supra-glottal pressure (P_e)		0 [Pa]	
Arytenoid length (ℓ_a)		13.20 [mm]	
Gender		male[173]	
Vocal tract		Takemoto /i/[157]	

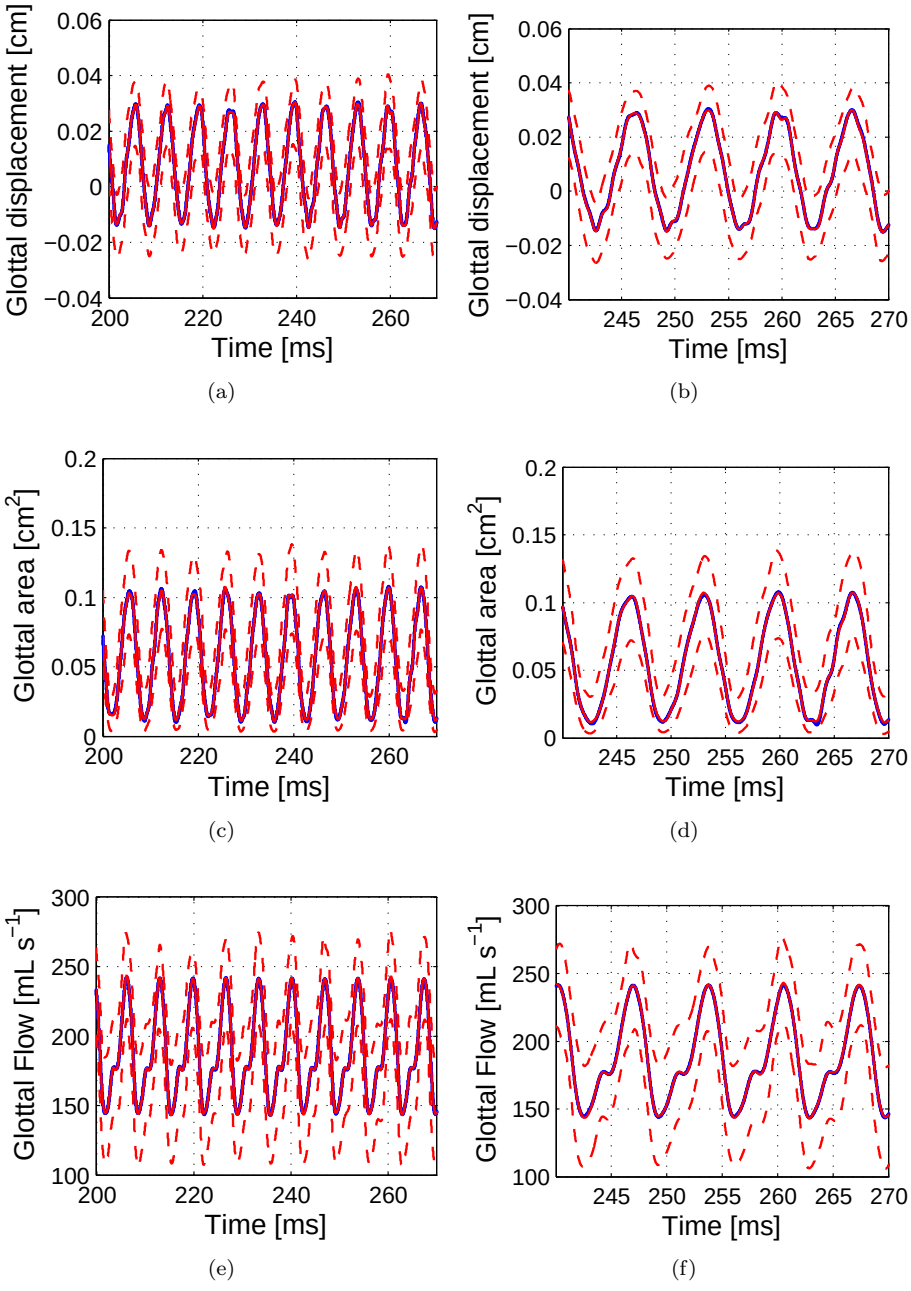


Figure 5.3. Synthetic observations: Displacement (a) time series, (b) zoom in 30[ms]. Glottal area (c) time series, (d) zoom in of 30[ms]. Glottal flow (e) time series, (f) zoom in of 30[ms]. [Legend: (Solid): Measurement, (Solid): MAP, (--): 95% credibility]

the different sampling rates observed in the clinical setup. The glottal area and the glottal displacement were decimated from 70[kHz] to 4[kHz] and contaminated with Gaussian noise of zero mean and standard deviation (SD) equal to 10% of the signal SD, while the glottal flow and microphone signals were interpolated from 70[kHz] to 100[kHz] and also contaminated with a Gaussian noise of zero mean and SD equal to the 10% of the raw signal SD. This interpolation step, however, will have minor effect in the estimation process, since it only uses a single sample per observation at any given time.

One of the objectives of this chapter is to validate the use of clinically acquired data with the proposed Bayesian technique. The use of a synthetic simulation with known parameters allows for directly compare estimation results, providing a first assessment of the estimator configuration, to be later used with clinically acquired data. For this reason, and as a proof of concept, not just the control parameters of CT muscle activation, and sub-glottal pressure were used. Additional outputs of microphone signal, and contact pressure, were also computed and compared to the reference values.

Figure 5.3(a) shows the observed edge displacement as seen in a endoscopic exam, i.e., the visualization of the minimum displacement measured from the glottal midline. It can be observed from Figure 5.3 that the displacement is sometimes negative, an impossible phenomena in the body-cover model (BCM) due to its parallel glottal shape configuration. The negative displacement in the TBCM is a product of the method used to calculate the displacement. Considering symmetric versions of the models, the use of strictly parallel masses in the BCM along with negative displacement, necessarily implies complete membranous glottal closure. Therefore, there is no negative displacement without complete membranous glottal closure. On the other hand, the TBCM does not require to have complete membranous closure on negative mass displacement. Thus, allowing negative displacement while maintaining an open glottis. It is important to clarify that the negative displacement observed in numerical models does not implies physiological overlap of the vocal folds, and it is only used to represent the restoring force and energy transfered during collision.

Figure 5.3(c) shows the observed membranous glottal area, which for the symmetric case is obtained from the minimum area between the lower masses and the upper masses. It can be appreciated the effect of including a PGO with a linked membranous glottal opening (MGO), which produces an offset in the area signal.

The main advantage of Bayesian estimation when compared with deterministic methods can be appreciated in Figure 5.3, where not only the most probable value is obtained, but also its credibility intervals. This credibility bands cannot be obtained in a deterministic framework. The maximum a *posteriori* (MAP), which is the value with maximum probability after the estimation *posterior* is calculated, is presented as a solid red line. The 95% credibility bands are illustrated in a dotted red line, and the true measurements are presented in solid blue. It can be appreciated that the proposed estimation method successfully follows the evolution of the synthetic measurements, were all the true values are contained within the estimated credibility bands.

Figure 5.4 presents the probability density function (pdf) of the parameters that were considered time-invariant, along with the signal of two hidden measurements. The parameters

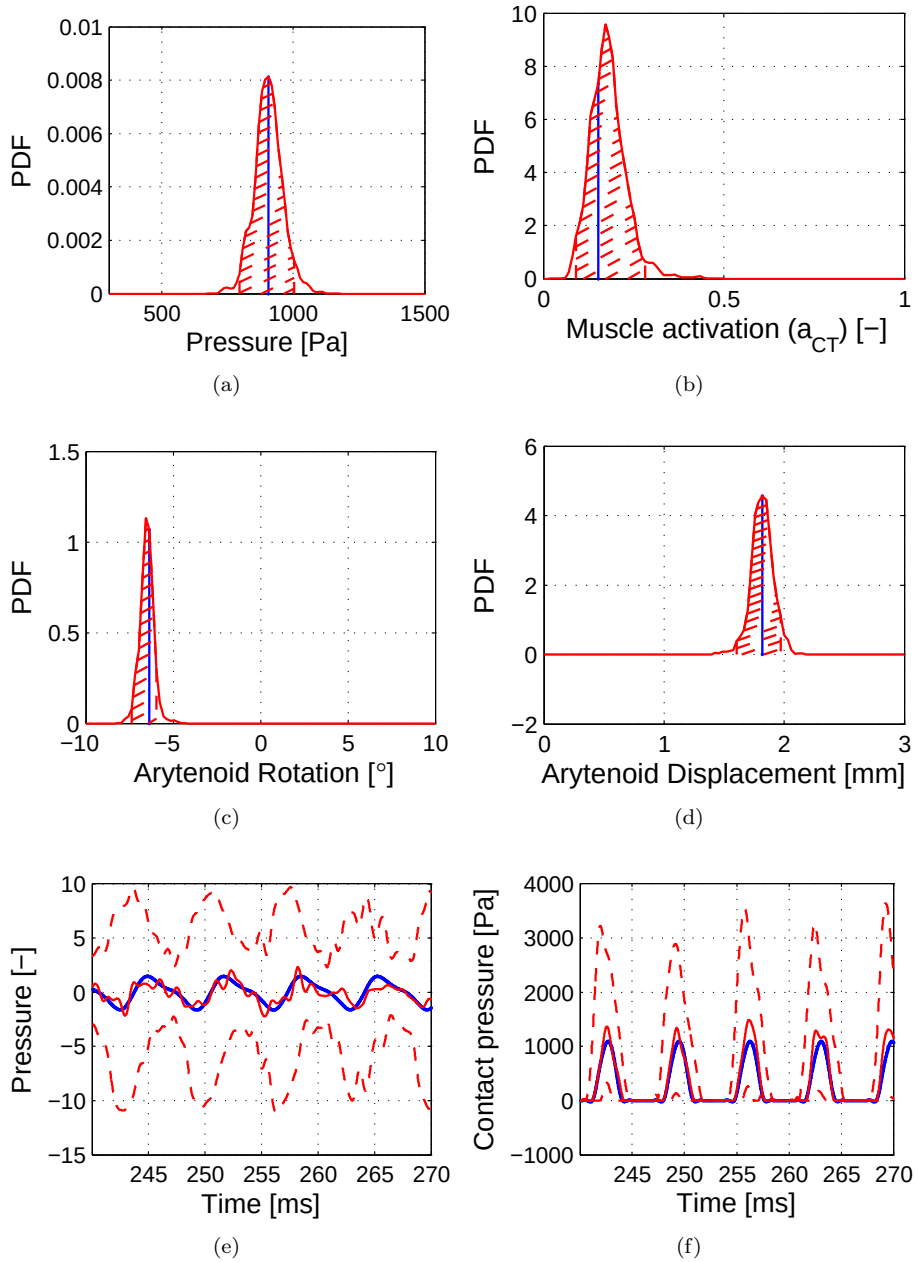


Figure 5.4. Synthetic estimation: model parameters and "virtual sensors". (a) sub glottal pressure, (b) CT activation, (c) Arytenoid rotation, (d) Arytenoid displacement, (e) Microphone Signal, (f) Contact Pressure. [Legend: (Solid): Reference, (Solid): Estimation, (--) : 95% credibility]

distribution, the 95% credibility region, and the true value are shown. It is noted that the true value is within the credibility range, and that the MAP is also a close approximation of the target value. In addition, it is important to mention that while the parameters are assumed to be time-invariant, they are not fixed values and they are not time constants. These time-invariant parameters, are state variables with a unitary state function, producing a constant value with process noise presented in Table 5.1.

Figure 5.4(e) and (f) shows the resulting observations of non-stationary outputs. This means that the evolution mechanism of those values have associated a transfer function different than unitary, therefore configuring a Markov chain model. In this figure, we see the estimation of the microphone signal and vocal fold contact. Given that the current scenario is synthetically generated, the reference values of both signals are shown (solid blue lines). Remarkably, these signals were obtained indirectly from the estimation process, i.e., no direct relationship exist more than the modeling approach. Figure5.4, also shows that the MAP clearly follows the reference values, particularly for the contact signal, where the root mean square (RMS) error is approximately 240[Pa], which is less then a 20% of the maximum collision pressure.

One particular advantage of this method over other similar approaches[17,36] is the ability to obtain hidden signals. This is, for example, the ability to obtain the contact pressure generated during the collision of the vocal folds. This particular signal could provide important information regarding the pathogenesis of vocal lesions, and it is believed to play an important role in VH related problems.

5.3 Clinical case

As mentioned before, the synthetic scenario and the clinical scenario are meant to be the same. The configurations used in the synthetic case are used without modification for this case as well. The noise consideration in the estimation process, the sample frequency, and the model inputs and outputs are the same. The main difference between the synthetic case and the clinical case is the data source. In the synthetic case, the source data was generated by a numerical model, therefore knowing the target values. On the other hand, the clinical case data was obtained directly from the analysis of the recorded HSV, flow, and microphone signal. As stated before, the video was recorded at a rate of 4000[fps], and the data was recorded at a rate of 100[kHz]. Both sources of information where later re-sampled to produce a single sample frequency of 70[kHz], which is the sampling frequency of the particular implementation of wave reflection analog (WRA)[90,157] scheme used in the model.

Figure 5.5 presents the signals that were used as information source (measurements). In the same Figure the estimated posteriors of the observations are shown. It can be appreciated that, compared to the synthetic case, the match between the MAP and the measured output for the area and flow is similar to the synthetic case, but slightly worse for the glottal displacement. The reasons behind this differences can be diverse. Several model parameters were assumed to be deterministic (e.g., default vocal fold length, tissue density, air viscosity, etc.), where they may be stochastic. In addition, the system modeling carries a modeling error due to simplification of the real system. Further efforts will look into increasing the model complexity to address this point.

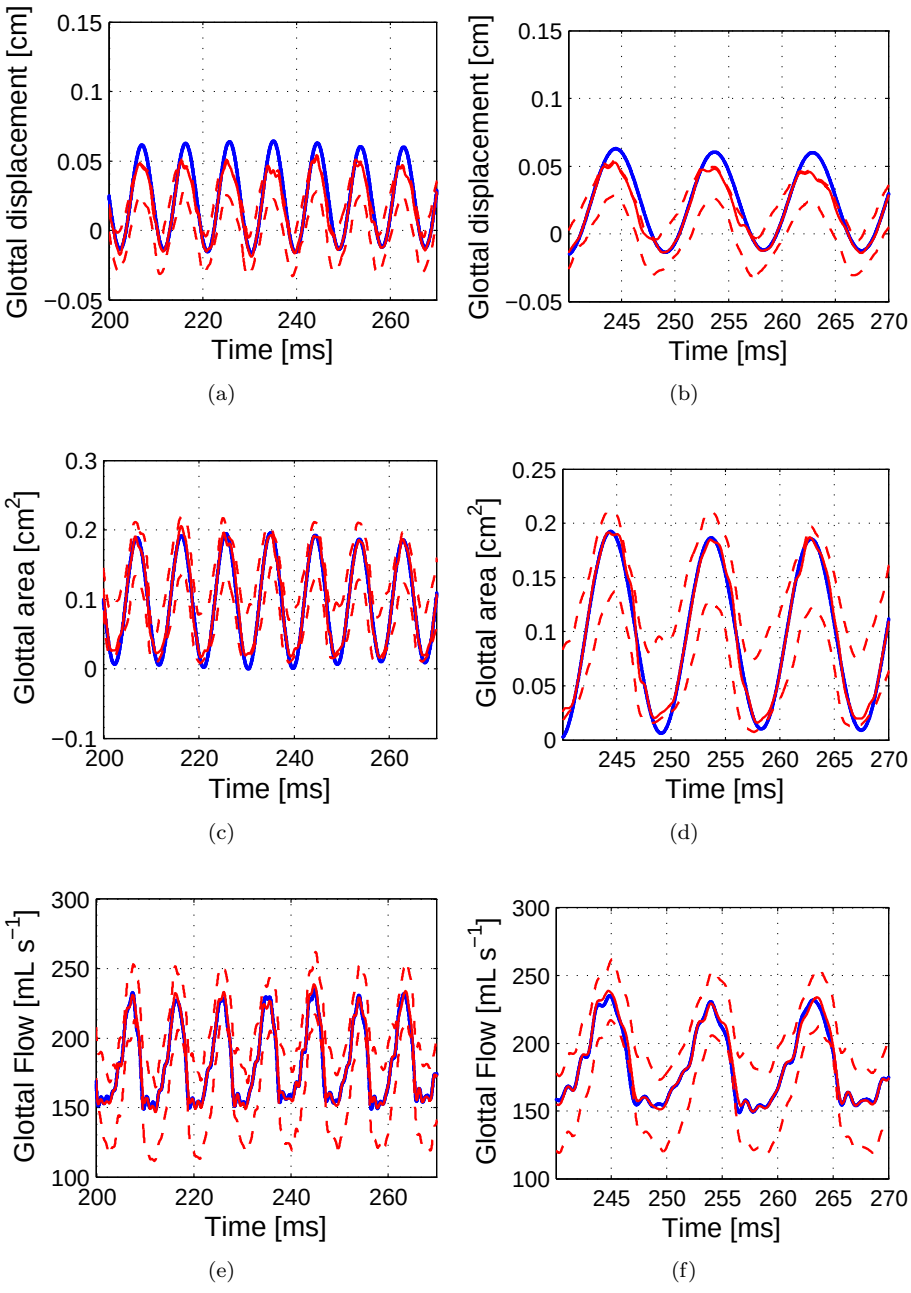


Figure 5.5. Clinical observations: Displacement (a) time series, (b) zoom in 30[ms]. Glottal area (c) time series, (d) zoom in of 30[ms]. Glottal flow (e) time series, (f) zoom in of 30[ms]. [Legend: (Solid): Measurement, (Solid): MAP, (--): 95% credibility]

However, the proposed estimation is still considered valid and more descriptive than deterministic methods, since these problems are model related, and are independent of the estimation method. It can be observed in Figure 5.5 that in almost all the sequence, the credibility bands contain the clinical observation, implying that the proposed estimation method does actually account for the observed value.

Figure 5.6(a)-(d) shows the stochastic time-invariant parameters of the models. Here, in contrast with the synthetic case, the targets are unknown, and only the pdfs with their respective credibility interval are presented. It is noted however, that not all the resulting estimates resemble Gaussian distributions. Figure 5.6(c) and (d) exhibit bi-modal characteristics, which can be a attributed to the PGO configuration. The posterior glottal area is the results of a two-degree of freedom system (arytenoid movement) with only one output (posterior glottal area). The ability to reproduce both modal responses is one of the advantages of the proposed method, allowing for a deeper analysis of the resulting estimates. Note that similar bimodal pdfs were obtained before for the same parameters in the synthetic scenario.

Figures 5.6(e) and (f) shows the signals from the "virtual sensors", where it can be compared the measured and estimated normalized microphone signal in Figure 5.6(e). The MAP estimate of the virtual sensor provides sufficient accuracy when compared with the measured signal, while the credibility intervals spans over the actual observed signal. Figure 5.6(f) does not have a reference value to compare the results. However, given the microphone signal comparison, and the simulated output result from the synthetic case, it is reasonable to assume that the estimated contact pressure spans over the actual contact pressure. In addition, it can be observed that the MAP estimates are in close relation with previous studies of contact pressures in vocal folds.[41] Consequently, we argue that the proposed estimator can provide information for further clinical analysis. From Figure 5.6, the large span of the credibility intervals is noted. These increased intervals are the product of uncertain measurements (e.g., low resolution video, measurement noise, calibration bias, etc.). The increased uncertainty in the observation produces, by the effect of sensitivity, larger uncertainties in the estimates.

5.4 Discussion

The aim of this chapter was to test of the proposed method using clinical data. Two sources of data were used with the same TBCM model. A test with clinical data was obtained through direct measurement of acoustic and aerodynamic signals, complemented with the acquisition of HSV. The synthetic source, on the other hand, utilized a data set that mimicked the clinical acquisition, with the additional advantage of knowing the true expected parameters.

The results obtained in this chapter allow for assesing the proposed estimation method on clinical data. Even though the results support the validity of the proposed method, it becomes evident that the model selection, the selection of the stochastic parameters, and the noise determination, are critical factors of the proposed Bayesian estimator. An incorrect model design will produce a deviation of the target anatomical system, yielding a representativity error. This types of errors will be translated into inaccurate estimation results, thus limiting the applicability of the proposed method. Incorrect selection of stochastic parameters produces alterations of the

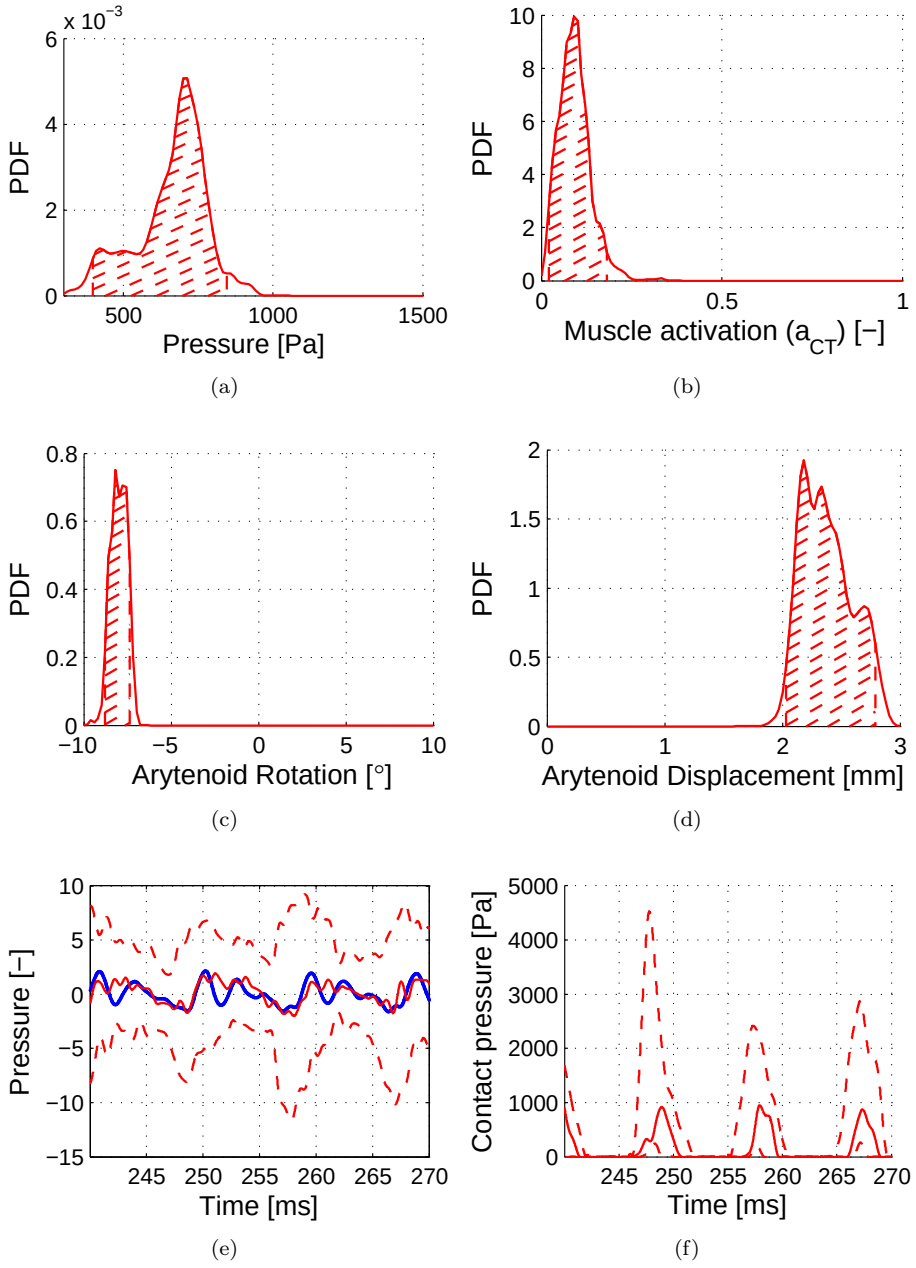


Figure 5.6. Clinical estimation: model parameters and "virtual sensors". (a) sub glottal pressure, (b) CT activation, (c) Arytenoid rotation, (d) Arytenoid displacement, (e) Microphone Signal, (f) Contact Pressure. [Legend: (Solid) Measurement, (Solid): Estimation, (--) : 95% credibility]

estimated parameter values. When a stochastic parameter is considered deterministic, the variation produced by the incorrect parameter selection will be translated as an overestimation of the remaining parameters, thus resulting in a misinterpretation of the data. As a trivial example, if the arytenoid displacement is considered deterministic, the PGO will be the only product of the arytenoid rotation. Therefore, any variation of the PGO will be attributed to the arytenoid rotation and not to the its displacement.

The noise selection is probably the most complex and least intuitive procedure in the estimation process. So far, for the voice model estimation, no automatic method has been developed. An incorrect assumption of the state and observation noise will produce incorrect estimation results. However, it is possible to select noise parameters by observing the behavior of the estimated model. Nonetheless, such approach would require parametric variations and therefore reducing the overall utility of the method. A comprehensive approach to estimate the noise in a more efficient fashion is pending.

While a good estimation can be assessed by comparing the credibility range and the observed signal, it becomes technically impossible to perform such analysis on non-observable data. Therefore, to further validate the method on hidden (non-observable) signals, it become essential to reproduce these results on a-fully observable environment, such as silicone models of the vocal folds.

All things noted, the Bayesian estimation allows for creating a subject-specific model of the voice production system, which is a new tool for a deeper model-based knowledge of the VF behavior. In addition, the creation of virtual sensors allows to "observe" time-varaying signals that are otherwise hidden from clinical setups. This additional information is expected to enhance the clinical practice.

In addition, the proposed estimation is not restricted to the TBCM. It can be applied for different model complexities. This characteristic allows for different levels of abstraction in the subject-specific modeling, allowing not only for the study of kinematic behavior, but eventually enabling multiple types of subject-characterization (e.g., simplified airflow models).

Chapter 6

CONCLUSIONS

Numerical models of voice production have been used to describe healthy and pathological phonation.[44] In general, these models have been designed to describe general trends rather than a single subject. Therefore, in this thesis it is hypothesized that a subject-specific model can provide additional relevant clinical information, and enhance the knowledge of the underlying physics. However, current techniques to construct subject-specific models of speech production are based on deterministic or stationary methods. The overall aim of this thesis work was to develop a subject-specific model of voice production, using a stochastic non-stationary estimator and test it with clinical data. If a stochastic subject-specific model is feasible, indirect model-related information can be drawn, thus increasing the knowledge about the vocal behavior, and yielding a tool for an improved clinical assessment. To develop such estimation technique, tree specific aims were set: (1) To improve the anatomical description of a well-stablished model of voice production while maintaining the low model complexity, (2) To develop a non-stationary Bayesian estimation on the proposed model to produce a subjec-specific set of model parameters, and (3) to assess the performance of the estimation process using clinically acquired data.

The body-cover model (BCM) is a numerical model of the vocal folds (VF) broadly used for the study of voice production and VF behavior.[44] The BCM however, makes several assumptions that hampers its ability to describe vocal hyperfunction (VH). The model was modified to include a posterior glottal opening (PGO) and a triangular glottal shape through the concept of arytenoid rotation and displacement. These modifications produced an improved description of collision forces, allowed for membranous and cartilaginous incomplete glottal closure, and improved the model representation of the onset condition, which is the VF posturing prior to phonation. This enhanced model is referred to as triangular body-cover model (TBCM). In addition to the anatomical modifications, several numerical improvements were made. A state space vocal tract propagation method was implemented to reduce the computational time used to propagate sound waves, a truncated Taylor series (TTS) was implemented to solve the equations of motion of the system, and a smooth flow solution was implemented to avoid numerical errors.[96] The ability of the proposed model to represent VH was investigated in detail. By incorporating auditory feedback and compensatory mechanism, the proposed model was capable to reproduce detrimental patterns that are believed to be linked to voice problems.[60] When mimicking incomplete glottal closure, compensatory mechanism were utilized to achieve a given auditory target, which in turn resulted in increased contact pressures and related aerodynamic measures.

The framework for the Bayesian subject-specific model was addressed on Chapter 4. Two different strategies were analyzed: (1) Stationary estimation, and (2) non-stationary estimation. Synthetic cases of vocal folds were used to provide a ground truth for the estimation process, and each method was investigated separately. Both methods were able to estimate stationary parameters, and as expected, the stationary method preformed poorly on the non-stationary scenario. With a windowing technique the stationary method was able to follow the non-stationary parameters. However, given that this estimation process is based on invariant assumptions, the ability to observe time-varying signals was absent. The non-stationary method performed well on both scenarios, and allowed for obtaining an estimate of clinically-hidden measurements. Thus, we selected this approach, which is referred to as sequential Markov chain Monte Carlo (SMCMC)

The clinical investigation of the proposed technique was addressed in Chapter 5. In this Chapter, the ability to estimate model parameters from clinical data was explored using a clinically acquired, and spatially-calibrated, high speed videoendoscopy (HSV) in synchrony with acoustic and flow measurements. This type of recording is unique in its nature and allowed for the investigation of the proposed method in great detail. From HSV, the glottal edge displacement, the glottal area, and the initial estimation of arytenoid position and the onset condition were obtained. A Bayesian estimation process was performed using the edge displacement, glottal area, and glottal flow, with the aim of capturing the underlying TBCM configuration parameters. To our knowledge, this is the first approach to subject-specific modeling that accounts for multiple signals. Since in clinical data there is no ground-truth reference for the model parameters, a cross comparison was performed using the synchronous microphone output. The comparison of this microphone signal and the maximum a *posteriori* (MAP) estimated model output ("virtual sensor") were in close resemblance and within the credibility intervals of the estimation, which illustrated the accuracy of the method. In addition, the contact pressure was obtained through the virtual sensing capabilities of the approach. The comparison between the estimated contact pressure, the simulated scenario, and previous studies, illustrated that accurate estimations can be obtained with the proposed scheme. Taking all the above in consideration, the proposed scheme using the Bayesian SMCMC estimator with the TBCM implementation and a set of calibrated data, it is possible to construct a subject-specific model with known uncertainty. The Bayesian estimation process allowed for estimating model parameters and their associated credibility intervals, with the additional advantage of producing unobservable measurement estimates from the same model. Chapter 4 illustrated that it is possible to estimate the model parameters probability density function (pdf) that originally generated such dataset, even considering a large number of hidden states and non-linearities of the system. In addition, when compared to deterministic approaches, the Bayesian estimation provided a pdf of the inverse problem, thus allowing a more comprehensive analysis of the clinical data. A drawback of the SMCMC method is that it is more computationally expensive than other methods, which limits its potential. Nevertheless, given the non-linearity of the voice production system, and the high complexity of the estimation, the SMCMC provided a solution that is not bounded to linear assumptions, Gaussian distributions, or strictly convex solutions. In addition, provided the correct assumptions, the stochastic estimation could be simplified through linearization of the model, or by the restriction of its pdf (e.g., normal distributions). In all scenarios, the estimation will be as good as the match between

the data and the model.

This thesis work covered topics from the development of an enhanced model to the development and application of a stochastic estimation method for its parameters. The level of complexity of the model was fixed in the Bayesian estimation framework. A topic of subsequent research includes the search for an optimal model complexity, based upon the observed data. Ideally, this would be incorporated within the core of the estimation process itself, rather than through parametric variations. Additional components for future work include a framework for an experimental validation of the Bayesian model predictions, which could be performed in a controlled environment with silicone models of the VF or excised larynx experiments, in which all states of the numerical-model can be accounted for and directly measured. Contrast between the estimated values and their realizations should provide enough evidence to support the applicability of this method on clinical environments. In this thesis, the subject-specific model does not ensures uniqueness of the model, this is because the estimated parameters are rather control parameters, and not anatomic features. Therefore, current estimator configuration doe not attributes the variation of the observation to a variation of the anatomical features. Therefore, while a unique-subject specific model is not feasible under current conditions, it would be interesting to analyze which minimum requirements must voice production model have to generate a unique-subject model estimation. Such features could be useful in subject identification application, parameter-based classification, forensic analysis, or any other field where the unique-subject analysis is important.

During this thesis, no information was provided about the robustness of the model-estimator in terms of the minimum amount of data required for this method to work. These questions are still unanswered and are topics for future research in the area. Knowing the minimum number of signal that would be required to achieve a certain credibility, and possibly exploring its feasibility with imaging methods other than HSV (e.g., stroboscopy) would make the approach more relevant for the general clinical practice. Another interesting branch for future investigation is the possibility of merging both stationary and non-stationary Bayesian methods into a single estimator. The idea behind such a combination is to utilize the benefits from each method and use them in a single time-varying procedure. Using the *a priori* information of the system to produce pre-computed time-varying pdfs would require less particles to estimate the remaining states (e.g., pre-compute the flow signal pdf, given the glottal pressure and area distribution). Finally, we aim to explore the connection between the laboratory and ambulatory assessments of vocal function through the proposed Bayesian framework. For this purpose, an eventual two-step estimation procedure could be implemented by first obtaining model parameters in clinical (laboratory) setup and applying the model in a time-varying framework to estimate additional signals (e.g., contact forces) in the ambulatory assessment. Such procedure could allow for the enhancement of ambulatory voice monitoring through the proposed virtual sensing in this thesis, which would be expected to provide further insights to the pathophysiology of vocal hyperfunction and other disorders.

REFERENCES

[1] F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 3003–3012, 2000.

[2] F. Alipour, D. Montequin, and N. Tayama, "Aerodynamic profiles of a hemilarynx with a vocal tract," *Annals of Otology, Rhinology & Laryngology*, vol. 110, no. 6, pp. 550–555, 2001.

[3] F. Alipour, R. C. Scherer, and E. Finnegan, "Pressure-flow relationships during phonation as afunction of adduction," *Journal of Voice*, vol. 11, no. 2, pp. 187–194, 1997.

[4] M. Behlau, *Voz: o livro do especialista*. Revinter, 2005.

[5] A. Behrman, L. D. Dahl, A. L. Abramson, and H. K. Schutte, "Anterior-posterior and medial compression of the supraglottis: signs of nonorganic dysphoniaor normal postures," *Journal of Voice*, vol. 17, no. 3, pp. 403–410, 2003.

[6] D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," *The Journal of the Acoustical Society of America*, vol. 95, no. 6, pp. 3595–3604, 1994.

[7] D. A. Berry, D. W. Montequin, R. W. Chan, I. R. Titze, and H. T. Hoffman, "An investigation of cricoarytenoid joint mechanics using simulated muscle forces," *Journal of Voice*, vol. 17, no. 1, pp. 47–62, 2003.

[8] D. A. Berry, K. Verdolini, D. W. Montequin, M. M. Hess, R. W. Chan, and I. R. Titze, "A quantitative output-cost ratio in voice production," *Journal of Speech, Language, and Hearing Research*, vol. 44, no. 1, pp. 29–37, 2001.

[9] P. Birkholz, B. J. Kröger, and C. Neuschaefer-Rube, "Articulatory synthesis of words in six voice qualities using a modified two-mass model of the vocal folds," in *First International Workshop on Performative Speech and Singing Synthesis*, 2011.

[10] ——, "Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis," in *Proc. of the Interspeech 2011*, Florence, Italy, 2011, pp. 2681–2684.

[11] P. Birkholz and B. Kröger, "A survey of self-oscillating lumped-element models of the vocal folds," *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*, pp. 47–58, 2011.

[12] T. Braunschweig, J. Flaschka, P. Schelhorn-Neise, and M. Döllinger, "High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias," *Medical engineering & physics*, vol. 30, no. 1, pp. 59–66, 2008.

[13] J. V. Candy, *Bayesian signal processing: Classical, modern and particle filtering methods*. John Wiley & Sons, 2011, vol. 54.

[14] O. Cappé, E. Moulines, and T. Rydén, "Inference in hidden markov models," in *Proceedings of EUSFLAT Conference*, 2009, pp. 14–16.

[15] A. Castillo, C. Casanova, D. Valenzuela, and S. Castañón, "Prevalencia de disfonía en profesores de colegios de la comuna de santiago y factores de riesgo asociados," *Ciencia & trabajo*, vol. 17, no. 52, pp. 15–21, 2015.

[16] E. Cataldo, C. Soize, and R. Sampaio, "Uncertainty quantification of voice signal production mechanical model and experimental updating," *Mechanical Systems and Signal Processing*, vol. 40, no. 2, pp. 718–726, 2013.

[17] E. Cataldo, C. Soize, R. Sampaio, and C. Desceliers, "Probabilistic modeling of a nonlinear dynamical system used for producing voice," *Computational Mechanics*, vol. 43, no. 2, pp. 265–275, 2009.

[18] T. Chen, J. Morris, and E. Martin, "Particle filters for state and parameter estimation in batch processes," *Journal of Process Control*, vol. 15, no. 6, pp. 665–673, 2005.

[19] D. K. Chhetri, J. Neubauer, and D. A. Berry, "Neuromuscular control of fundamental frequency and glottal posture at phonation onset," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. 1401–1412, 2012.

[20] 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, pp. 1735–1745, 2007.

[21] B. Choi, *ARMA model identification*. Springer Science & Business Media, 2012.

[22] W. Conrad, "A new model of the vocal cords based on a collapsible tube analogy." *Medical research engineering*, vol. 13, no. 2, pp. 7–10, 1979.

[23] D. D. Cook and D. J. Robertson, "The generic modeling fallacy: Average biomechanical models often produce non-average results!" *Journal of Biomechanics*, vol. 49, no. 15, pp. 3609–3615, 2016.

[24] A. Cooke, C. L. Ludlow, N. Hallett, and W. S. Selbie, "Characteristics of vocal fold adduction related to voice onset," *Journal of Voice*, vol. 11, no. 1, pp. 12–22, 1997.

[25] G. Cornut and J. Lafon, "Vibrations neuro-musculaires des cordes vocales et theories de la phonation," *J Fr ORL*, vol. 9, pp. 317–324, 1960.

[26] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[27] B. Cranen and L. Boves, "On subglottal formant analysis," *The Journal of the Acoustical Society of America*, vol. 81, no. 3, pp. 734–746, 1987.

[28] B. Cranen and J. Schroeter, "Modeling a leaky glottis," *Journal of Phonetics*, vol. 23, no. 1-2, pp. 165–177, 1995.

[29] ——, "Physiologically motivated modelling of the voice source in articulatory analysis/synthesis," *Speech Communication*, vol. 19, no. 1, pp. 1–19, 1996.

[30] L. Cveticanin, "Review on mathematical and mechanical models of the vocal cord," *Journal of Applied Mathematics*, vol. 2012, 2012.

[31] S. H. Dailey, J. B. Kobler, R. E. Hillman, K. Tangrom, E. Thananart, M. Mauri, and S. M. Zeitels, "Endoscopic measurement of vocal fold movement during adduction and abduction," *The Laryngoscope*, vol. 115, no. 1, pp. 178–183, 2005.

[32] P. H. Dejonckere and M. Kob, "Pathogenesis of vocal fold nodules: new insights from a modelling approach," *Folia Phoniatrica et Logopaedica*, vol. 61, no. 3, pp. 171–179, 2009.

[33] D. D. Deliyski and P. Petrushev, "Methods for objective assessment of high-speed videoendoscopy," *Proc. Advances in Quantitative Laryngology (AQL)*, pp. 1–16, 2003.

[34] J. E. Dennis Jr and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*. Siam, 1996, vol. 16.

[35] J. DiStefano III, *Dynamic systems biology modeling and simulation*. Academic Press, 2015.

[36] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 773–781, 2002.

[37] ——, "Vibration parameter extraction from endoscopic image series of the vocal folds," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 8, pp. 773–781, 2002.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

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

[40] B. D. Erath and M. W. Plesniak, "The occurrence of the coanda effect in pulsatile flow through static models of the human vocal folds," *The Journal of the Acoustical Society of America*, vol. 120, pp. 1000–1011, 2006.

[41] B. D. Erath, M. Zañartu, and S. D. Peterson, "Modeling viscous dissipation during vocal fold contact: the influence of tissue viscosity and thickness with implications for hydration," *Biomechanics and modeling in mechanobiology*, vol. 16, no. 3, pp. 947–960, 2017.

[42] B. D. Erath, M. Zañartu, S. D. Peterson, and M. W. Plesniak, "The impact of a refined theoretical flow solver on nonlinear vocal fold dynamics in an asymmetric two-mass model of speech," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, pp. 033 113–1–8, 2011.

[43] ——, "Nonlinear vocal fold dynamics resulting from asymmetric fluid loading on a two-mass model of speech," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 21, no. 3, p. 033113, 2011.

[44] B. D. Erath, M. Zañartu, K. C. Stewart, M. W. Plesniak, D. E. Sommer, and S. D. Peterson, "A review of lumped-element models of voiced speech," *Speech Communication*, vol. 55, no. 5, pp. 667–690, 2013.

[45] L. J. Eriksson, "Higher order mode effects in circular ducts and expansion chambers," *The Journal of the Acoustical Society of America*, vol. 67, no. 2, pp. 545–550, 1980.

[46] 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*, pp. 1–11, 2017.

[47] G. Fant, *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*. The Hague: Mouton, 1960.

[48] J. L. Flanagan, *Speech analysis; synthesis and perception*, 2nd ed. New York: Springer-Verlag, 1972.

[49] J. L. Flanagan and L. L. Landgraf, "Self-oscillating source for vocal tract synthesizers," *IEEE Trans. Audio Electroacous.*, vol. AU-16, pp. 57–64, 1968.

[50] G. E. Galindo, S. D. Peterson, B. D. Erath, C. Castro, R. E. Hillman, and M. Zañartu, "Modeling the pathophysiology of phonotraumatic vocal hyperfunction with a triangular glottal model of the vocal folds," *Journal of Speech, Language, and Hearing Research*, 2017.

[51] G. E. Galindo, M. Zañartu, and J. I. Yuz, "A discrete-time model for the vocal folds," in *2014 IEEE EMBS International Student Conference (ISC)*, 2014, pp. 72–76.

[52] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE Proceedings F-Radar and Signal Processing*, vol. 140, no. 2. IET, 1993, pp. 107–113.

[53] E. U. Grillo, K. V. Abbott, and T. D. Lee, "Effects of masking noise on laryngeal resistance for breathy, normal, and pressed voice," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 4, pp. 850–861, 2010.

[54] F. H. Guenther, "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production." *Psychological review*, vol. 102, no. 3, p. 594, 1995.

[55] H. E. Gunter, "A mechanical model of vocal-fold collision with high spatial and temporal resolution," *The Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 994–1000, 2003.

[56] P. J. Hadwin, G. E. Galindo, K. J. Daun, M. Zañartu, B. D. Erath, E. Cataldo, and S. D. Peterson, "Non-stationary bayesian estimation of parameters from a body cover model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2683–2696, 2016.

[57] H. M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *The Journal of the Acoustical Society of America*, vol. 101, no. 1, pp. 466–481, 1997.

[58] H. M. Hanson, K. N. Stevens, H. K. J. Kuo, M. Y. Chen, and J. Slifka, "Towards models of phonation," *Journal of Phonetics*, vol. 29, no. 4, pp. 451–480, 2001.

[59] 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, pp. 543–550, May 2001.

[60] 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.

[61] ——, "Phonatory function associated with hyperfunctionally related vocal fold lesions," *Journal of Voice*, vol. 4, no. 1, pp. 52–63, 1990.

[62] M. Hirano, T. Kurita, and T. Nakashima, "The structure of the vocal folds," in *Vocal fold physiology*, K. Stevens and M. Hirano, Eds. University of Tokyo Press, 1981, pp. 33–41.

[63] M. Hirano, W. Vennard, and J. Ohala, "Regulation of register, pitch and intensity of voice. an electromyographic investigation of intrinsic laryngeal muscles," *Folia Phoniatrica et Logopaedica*, vol. 22, no. 1, pp. 1–20, 1970.

[64] 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.

[65] ——, "Structure and vibratory behavior of the vocal folds," *Dynamic aspects of speech production*, pp. 13–27, 1977.

[66] 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.

[67] 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, pp. 269–282, 2003.

[68] E. B. Holmberg, R. E. Hillman, J. S. Perkell, and P. C. Guiod, "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 6, pp. 1212–1223, 1995.

[69] K. Honda, H. Takemoto, T. Kitamura, S. Fujita, and S. Takano, "Exploring human speech production mechanisms by MRI," *IEICE Info. & Syst.*, vol. E87-D, pp. 1050–1058, 2004.

[70] J. Horáček, P. Šidlof, and J. G. Švec, "Numerical simulation of self-oscillations of human vocal folds with hertz model of impact forces," *Journal of Fluids and Structures*, vol. 20, pp. 853–869, 2005.

[71] D. G. Hottinger, C. Tao, and J. J. Jiang, "Comparing phonation threshold flow and pressure by abducting excised larynges," *The Laryngoscope*, vol. 117, no. 9, pp. 1695–1699, 2007.

[72] C. R. Houck, J. Joines, and M. G. Kay, "A genetic algorithm for function optimization: a matlab implementation," *NCSU-IE TR*, vol. 95, no. 09, 1995.

[73] M.-W. Hsiung and Y.-C. Hsiao, "The characteristic features of muscle tension dysphonia before and after surgery in benign lesions of the vocal fold," *ORL*, vol. 66, no. 5, pp. 246–254, 2004.

[74] E. J. Hunter, I. R. Titze, and F. Alipour, "A three-dimensional model of vocal fold abduction/adduction," *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1747–1759, 2004.

[75] R. Husson, *Physiologie de la phonation*. Masson, 1962.

[76] L. Ingber, A. Petraglia, M. R. Petraglia, M. A. S. Machado *et al.*, "Adaptive simulated annealing," in *Stochastic global optimization and its applications with fuzzy adaptive simulated annealing*. Springer, 2012, pp. 33–62.

[77] K. Ishizaka and M. Matsudaira, "Fluid mechanical considerations of vocal fold vibration," in *Speech Communication Research Laboratory, Mongraph No. 8*, Santa Barbara, CA, 1972.

[78] 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.

[79] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[80] ——, "Information theory and statistical mechanics. ii," *Physical review*, vol. 108, no. 2, p. 171, 1957.

[81] J. J. Jiang, C. E. Diaz, and D. G. Hanson, "Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: implications for the pathogenesis of vocal nodules," *Annals of Otology, Rhinology & Laryngology*, vol. 107, no. 7, pp. 603–610, 1998.

[82] J. J. Jiang and I. R. Titze, "Measurement of vocal fold intraglottal pressure and impact stress," *Journal of Voice*, vol. 8, no. 2, pp. 132–144, 1994.

[83] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*. Springer Science & Business Media, 2006, vol. 160.

[84] ——, "Statistical inverse problems: discretization, model reduction and inverse crimes," *Journal of computational and applied mathematics*, vol. 198, no. 2, pp. 493–504, 2007.

[85] J. L. Kelly and C. C. Lochbaum, "Speech synthesis," in *Proceedings of the Fourth International Congress on Acoustics*, Copenhagen, 1962, pp. 1–4.

[86] M. J. Kim, E. J. Hunter, and I. R. Titze, "Comparison of human, canine, and ovine laryngeal dimensions," *Annals of Otology, Rhinology & Laryngology*, vol. 113, no. 1, pp. 60–68, 2004.

[87] D. H. Klatt and L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers," *The Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 820–856, 1990.

[88] T. Koizumi, S. Taniguchi, and S. Hiromitsu, "Two-mass models of the vocal cords for natural sounding voice synthesis," *The Journal of the Acoustical Society of America*, vol. 82, no. 4, pp. 1179–1192, 1987.

[89] H.-K. J. Kuo, "Voice source modeling and analysis of speakers with vocal-fold nodules." Ph.D. dissertation, Harvard-MIT Division of Health Sciences and Technology, 1998.

[90] J. Liljencrants, "Speech synthesis with a reflection-type line analog," Ph.D. dissertation, Dept. of Speech Commun. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden, Stockholm, Sweden, 1985.

[91] S. E. Linville, "Glottal gap configurations in two age groups of women," *Journal of Speech, Language, and Hearing Research*, vol. 35, no. 6, pp. 1209–1215, 1992.

[92] J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[93] A. F. Llico, M. Zañartu, A. J. González, G. R. Wodicka, D. D. Mehta, J. H. Van Stan, and R. E. Hillman, "Real-time estimation of aerodynamic features for ambulatory voice biofeedback," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. EL14–EL19, 2015.

[94] J. C. Lucero and L. L. Koenig, "Simulations of temporal patterns of oral airflow in men and woman using a two-mass model of the vocal folds under dynamic control," *The Journal of the Acoustical Society of America*, vol. 117, no. 3, pp. 1362–1372, 2005.

[95] J. C. Lucero, "Oscillation hysteresis in a two-mass model of the vocal folds," *Journal of sound and vibration*, vol. 282, no. 3, pp. 1247–1254, 2005.

[96] J. C. Lucero and J. Schoentgen, "Smoothness of an equation for the glottal flow rate versus the glottal area," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2970–2973, 2015.

[97] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Commun.*, vol. 1, no. 3-4, pp. 199–229, 1982.

[98] R. S. McGowan, "An aeroacoustic approach to phonation," *The Journal of the Acoustical Society of America*, vol. 83, no. 2, pp. 696–704, 1988.

[99] R. S. McGowan, L. L. Koenig, and A. Löfqvist, "Vocal tract aerodynamics in /aca/ utterances: Simulations," *Speech Communication*, vol. 16, no. 1, pp. 67–88, 1995.

[100] D. D. Mehta, D. D. Deliyski, S. M. Zeitels, M. Zañartu, and R. E. Hillman, "Integration of transnasal fiberoptic high-speed videoendoscopy with time-synchronized recordings of vocal function," in *"Normal & Abnormal Vocal Folds Kinematics: High Speed Digital Phonoscopy (HSDP), Optical Coherence Tomography (OCT) & Narrow Band Imaging (NBI®)", Volume I: Technology*. Pacific Voice & Speech Foundation, San Francisco, CA, 2015, pp. 105–114.

[101] D. D. Mehta and R. E. Hillman, "Use of aerodynamic measures in clinical voice assessment," *Perspectives on Voice and Voice Disorders*, vol. 17, no. 3, pp. 14–18, 2007.

[102] D. D. Mehta, D. Rudoy, and P. J. Wolfe, "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking a," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1732–1746, 2012.

[103] 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, 2015.

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

[105] L. Mongeau, N. Franchek, C. H. Coker, and R. A. Kubli, "Characteristics of a pulsating jet through a small modulated orifice, with application to voice production," *The Journal of the Acoustical Society of America*, vol. 102, no. 2, pp. 1121–1133, 1997.

[106] M. D. Morrison, H. Nichol, and L. A. Rammage, "Diagnostic criteria in functional dysphonia," *The Laryngoscope*, vol. 96, no. 1, pp. 1–8, 1986.

[107] M. D. Morrison and L. A. Rammage, "Muscle misuse voice disorders: description and classification," *Acta oto-laryngologica*, vol. 113, no. 3, pp. 428–434, 1993.

[108] P. R. Murray and S. L. Thomson, "Vibratory responses of synthetic, self-oscillating vocal fold models," *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3428–3438, 2012.

[109] L. R. Neils and E. Yairi, "Effects of speaking in noise on vocal fatigue and vocal recovery," *Folia Phoniatrica et Logopaedica*, vol. 39, no. 2, pp. 104–112, 1987.

[110] D. O'Shaughnessy, "Linear predictive coding," *IEEE potentials*, vol. 7, no. 1, pp. 29–32, 1988.

[111] J. B. Park and L. Mongeau, "Experimental investigation of the influence of a posterior gap on glottal flow and sound," *The Journal of the Acoustical Society of America*, vol. 124, no. 2, pp. 1171–1179, 2008.

[112] H. M. Paynter, *Analysis and design of engineering systems*. MIT press, 1961.

[113] X. Pelorson, A. Hirschberg, A. P. J. Wijnands, and H. M. A. Bailliet, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation," *The Journal of the Acoustical Society of America*, vol. 96, pp. 3416–3431, 1994.

[114] C. Pemberton, A. Russell, J. Priestley, T. Havas, J. Hooper, and P. Clark, "Characteristics of normal larynges under flexible fiberscopic and stroboscopic examination: an australian perspective," *Journal of Voice*, vol. 7, no. 4, pp. 382–389, 1993.

[115] J. Perello, "The muco-undulatory theory of phonation," *Ann Otolaryngol*, vol. 79, pp. 722–725, 1962.

[116] 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.

[117] M. F. Regner, C. Tao, D. Ying, A. Olszewski, Y. Zhang, and J. J. Jiang, "The effect of vocal fold adduction on the acoustic quality of phonation: ex vivo investigations," *Journal of Voice*, vol. 26, no. 6, pp. 698–705, 2012.

[118] D. Robertson, M. Zañartu, and D. Cook, "Comprehensive, population-based sensitivity analysis of a two-mass vocal fold model," *PloS one*, vol. 11, no. 2, p. e0148309, 2016.

[119] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *The Journal of the Acoustical Society of America*, vol. 53, no. 6, pp. 1632–1645, 1973.

[120] ——, "An interactive model for the voice source," *Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm*, vol. 22, no. 4, pp. 001–017, 1981.

[121] ——, "Source-tract acoustic interaction in breathy voice," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, I. R. Titze and R. C. Scherer, Eds. The Denver Center for the Performing Arts, 1984, pp. 465–481.

[122] M. Rothenberg and S. Zahorian, "Nonlinear inverse filtering technique for estimating the glottal-area waveform," *The Journal of the Acoustical Society of America*, vol. 61, no. 4, pp. 1063–1070, 1977.

[123] M. Rothenberg, "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing," *The Journal of the Acoustical Society of America*, vol. 53, no. 6, pp. 1632–1645, 1973.

[124] N. Roy, R. M. Merrill, S. Thibeault, S. D. Gray, and E. M. Smith, "Prevalence of voice disorders in teachers and the general population," *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 2, pp. 281–293, 2004.

[125] 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.

[126] S. J. Rupitsch, J. Ilg, A. Sutor, R. Lerch, and M. Döllinger, "Simulation based estimation of dynamic mechanical properties for viscoelastic materials used for vocal fold models," *Journal of Sound and Vibration*, vol. 330, no. 18, pp. 4447–4459, 2011.

[127] S. Sahoo and A. Routray, "A novel method of glottal inverse filtering," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 7, pp. 1230–1241, 2016.

[128] A. Sama, P. N. Carding, S. Price, P. Kelly, and J. A. Wilson, "The clinical features of functional dysphonia," *The Laryngoscope*, vol. 111, no. 3, pp. 458–463, 2001.

[129] R. A. Samlan, B. H. Story, and K. Bunton, "Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 4, pp. 1209–1223, 2013.

[130] C. M. Sapienza and E. T. Stathopoulos, "Respiratory and laryngeal measures of children and women with bilateral vocal fold nodules," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 6, pp. 1229–1243, 1994.

[131] C. M. Sapienza, E. T. Stathopoulos, and W. Brown, "Speech breathing during reading in women withvocal nodules," *Journal of Voice*, vol. 11, no. 2, pp. 195–201, 1997.

[132] R. C. Scherer, F. Alipour, E. Finnegan, and C. G. Guo, "The membranous contact quotient: a new phonatory measure of glottal competence," *Journal of Voice*, vol. 11, no. 3, pp. 277–284, 1997.

[133] R. C. Scherer, B. Frazer, and G. Zhai, "Modeling flow through the posterior glottal gap," in *Proceedings of Meetings on Acoustics*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 060240.

[134] R. C. Scherer, I. R. Titze, and J. F. Curtis, "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," *The Journal of the Acoustical Society of America*, vol. 73, no. 2, pp. 668–676, 1983.

[135] J. Schoentgen and J. C. Lucero, "Synthesis by rule of disordered voices," in *International Conference on Nonlinear Speech Processing*. Springer, 2013, pp. 120–127.

[136] R. Schwarz, M. Döllinger, T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 2717–2732, 2008.

[137] R. Schwarz, U. Hoppe, M. Schuster, T. Wurzbacher, U. Eysholdt, and J. Lohscheller, "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model," *IEEE transactions on biomedical engineering*, vol. 53, no. 6, pp. 1099–1108, 2006.

[138] A. F. Smith and A. E. Gelfand, "Bayesian statistics without tears: a sampling–resampling perspective," *The American Statistician*, vol. 46, no. 2, pp. 84–88, 1992.

[139] E. Smith, K. Verdolini, S. Gray, S. Nichols, J. Lemke, J. M. Barkmeier-Kraemer, H. Dove, and H. Hoffman, "Effect of voice disorders on quality of life," *Journal of Medical Speech-Language Pathology*, vol. 4, no. 4, pp. 223–244, 1996.

[140] S. L. Smith and E. J. Hunter, "A viscoelastic laryngeal muscle model with active components," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 2041–2051, 2014.

[141] M. Sodersten and P.-A. Lindestad, "Glottal closure and perceived breathiness during phonation in normally speaking subjects," *Journal of Speech, Language, and Hearing Research*, vol. 33, no. 3, pp. 601–611, 1990.

[142] E. Sperry, R. E. Hillman, and J. S. Perkell, "The use of inductance plethysmography to assess respiratory function in a patient with vocal nodules," *Journal of Medical Speech-Language Pathology*, vol. 2, pp. 137–145, 1994.

[143] S. V. Stager, S. A. Bielamowicz, J. R. Regnell, A. Gupta, and J. M. Barkmeier, "Supraglottic activityevidence of vocal hyperfunction or laryngeal articulation," *Journal of speech, language, and hearing research*, vol. 43, no. 1, pp. 229–238, 2000.

[144] S. V. Stager, R. Neubert, S. Miller, J. R. Regnell, and S. A. Bielamowicz, "Incidence of supraglottic activity in males and females: a preliminary report," *Journal of Voice*, vol. 17, no. 3, pp. 395–402, 2003.

[145] E. T. Stathopoulos, J. E. Huber, K. Richardson, J. Kamphaus, D. DeCicco, M. Darling, K. Fulcher, and J. E. Sussman, "Increased vocal intensity due to the lombard effect in speakers with parkinson's disease: Simultaneous laryngeal and respiratory strategies," *Journal of communication disorders*, vol. 48, pp. 1–17, 2014.

[146] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *The Journal of the Acoustical Society of America*, vol. 97, no. 3, pp. 1874–1884, 1995.

[147] C. E. Stepp, R. E. Hillman, and J. T. Heaton, "The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 5, pp. 1220–1226, 2010.

[148] K. N. Stevens, *Acoustic phonetics*, 1st ed. Cambridge, Mass.: MIT Press, 1998.

[149] K. N. Stevens and A. S. House, "Development of a quantitative description of vowel articulation," *The Journal of the Acoustical Society of America*, vol. 27, no. 3, pp. 484–493, 1955.

[150] W. J. Stewart, *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press, 2009.

[151] B. H. Story, "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. dissertation, University of Iowa, Iowa City, IA, 1995.

[152] ——, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 327–335, 2008.

[153] B. H. Story and K. Bunton, "Production of child-like vowels with nonlinear interaction of glottal flow and vocal tract resonances," in *Proceedings of Meetings on Acoustics*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 060303.

[154] 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.

[155] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 537–554, 1996.

[156] J. Sylvestre and P. MacLeod, "Le muscle vocal humain est-il asynchrone," *Journal Physiol*, vol. 5, pp. 373–389, 1968.

[157] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto, "Measurement of temporal changes in vocal tract area function from 3d cine-mri data," *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1037–1049, 2006.

[158] C. Tao, Y. Zhang, D. G. Hottinger, and J. Jiang, "Asymmetric airflow and vibration induced by the coanda effect in a symmetric model of the vocal folds," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2270–2278, 2007.

[159] C. Tao and J. J. Jiang, "Mechanical stress during phonation in a self-oscillating finite-element vocal fold model," *Journal of biomechanics*, vol. 40, no. 10, pp. 2191–2198, 2007.

[160] C. Tao, J. J. Jiang, and Y. Zhang, "Simulation of vocal fold impact pressures with a self-oscillating finite-element model," *The Journal of the Acoustical Society of America*, vol. 119, no. 6, pp. 3987–3994, 2006.

[161] C. Tao, Y. Zhang, and J. J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 794–801, 2007.

[162] S. L. Thomson, L. Mongeau, and S. H. Frankel, "Aerodynamic transfer of energy to the vocal folds," *The Journal of the Acoustical Society of America*, vol. 118, no. 3, pp. 1689–1700, 2005.

[163] I. R. Titze, "Parameterization of the glottal area, glottal flow, and vocal fold contact area," *The Journal of the Acoustical Society of America*, vol. 75, no. 2, pp. 570–580, 1984.

[164] ——, *Principles of voice production*. Iowa City, IA: National Center for Voice and Speech, 2000, 293-301.

[165] ——, "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice," *Journal of Voice*, vol. 18, no. 3, pp. 292–298, 2004.

[166] ——, "Nonlinear source-filter coupling in phonation: Theory," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 2733–2749, 2008.

[167] I. R. Titze and F. Alipour, *The Myoelastic Aerodynamic Theory of Phonation*. Denver, CO ; Iowa City, IA: The National Center for Voice and Speech, 2006.

[168] I. R. Titze, E. Hunter, and J. G. Švec, "Voicing and silence periods in daily and weekly vocalizations of teachers," *The Journal of the Acoustical Society of America*, vol. 121, no. 1, pp. 469–478, 2007.

[169] I. R. Titze and E. J. Hunter, "A two-dimensional biomechanical model of vocal fold posturing," *The Journal of the Acoustical Society of America*, vol. 121, no. 4, pp. 2254–2260, 2007.

[170] I. R. Titze, J. Jiang, and D. G. Drucker, "Preliminaries to the body-cover theory of pitch control," *Journal of Voice*, vol. 1, no. 4, pp. 314–319, 1988.

[171] I. R. 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, pp. 1902–1915, 2008.

[172] I. R. Titze and B. H. Story, "Acoustic interactions of the voice source with the lower vocal tract," *The Journal of the Acoustical Society of America*, vol. 101, no. 4, pp. 2234–2243, 1997.

[173] ——, "Rules for controlling low-dimensional vocal fold models with muscle activation," *The Journal of the Acoustical Society of America*, vol. 112, no. 3, pp. 1064–1076, 2002.

[174] I. R. Titze, J. G. Švec, and P. S. Popolo, "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 4, pp. 919–932, 2006.

[175] I. R. Titze and A. S. Worley, "Modeling source-filter interaction in belting and high-pitched operatic male singing," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1530–1540, 2009.

[176] I. T. Tokuda and H. Herzel, "Detecting synchronizations in an asymmetric vocal fold model from time series data," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 15, pp. 013 702:1–11, 2005.

[177] I. T. Tokuda, M. Zemke, M. Kob, and H. Herzel, "Biomechanical modeling of register transitions and the role of vocal tract resonators," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, pp. 1528–1536, 2010.

[178] A. van der Schaft, "Port-hamiltonian systems," in *L2-Gain and Passivity Techniques in Nonlinear Control*. Springer, 2017, pp. 113–171.

[179] H. Von Leden, "A cultural history of the larynx and voice," *Professional voice: the science and art of clinical care. 3rd ed. San Diego (CA): Plural Publishing, Inc*, pp. 9–88, 2005.

[180] J. G. Švec, F. Šram, and H. K. Schutte, "Videokymography in voice disorders: What to look for?" *Annals of Otology, Rhinology & Laryngology*, vol. 116, no. 3, pp. 172–180, 2007.

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

[182] D. Wong, M. R. Ito, N. B. Cox, and I. R. Titze, "Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 383–394, 1991.

[183] T. Wurzbacher, M. Döllinger, R. Schwarz, U. Hoppe, U. Eysholdt, and J. Lohscheller, "Spatiotemporal classification of vocal fold dynamics by a multimass model comprising time-dependent parameters," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. 2324–2334, 2010.

[184] T. Wurzbacher, R. Schwarz, M. Döllinger, U. Hoppe, U. Eysholdt, and J. Lohscheller, "Model-based classification of nonstationary vocal fold vibrations," *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 1012–1027, 2006.

[185] A. Yang, D. A. Berry, M. Kaltenbacher, and M. Döllinger, "Three-dimensional biomechanical properties of human vocal folds: Parameter optimization of a numerical model to match in vitro dynamics," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. 1378–1390, 2012.

[186] A. Yang, J. Lohscheller, D. A. Berry, S. Becker, U. Eysholdt, D. Voigt, and M. Döllinger, "Biomechanical modeling of the three-dimensional aspects of human vocal fold dynamics," *The Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 1014–1031, 2010.

[187] A. Yang, M. Stingl, D. A. Berry, J. Lohscheller, D. Voigt, U. Eysholdt, and M. Döllinger, "Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model," *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. 948–964, 2011.

[188] Y. Yang and D. B. Dunson, "Sequential markov chain monte carlo," *arXiv preprint arXiv:1308.3861*, 2013.

[189] J. I. Yuz and G. C. Goodwin, *Sampled-data models for linear and nonlinear systems*. Springer, 2014.

[190] J. I. Yuz-Eissmann, *Sampled-data Models for Linear and Non-linear Systems*. University of Newcastle, 2006.

[191] M. Zañartu, "Influence of acoustic loading on the flow-induced oscillations of single mass models of the human larynx," Master's thesis, School of Electrical and Computer Engineering, Purdue University, 2006.

[192] M. Zañartu, G. E. Galindo, B. D. Erath, S. D. Peterson, G. R. Wodicka, and R. E. Hillman, "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction," *The Journal of the Acoustical Society of America*, vol. 136, no. 6, pp. 3262–3271, 2014.

[193] M. Zañartu, J. C. Ho, D. D. Mehta, R. E. Hillman, and G. R. Wodicka, "Acoustic coupling during incomplete glottal closure and its effect on the inverse filtering of oral airflow," *The Journal of the Acoustical Society of America*, vol. POMA 19, pp. 1–7, 2013.

[194] M. Zañartu, D. D. Mehta, J. C. Ho, G. R. Wodicka, and R. E. Hillman, "Observation and analysis of *in vivo* vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study," *The Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 326–339, 2011.

[195] 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, pp. 1119–1129, 2007.

[196] Y. Zhang and J. J. Jiang, "Asymmetric spatiotemporal chaos induced by polypoid mass in the excised larynx," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 18, pp. 043 102:1–6, 2008.

[197] Y. Zhang, C. Tao, and J. J. Jiang, "Parameter estimation of an asymmetric vocal-fold system from glottal area time series using chaos synchronization," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 16, no. 2, pp. 023 118:1–8, 2006.

[198] Z. Zhang, L. Mongeau, and S. Frankel, "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, p. 1652ï¿½1663, 2002.

[199] 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, pp. 1558–1569, 2006.

[200] Z. Zhang, "Regulation of glottal closure and airflow in a three-dimensional phonation model: Implications for vocal intensity control," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 898–910, 2015.

[201] W. Zhao, C. Zhang, S. H. Frankel, , and L. Mongeau, "Computational aeroacoustics of phonation, part i: Computational methods and sound generation mechanisms," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2134–2146, 2002.