ACOUSTIC COUPLING IN PHONATION AND ITS EFFECT ON INVERSE FILTERING OF ORAL AIRFLOW AND NECK SURFACE ACCELERATION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Matías Zañartu Salas

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2010

Purdue University

West Lafayette, Indiana

ACKNOWLEDGMENTS

I am truly grateful to Professor George R. Wodicka for providing me the opportunity and freedom to pursue this thesis research. His continuous guidance, support, encouragement, and enthusiasm made my doctoral studies at Purdue a wonderful experience. I am also indebted to Professor Robert E. Hillman for inviting me to the Center for Laryngeal Surgery & Voice Rehabilitation and for providing me with valuable advice regarding many clinical components of my research.

I extend my thanks to the members of my thesis committee Professors Luc Mongeau, Jessica Huber, and Thomas Talavage for pushing me to go the extra mile and yet helping me to stay focused along the way. Dr. Hans Pasterkamp and Dr. Steve Kraman provided many important insights and contributions to the biosensing component of this thesis in the early stages on this research. Professors Brad Story and Ingo Titze contributed to the implementation of a parallel glottal gap in the self-sustained simulations.

Many other friends and collaborators contributed to this thesis research, including Daryush Mehta, Julio Ho, Steven Lulich, Byron Erath, Douglas Cook, Sean Peterson, Harold Cheyne, Maia Braden, Jim Kobler, Dimitar Delyski, Thomas Quatieri, and Janet Slifka, for which I am much obliged. I am also thankful for the fruitful discussions I had with the members of the voice quality study group at MIT, which influenced many aspects of my work.

Finally, my deepest gratitude goes to my wife, Sandra, and my daughter, Julieta, for their support, patience, and unconditional love that made this complete journey possible.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

SYMBOLS

A_g	Glottal area
c_d	Orifice discharge coefficient
F	Applied force for vocal folds model
k_t	Empirically-determined glottal coefficient
ΔP_g	Transglottal Pressure
p_e	unsteady supraglottal (acoustic) pressure
P_s	Steady subglottal (lung) pressure
p_s	unsteady subglottal (acoustic) pressure
T	Transfer function
T_u	Uncoupled transfer function
T_c	Coupled transfer function
U	Airflow
dU	Airflow derivative
U_g	Glottal airflow
U_o	Uncoupled glottal airflow
U_{supra}	Airflow entering the vocal tract
U_{sub}	Airflow entering the subglottal system
v	Particle velocity
x	Lateral axis of displacement of vocal fold motion
z	Superior-inferior axis vocal folds
Z	Acoustic impedance
Z_g	Glottal acoustic impedance
Z_{supra}	Supraglottal tract acoustic impedance
Z_{sub}	Subglottal tract acoustic Impedance

ABBREVIATIONS

AC	Alternating current, unsteady component of a signal
ACC	Neck surface acceleration signal
AP	Anterior-posterior axis of the vocal folds
BLEAP	Boundary layer estimation of asymmetric pressures
BMI	Body Mass Index
CPIF	Closed phase inverse filtering
dB	Decibel
DC	Direct current, steady component of a signal
EGG	Electroglottograph signal
HRF	Harmonic richness factor
HSV	Laryngeal high-speed videoendoscopy
IBIF	Impedance-based inverse filtering
IF72	Bibliographic reference: Ishizaka and Flanagan (1972)
IOP	Intraoral pressure signal
LPC	Linear predictive coding
LR	Left and right vocal folds
MADR	Maximum area declination rate
MFDR	Maximum flow declination rate
MIC	Microphone signal
MRI	Magnetic resonance imaging
RMSE	Root mean squared error
SPL	Sound pressure level
ST95	Bibliographic reference: Story and Titze (1995)
OVV	Oral airflow signal

ABSTRACT

Zañartu Salas, M. Ph.D., Purdue University, August 2010. Acoustic Coupling in Phonation and its Effect on Inverse Filtering of Oral Airflow and Neck Surface Acceleration. Major Professor: George R. Wodicka.

Many aspects of the coupling between vocal fold motion, glottal airflow, and resulting sound waves remain unexplored, particularly during incomplete glottal closure. In addition, this coupling has been associated with distortions in the estimates of glottal parameters obtained from neck acceleration measurements, thus affecting the ambulatory assessment of vocal function. This acoustic coupling and its effects on inverse filtering of oral airflow and neck surface acceleration during vowel production were investigated using numerical and experimental methods. A lumped element self-oscillating model was extended to include different types of incomplete glottal closure, which provided an enhanced simulation tool that mimicked normal and hyperfunctional voices. Novel *in vivo* measurements and analyses of vocal fold tissue instabilities due to acoustic coupling were obtained using synchronous laryngeal high-speed videoendoscopy recordings. These observations suggested that instabilities differed at the vocal fold tissue level from those where no strong coupling was present, but classified each as subcritical Hopf bifurcations. These findings supported the concept of describing interactions within the voice production system based on acoustic impedance representations. In addition, an impedance-based inverse filtering approach with time-varying glottal coupling demonstrated how source-tract interactions skewed and altered the glottal pulses. Based on these observations, a subglottal inverse filtering scheme for neck surface acceleration was proposed and evaluated, yielding estimates of glottal aerodynamic parameters with accuracy comparable to that of standard pneumotachograph measurements and closed-phase inverse filtering.

1. INTRODUCTION

This chapter introduces the motivation for the proposed research, indicating the problem and its significance. The objectives and hypotheses of the research and its expected contributions are stated. The structure of the thesis document is then described.

1.1 Motivation

Linear source-filter theory, as applied to voice production [1], assumes that the glottal impedance is much larger than the acoustic input impedance of the vocal tract and subglottal system, with no interaction or coupling between components taken into account. Different studies have elucidated the importance of the interaction, noting that it affects glottal airflow introducing skewing and ripples in the open phase of its pulses [2–5], which has been referred to as "true glottal source" [6–8]. Source-filter interaction has also been related to the effects of incomplete glottal closure, which provides the basis of the current understanding of breathy and female voices. Indeed, it was shown that incomplete glottal closure is related to: changes in the harmonic decay of the source spectra and changes in the vocal tract formant and bandwidths [9–12], the introduction of turbulent noise [10,13], and fluctuations during the closed phase of the cycle [14, 15].

More recently, researchers have revised principles governing source-filter interaction [16, 17] based on enhanced vocal tract geometries [18–22] that suggest the tract has an impedance comparable or even larger than that of the glottis [23]. A recent theory [17] proposed that irregularities, instabilities, and nonlinear phenomena in voice production can be highly affected or triggered by the acoustic coupling between

components. It was also stated that this interaction could be completely described in terms of the system impedances.

Many questions regarding source-filter interactions remain unanswered, as supporting evidence with human subjects has been partial and many theoretical aspects have not been covered. The foundations supporting the interaction are constituted by numerical simulations or recordings of radiated pressure. Some of the aspects that remain largely untested or not fully validated with human subject data include the presence of skewing and ripples in the glottal pulses and *in vivo* visualizations of acoustically-induced instabilities in vocal fold tissue motion. Furthermore, nonlinear source-filter coupling theory [17] is considered incomplete, as it only refers to the tract impedance without accounting for the effects of different glottal configurations. In particular, it is currently unclear how the source-filter coupling affects the tissue dynamics, energy transfer, and glottal aerodynamics during the presence of incomplete glottal closure, where coupling between the subglottal and supraglottal tracts is also present.

An important application where nonlinear, interactive principles in voice production affect current practices is in the inverse filtering of speech sounds. This technique is used to estimate the source of excitation at the glottis and it is based on source-filter theory principles to separate and remove the acoustic effects of the tracts. This technique is primarily performed for the vocal tract using recordings of oral airflow or radiated pressure. A wide range of applications make use of inverse filtering, including the assessment of vocal function [24–26], vocal training [27], speech enhancement [28], speech synthesis [29], speech coding [30], and speaker identification [31]. Most recent methods in inverse filtering make use of the closed phase of the vocal fold oscillation, where the glottis remains closed and the flow-sound interaction is assumed to be largely suppressed [31, 32]. This approach has been suggested to overcome problems found in traditional and alternative methods [33, 34], and it is expected to preserve more closely the "true glottal source". However, the effects of incomplete glottal closure on this technique remain unexplored. It is possible that closed-phase based

methods incorrectly suppress ripples during the closed phase and distort the glottal airflow estimates due to subglottal coupling. This type of tract coupling has not been explored in the context of inverse filtering before.

A different type of inverse filtering technique can be implemented through acceleration measured on the skin overlying the suprasternal notch to obtain estimates of glottal parameters [35–38]. This technique has been shown of great importance for the ambulatory assessment of vocal function. This subglottal inverse filtering requires a different approach, as zeros are present in its impedance, making standard vocal tract based methods unapplicable. In addition, the acoustic coupling between the subglottal system and the vocal tract is expected to be more pronounced in this case due to the much larger impedance of the latter. Initial efforts have indicated that the estimation of glottal parameters by means of inverse filtering via the subglottal system is possible [35]. However, these attempts were limited by the partial understanding of the underlying physical phenomena and the factors that could distort the estimates. Among them, the effects of incomplete glottal closure were suggested to increase the acoustic coupling with the vocal tract and thus distort the estimates of glottal airflow. The appropriate way to handle the coupling effects and obtain estimates of the "true glottal source" in this case thus remain unexplored.

Incomplete glottal closure has been shown to be common in normal and disordered voices [25, 26, 39], and female voices can have no glottal contact at all [12, 13, 40]. However, there have been only a few studies on the effects of incomplete closure on the vocal folds dynamics. Only recently the influence of a posterior gap on the airflow and its net energy transferred was investigated using a driven synthetic model [41]. However, no theoretical or numerical studies have explored the effects the incomplete closure on the tissue dynamics and energy transfer. Further understanding of the underlying physical phenomena during incomplete glottal closure can impact inverse filtering studies.

The governing principles of acoustic interaction and its implications on the inverse filtering of speech sounds obtained from both oral airflow and neck surface acceler-

ation signals are explored in this thesis research. Particular emphasis is placed on improving methods for sensing and analyzing the acceleration signal from the neck surface because of growing interest in the use of this signal as an unobtrusive approach for ambulatory monitoring of vocal function. This approach has the potential to improve the diagnosis and treatment of commonly-occurring voice disorders that are believed to result from faulty and/or abusive patterns of vocal behavior. Such patterns are difficult to assess in the clinic and could potentially be much better characterized by long-term ambulatory monitoring of vocal function as individuals engage in their typical daily activities. It is especially critical to understand the impact of source-tract interactions on the acceleration signal since common voice disorders are often associated with incomplete glottal closure.

1.2 Research objectives

The general scope of this thesis research is the better understanding of source-filter interaction and its implications on the processing of speech signals that aim to retrieve glottal source properties. An important goal is to develop signal processing tools that incorporate the knowledge of the interactions for the enhancement of the vocal function assessment using neck surface acceleration. In order to achieve this goal, the following specific aims were proposed:

1. Investigate, through numerical simulations, the effects of acoustic coupling between the subglottal and supraglottal tracts during sustained vowels with complete and incomplete glottal closure. It was hypothesized that the use of lumped element representations for the different system components could provide new insights on the complex fluid-structure and flow-sound interactions observed during voice production.

2. Study the effects of acoustic coupling on tissue dynamics *in vivo* with human subjects. It was hypothesized that the sub- and supra-glottal tracts act as lumped impedance elements that can affect tissue dynamics. Verification of

this hypothesis would validate current theories of source-filter interaction based on lumped representations of the system.

3. Use the knowledge obtained from studying the acoustic coupling between tracts to develop an inverse filtering scheme that allows for an enhanced estimation of glottal airflow from the acceleration of the skin overlying the sternal notch. It was hypothesized that a scheme that accounts for the coupling effects would provide more reliable estimation of clinical parameters, thus enhancing the ambulatory assessment of vocal function.

4. Explore biosensing issues associated with neck surface acceleration measurements. It was hypothesized that bioacoustic sensors used in the skin surface may capture corrupting components from multiple pathways, for which sensor optimization may be needed.

5. Evaluate the aforementioned inverse filtering scheme for controlled scenarios that represent different quantifiable glottal configurations in normal phonation during sustained vowels. It was hypothesized that simple and repeatable vocal maneuvers reduce the uncertainty during the evaluation, thus providing more insights on the performance of the estimation scheme. In addition, it was expected that a combination of theoretical and experimental evaluation methods would provide a more complete and insightful assessment.

1.3 Expected contributions

The thesis research strives to make several scientific contributions. First, it adds to research efforts aimed at better understanding of acoustic coupling in phonation as well as its relation with other measurable signals used in the assessment of vocal function. The efforts also contribute to the development of more realistic modeling of speech production, inverse filtering algorithms of human speech, and the use of neck acceleration in the ambulatory assessment of vocal function.

1.4 Contents

The aforementioned specific aims can be divided in four research components: 1) numerical simulations of acoustic coupling, 2) experiments of tissue instabilities due to acoustic coupling, 3) biosensing evaluation of neck surface acceleration, and 4) experiments on inverse filtering with and without acoustic coupling. Thus, the remainder of this thesis document is organized as follows. In Chapter 2, the key points pertaining to the literature for these components are summarized. Chapters 3-6 individually address each of these research components, specific aims, and approaches taken. These chapters are relatively self-contained and include additional rationale, methods, results, discussion, and conclusion sections within. Chapter 3 explores the problem of source-filter interactions for complete and incomplete closure via numerical simulations of driven glottal area and an extended lumped mass representation of the body-cover model [42], thus addressing specific aim 1. Chapter 4 deals with specific aim 2, wherein the validity of current theories of source-filter interaction based on their lumped representations is explored *in vivo*. Thus, comparison of tissue instabilities with and without the presence of strong acoustic coupling is investigated with the aim of observing if the coupling addition changes the dynamics of the vocal fold oscillation. Chapter 5, covers biosensing issues related to neck surface acceleration recordings, thus addressing specific aim 4. The experiments of this chapter were used to support further inverse filtering investigations using this measurement technique. Chapter 6 introduces a number of investigations related to the inverse-filtering of speech sounds and how they are affected by source-filter interactions, thus addressing specific aims 3 and 5. This chapter introduces a model-based inverse filtering with glottal coupling representing complete and incomplete glottal closure. Synthetic speech and comprehensive high-speed video recordings performed in two normal subjects are used for different vowels and glottal configurations. Finally, chapter 7 captures the overall contributions and elucidates major findings, discussion points, and future work.

2. BACKGROUND

The increasing interest in the complexities observed during speech production and the higher demand for better tools for the assessment of vocal function have been evident in recent years. In order to provide an overview of the well-established and more recent notions pertaining to this thesis research, a brief literature review is presented in this chapter. The foundation for the research and important concepts are also described.

2.1 Theories and models of voice production

Early theories of speech production defined the fundamental principles governing vocal folds mechanics, aerodynamics, vocal tract acoustics, and the relation between them. The myoelastic theory of phonation [43] defined the fundamental principles associated to the self-sustained oscillations of the vocal folds. Linear source-filter theory [1] provided an important feature that unified the role of the additional components. In this representation, the subglottal system was considered as simple pressure source that provided the energy necessary for the self-sustained oscillation of the vocal folds, and the vocal tract acted as a simple filter that modulated the source sound. No interaction between the system components was considered and the effects of the subglottal tract were neglected. Even though it contains these simplifying assumptions, this representation has dominated the field of speech communication for more than 40 year and found many useful applications.

Subsequent theoretical studies acknowledged changes in the glottal airflow due to its interaction with the subglottal system and vocal tract [5,8,44–46], suggesting that this interaction can affect the source and filter properties significantly. Evidence of

these early observations were reported in human subjects, as seen, for example, in female and breathy voice [10, 12], and inverse filtering of oral volume velocity [24, 47].

Recent studies have suggested that source-filter interactions can be much stronger and completely alter the dynamics of vocal fold tissue motion. This claim is primarily based on advances in vocal tract imaging [18, 20–22] that have elucidated that the vocal tract impedance can be comparable to that of the glottis [16,17,23]. The nonlinear source-filter coupling theory [17] unifies all previous observations on source-filter interactions and introduces the concept of interactions due to impedance relations. Supporting evidence for the latter with human subjects include register transitions in singers [48] and dynamic vocal exercises [49, 50].

Understanding of the impedance represented by the subglottal system, glottis, and vocal tract is fundamental to explore the interaction between these complex systems. In the subsequent subsections, the fundamental ideas and assumptions behind the models of all three components are described in detail.

2.1.1 Subglottal and supraglottal impedances

Initial models of the vocal tract considered it as a simple system with three tube sections and a radiation impedance [1]. This representation allowed differentiation between vowels, where the formant frequencies were computed by combining models of uniform tubes and Helmholtz resonators. A more accurate representation made use of a larger numbers of tube sections and a transmission line method consisting of cascaded T-sections that accounted for losses in each tube section [51]. Further adjustments were made to the transmission line approach to account for yielding walls, nasal coupling, and additional losses [3, 52]. This technique has been the standard to define complex geometries and represent the input impedance or the vocal tract given its detailed frequency domain nature. Simplified versions of transmission line method used a reduced number of tube sections and loop equations to represent the

flow in the vocal tract in time domain [45, 46]. Only recently a full representation of this method has been adapted to handle complex geometries in time domain [53, 54].

An alternative representation of the tract acoustics in time domain known as "wave reflection analog" was established decades ago [55]. This technique followed a multi-tube approach to compute reflection coefficients between junctions and performed synchronous computations of the acoustic wave propagation. Several improvements have been introduced since its original design to account for multiple losses and yielding walls [56–58]. This technique has been more widely adopted in self-sustained oscillating models of phonation [42, 59, 60].

These methods assume planar waves within the tract, and base the overall tract response on the equivalent area function. These assumptions limit the frequency range to the length of each tube section, discard other possible modes of vibration within the tract, and neglect the effect of tube bending that are known to introduce turbulence and additional effects when bulk flow is present. Even with these limitations, the transmission line and wave reflection schemes remain as the fundamental representations to express the vocal tract acoustics and its associated acoustic impedance.

Accurate representations of the vocal tract are fundamental for these multi-tube models to represent the inherent acoustic characteristics well. Early area functions were based on X-ray imaging [1, 51]. More recent advances in vocal tract descriptions using magnetic resonance imaging [18, 20–22] have provided more reliable estimates that exhibit narrower area functions with respect to those from previous studies. These newer imaging representations allowed researchers to revaluate the impedance relations between the glottis and the vocal tract. A detailed analysis of the input impedance of the vocal tract observed from the glottis has been performed [23, 61]. The length of the epilarynx tube and its cross-sectional area were suggested to be the main factors in controlling the impedance of the vocal tract, the latter being more sensitive. Narrow entry sections in the vocal tract result in a high impedance, comparable with the impedance of the glottis, which favors flow-sound interactions.

Initial interest in subglottal acoustics has its origins in respiratory physiology. Even though the subglottal system was neglected for years in speech production research, recent studies have illustrated that it can significantly affect the system dynamics and source characteristics [17,60,62,63] and thus subglottal acoustics is now gaining more attention in this field. The impedance of the subglottal system observed from near the glottis was first investigated in laryngectomized patients and then modeled using a transmission line approach [64]. Subsequent investigations refined these initial efforts and explored the ability of the transmission line model to predict signals captured by accelerometers placed on the neck and chest wall surfaces due to the presence of an external sound source at the mouth [65–67]. Further refinements were made to incorporate turbulent sources in the model to mimic respiratory sounds [68, 69]. All these studies have shown a consistent behavior when describing the subglottal impedance, exhibiting prominent poles (but with lower amplitude with respect to those from the vocal tract) and marked zeros.

As in the vocal tract impedance, the geometry of the system plays a significant role in its overall characteristics. The aforementioned models have been constructed with different symmetric and asymmetric anatomical representations from male adult subjects [70–73]. It has been recently noted from magnetic resonance imaging data that the trachea can have significant variations in its length and diameter between subjects and speech gestures (Story, personal communication). These variations can have a significant impact when estimating the subglottal impedance and limit the application of models based on male anatomical representations.

To date, no efforts have been made to incorporate a full branching scheme for the subglottal tract using the wave reflection analog. Most studies based on this technique use simplified geometries representing the total subglottal area of the trachea and bronchi with a resistive termination impedance [32, 60, 74, 75]. The extent to where this assumption is appropriate is a topic of current investigations.

2.1.2 The glottal impedance

The glottal impedance differs from traditional acoustical or mechanical impedance expressions in that it is defined as the ratio between the transglottal pressure and glottal airflow. It is a time-varying quantity with both steady (DC) and unsteady (AC) components. Many studies have helped to quantify each of these components using two main approaches: experiments with either steady or pulsatile flow. The former has been more successful in describing glottal impedance by directly estimating its steady behavior and making use of the quasi-steady assumption to describe its unsteady component. The quasi-steady assumption states that unsteady representations can be described as a sequence of steady scenarios. Attempts to estimate the unsteady component directly using pulsating flow, normally obtained by an external acoustic excitation and a fixed glottal orifice, are not able to reproduce important effects that are a product of the periodic interruption of the flow at the glottis. Thus, pressure buildups and other important effects in the system are not taken into account, which typically makes these investigations less physiologically relevant.

The first experiment to describe the transglottal pressure drop using steady flow [76] was known to misrepresent the geometry of the larynx and was later found to contradict results from more detailed and adequate experiments [77]. The early studies from van den Berg [78] provided the foundations for many future investigations, as the study linked its experimental results to the Bernoulli equation to describe the transglottal pressure relations. In these experiments, the properties of the glottal impedance were investigated using a model casted from a normal human cadaveric larynx. The glottis was constructed to form a rectangular orifice that was maintained static during the experiments. The resulting glottal impedance from this experiment consisted of a nonlinear glottal resistance (flow dependent) with viscous and kinetic components. The equation describing these results is given by:

$$Z_g = \frac{12\mu h}{l_g d^3} + k\frac{\rho U_g}{2(l_g d)^2} = \frac{12\mu h l_g^2}{A_g^3} + k\frac{\rho U_g}{2A_g^2}, \tag{2.1}$$

where ΔP_g = transglottal pressure drop, U_g = glottal volume velocity, μ = air viscosity coefficient, h = thickness of the glottal slit, l_g = length of glottal slit, d = glottal width, $A_g = l_g d$ = glottal area, ρ = air density, and k= empirical constant that depends on the shape of the glottal slit, where for a rectangular shape $k = 0.875$.

The technique was adopted by the early self-oscillating models of the vocal folds [45, 46], which added a linear glottal inertance to the glottal impedance. Further studies on static configurations were performed using mechanical models by [77,79,80]. These studies were later used to question the validity of the viscous resistance and glottal inertance for unsteady conditions. It was proposed that the air inertia was negligible and that viscous losses could be included in the transglottal kinetic loss by a pressure coefficient [5, 6, 59]. This representation continues to be accepted and used in current studies [22, 42, 74, 81]. Thus, for incompressible, quasi-steady glottal flow, the time-varying transglottal pressure was expressed aerodynamically as [17, 59]

$$\Delta P_g(t) = k_t \frac{1}{2} \frac{U_g(t)^2}{A_g(t)^2}, \tag{2.2}$$

where k_t was an empirically-determined coefficient with an average value of about 1.1 [17, 77, 82], $U_g(t)$ was the time-dependent flow, and $A_g(t)$ was the time-varying glottal area. This previous expression reduces the glottal impedance to

$$Z_g(t) = k_t \frac{1}{2} \frac{U_g(t)}{A_g(t)^2}. \tag{2.3}$$

Additional studies have shown that even though the quasi-steady assumption holds during most of the duty cycle (not including the beginning and end of the glottal pulse, i.e., about 80% of the cycle), important differences in the glottal aerodynamics are observed between convergent (opening) and divergent (closing) glottal configurations [83,84]. It was suggested that these changes could be accounted for via a time-varying orifice discharge coefficient term that was measured experimentally in physical rubber models of the vocal folds [41,84]. These observations are well-aligned

with previous considerations on the role of the inertance and viscous losses [5, 6, 59], and relate the orifice discharge coefficient with a time-varying k_t coefficient in equation (2.3). Similar discrepancies between the aerodynamics of the opening and closing portions have been observed [85]. The separation point of the glottal airflow and the vocal fold walls during the closing portion of the cycle was observed to occur after the minimum glottal area from Bernoulli's equation. Further along these lines, it was suggested that the Coanda effect, where the jet remains arbitrarily attached to one of the vocal folds walls, can occur during divergent configurations [86, 87]. Despite the inability of the Bernoulli regime to properly describe the closing portion of the cycle, all of these recent studies support that its accuracy during the convergent portion is acceptable.

2.1.3 Additional impedance considerations

Two different core options are normally used to represent the origin of the source when expressing the complete system in term of lumped impedances: the use of an ideal pressure or ideal flow source. Although these two are technically interchangeable, they are typically used for different purposes.

Located at the alveolar level, an ideal pressure source tends to produce a more intuitive idea of voice production. However, it fails to represent the dipole nature of the source at the glottis in a circuit analogy. Although some ad hoc measures have been proposed to represent a dipole source using a pressure source (cf. [13] page 92), these require the use of a pressure source at the glottis and no subglottal pressure, thus distorting its original intuitive nature.

The use of an ideal flow source to represent the glottal excitation has also been explored in electrical analog models of voice production [3, 13, 47, 88]. The traditional approach has been to start with a ideal pressure model and convert it into its Norton equivalent, yet the equivalence is valid only for linear elements. Thus, the nonlinear

glottal impedance from equation (2.1) is normally linearized to allow for an ideal volume velocity source.

Two main linearization schemes of the glottal impedance have been used in voice research, differing in that the resulting impedance is either a time-invariant or time-varying term. A small-signal (AC) equivalent source for the glottal resistance proposed by Flanagan (cf. [3] pages 50-53) uses a Taylor series expansion with respect to the mean (DC) components of the signals. This method allows the glottal impedance to be represented by a time-invariant quantity. The flow source is then proportional to the time-varying glottal area with no DC component (i.e., an unloaded glottal airflow source). The main assumption behind this method is that the DC components are expected to be much larger than the time-varying ones. This assumption is expected to hold when there is incomplete glottal closure and very large DC flow component. Therefore, this method has been used by researchers to explore the effects of incomplete glottal closure, female voices, and breathy voice [9, 12, 89].

Another type of linearization uses similar considerations, but linearize the terms with respect to the instantaneous area and unloaded glottal volume velocity. The resulting flow source is again expected to be proportional to the time-varying glottal area but the glottal impedance now becomes a time-varying quantity [47]. It has been stated that this linearization holds only if the glottal impedance is larger than that of the vocal tract, or equivalently, if the supraglottal pressure variations are small compared with the subglottal static pressure [47].

To represent a dipole source in an electric analog representation, researchers have opted for the use of two ideal flow sources in a "push-pull" configuration (see figure 6.1). This representation has been mainly used to study the effects of subglottal cavities their related quantal effects [11, 13, 90, 91]. These studies have focused on the effects of the subglottal coupling in the transfer functions and formant bandwidths and jumps, without exploring their effects in the actual time based volumetric flow signals across the system.

2.1.4 Self-oscillating models of phonation

Self-sustained models of the vocal folds are designed to provide insights into the mechanisms that control phonation in normal and pathological cases. Low-dimensional models are more commonly used, as they efficiently capture the most dominant modes of vibration and are expected to reproduce many fundamental aspects of phonation with acceptable accuracy. Validation of these and other models of phonation continues to be an important research topic in the field.

The first low-order models were the single-mass [45] and two-mass models [46], that made use of van den Berg's glottal resistance [78] along with a glottal inertance and simplified versions of the vocal tract and subglottal system. Even though many components of the original two-mass model (herein referred to as IF72) have been shown to misrepresent the glottal aerodynamics, mechanical properties, and energy transfer [5, 6, 59], the model continues to be one of the most significant and highly referenced models of phonation.

Some improvements of IF72 have included a different solution for the flow equation, that was shown to be more stable and better handled time-varying flow and irregular geometries [59]. A method to correct for discrepancies found in the location of the separation point was implemented in a modified version of IF72 [85]. In addition, this study proposed the idea of a pre-collision mechanism to minimize the excessively abrupt closing observed in IF72. An asymmetric implementation of IF72 was developed to model superior and recurrent nerve paralysis [92]. The model was simplified version of IF72 that neglected its viscous effects, glottal inertance, nonlinear spring constants, and acoustic coupling. However, this model has been repeatedly used to study asymmetric and chaotic behavior within the glottis [93,94] and has been extended using more complex flow solvers [95].

A complete adaptation of IF72 was performed to include a third degree of freedom [42], which helped to better modeling of the mucosal wave and more accurately represent physiological aspects [96]. This "body-cover" model (herein referred to as

ST95) corrected discontinuity problems observed in IF72 and included the aforementioned enhanced flow solution [59]. However, ST95 failed to describe the subglottal system used and was not able to achieve self-sustained oscillations using MRI-based area function reported short after by the same authors [18]. In order to address these issues, a different approach to obtain material properties was proposed using muscle activation rules [97], and the aerodynamic driving forces were updated [74]. In addition, a bar-plate version of the same model was introduced [74].

In a study that simulated the effect of vocal nodules using low-order models, both IF72 and ST95 were compared [98]. It was found that ST95 yielded more realistic representations of the glottal aerodynamics. The study also made use of pre-contact ideas [85] to model the presence of nodules. Good agreement was found between recordings of pathological cases and these simulations.

Even though they were the first representation of self-sustained models, single mass models were essentially neglected since the introduction or IF72. Recent studies have shown that these models can still capture high-order effects if ad-hoc corrections are included, such as a time-varying orifice discharge coefficient [60] or time-varying separation points [99, 100].

Limited efforts have been made to include the effects of incomplete glottal closure in self-sustained models. With a goal of modeling consonant-vowel-consonant gestures, two studies proposed ways to minimize the unrealistic large amplitudes observed in IF72 when the degree of abduction was an increased in the model. These corrections included an anterior-posterior feature that restricted the vibration of the vocal folds [101], and the use of a nonlinear damping coefficient [102]. However, no insights were provided on the energy transfer due to abduction, and no other types of incomplete closure was explored.

High-order models with more degrees of freedom have also been used to study the self-sustained oscillations of the vocal folds. These models use modal analysis and fluid-structure interactions generally via finite elements method [103–105] or a large number of point masses [106], but neglect the effects of acoustic coupling. Only a

few high-order models are known to combine vocal fold tissue mechanics, laryngeal aerodynamics, and acoustic interactions with the subglottal and supraglottal tracts [17, 107, 108]. Predictions using these models confirm the trends observed in low-order models. The effects of incomplete glottal on the source-filter coupling, energy transfer, and glottal airflow have not been explored with these higher-order models.

2.2 Source-filter coupling in voice production

Due to the nonlinear nature of the glottal impedance, any source-filter interaction is a nonlinear phenomenon. A possibly "linear source-filter coupling" has been described [17], where the glottal impedance is much higher than those of the vocal tract and subglottal system and the respective acoustic pressures have no influence on the glottal flow or driving forces for the vocal fold oscillation. This classification essentially describes a no-coupling scenario and thus will not be used in this study. It has been suggested that humans have the ability to operate their voices with and without source filter interactions by changing the degree of adduction of the vocal folds and narrowing of the epilarynx tube [17]. Two possible types of source-filter interactions have been proposed [17]: a "level 1" of interaction, one where only flow and sound interacts, and a "level 2" of interaction, where the tissue is now part of a three-way interaction that affects the tissue motion.

2.2.1 Source-filter interactions: Level 1

Early studies noted that source-filter interactions essentially affected the source properties introducing skewing of the glottal pulse (a delay in its peak with respect to that of the glottal area) and ripples in the open phase [4, 5, 45, 46, 81]. Due to this coupling, the interactive airflow is also referred to as the "true glottal source" [6, 7].

Nonlinear coupling theory [17] summarizes these observations based on the relation between the tract impedances and the harmonic composition of the glottal source, i.e., those harmonics located within the inertive portion of the total tract

impedance are reinforced and those located within the compliant portion are reduced in amplitude. The total tract impedance is associated with the series connection between the sub- and supra-glottal tracts. In addition, harmonic distortion is possible, where, for instance, a sinusoidal glottal area can produce a glottal airflow with additional harmonic components. Even though its is clear that the coupling varies as a function of the tract impedances, no insights on how these harmonic variations and distortions change as a function of glottal impedance are provided in the theory. In addition, it is unclear what "glottal source" is being affected by the tract impedances. Possibly, an ideal, uncoupled source is the one experiencing the harmonic variations. However, this source may not be simply proportional to the glottal area, due to the added harmonic distortion.

Most of the supporting evidence on that the skewing and ripples are related to the vocal tract is based on numerical simulations [5–7, 45, 46, 81]. An alternative explanation of the skewing of the glottal pulse was based on flow separation and vortex shedding [109]. It was also suggested that a degree of skewing of the glottal pulse could be introduced by asymmetric flow behavior within the glottis [86]. Even though it is expected that these phenomena play a minor effect compared with the supraglottal inertia [17], no study has fully demonstrated that the skewing and ripples are due to the vocal tract by decoupling the two systems using in human subject recordings. An attempt to perform this decoupling was performed using a time-varying Norton equivalent airflow representation of the glottal source and a feedback approach [47]. The method had a number of limitations: it only considered a single supraglottal formant and no subglottal coupling, manual intervention was needed to obtain the desired behavior, and only a low degree of de-skewing of the glottal pulse was achieved. Nevertheless, this study is perhaps the strongest supporting evidence using human subject data on that the skewing of the glottal airflow pulse is due to the vocal tract loading.

When the glottis does not fully close during the oscillation, a much lower impedance is observed during the closed phase. This type of voice production is

known as non-modal phonation. Non-modal phonation is observed in different degrees in normal and pathological subjects, and it is associated with different acoustic characteristics, such as breathy voice, irregular vibrations, and failure to phonate [110]. The problem of breathiness was initially explored using a Norton equivalent of the volume velocity source with a parallel glottal impedance [9]. It has been shown that incomplete glottal closure can be related with a clear decay of harmonic energy in the source spectra and an increase of the formant's bandwidth [10–12], as well as the introduction of turbulent noise [10, 12, 13]. A number of descriptors to identify possible non-modal behavior from microphone recordings have been proposed [12].

The effects of incomplete closure and posterior glottal opening on the glottal pulses were studied [14,15] using previously described aerodynamic equations and geometric considerations [46]. These investigations provide the most insightful review on the effects of subglottal coupling on the closed phase portion of the glottal airflow pulse. The studies showed that strong fluctuations could be present during the closed portion of the cycle and described the structure and origin of them. The acoustic coupling between the subglottal and supraglottal tracts is also a consequence of finite glottal impedances. Subsequent studies on this topic illustrated that the acoustic coupling between the vocal tract and the subglottal system results in the introduction of pole-zero pairs in the vocal tract transfer function that correspond to resonances of the uncoupled subglottal tract [13,69,90,91]. Whether this relation holds in the opposite direction, i.e., how the resonances of the vocal tract affect the subglottal ones, has not been explored.

2.2.2 Source-filter interactions: Level 2

In this second type of interaction, the subglottal and supraglottal systems can affect the dynamics of vocal fold vibration. The general conditions where the tract impedances favor or hamper the oscillation have been described in terms of the series superposition of the tract reactances [17]. Even though a complete discussion of the

different combinations and their relation to human speech or singing is provided, no insights are provided on how this type of acoustic coupling changes as a function of the glottal impedance.

Nonlinear theory [17] suggests that for normal speech the vocal tract generally provides an inertive impedance, which favors the oscillation and raises the phonation threshold pressure. Even though the subglottal tract also acts as a inertive impedance, it normally hampers the oscillation. Even with the same tract configurations, different impedance relations can be observed for higher harmonics or high pitch conditions. In any case, it was considered that the most favorable loading conditions were inertive (positive reactance) for the supraglottal tract and compliant (negative reactance) for the subglottal tract.

Different quantitative studies agree with these observations, as seen in analytical models [81, 111], pressure-controlled valves in gas flows [112], numerical simulations [42,60], excised larynx experiments with a vocal tract [113] and a subglottal tract [62], and rubber physical models with different subglottal tracts [63, 114] and supraglottal structures [115, 116].

Favorable conditions for the self-sustained oscillation are normally expressed in terms of a reduced phonation threshold, where a lower lung pressure is needed to initiate the vocal folds oscillation. Likewise, unfavorable conditions raise the phonation threshold. Another evidence of the acoustic contribution to the driving forces could be seen as an increased amplitude of vocal fold oscillation obtained with a normal lung pressure. This effect also relates with the fact that this type of interaction has been related to voice instabilities and bifurcations. However, these effects will be discussed separately in the following sections in the context of voice breaks during phonation.

2.3 Bifurcations in voice production

Voice instabilities and bifurcations, also referred to as voice breaks, have been a long-standing topic of interest in speech science. Even though these singularities are encountered in different circumstances, including the male voice at puberty [117], singing [16, 50], and many different voice pathologies [118], they are not completely understood. In the following sections, the main aspects behind voice bifurcations are described with the aim of differentiating the role that source-filter coupling may play in this phenomenon.

Irregular vocal fold vibration and voice bifurcations are generally considered to be produced by desynchronization between vibratory modes [103, 119, 120], strong asymmetries between left and right vocal folds and/or excessively high subglottal pressure [92, 93], changes in vocal fold tension [121–123], chaotic behavior near limit cycles between registers [124], and nonlinear acoustic coupling [17, 49, 63, 125, 126].

Titze [49] proposes differentiating the origin for voice bifurcation between source-induced and acoustically-induced. This classification was adopted in this thesis research, noting that it is unlikely that the two can be truly separated as unique factors leading to the unstable behavior. Even though it is likely that a combination of components contribute in all cases, it is considered that "source" or "acoustic" ones may be more dominant in certain cases.

2.3.1 Source-induced bifurcations

Observations of unintentional register transitions have been consistently reported in the frequency range of 150 Hz - 200 Hz in males and 300 Hz - 350 Hz in both males and females [127]. Different studies have suggested that instabilities may occur given a physiologic limit in the maximum active stress in the thyroarytenoid (TA) muscle, which controls the medial surface shape of the glottis and thus its main modes of vibration [121, 122, 127]. In this context, a bifurcation explanation from the theory of nonlinear dynamics was proposed to justify jumps occurred when

there was gradual tension transition and even when glottal parameters were held constant [119]. Under this view, frequency jumps exhibiting amplitude differences and hysteresis are classified as subcritical Hopf bifurcations, and frequency jumps exhibiting smooth transitions with no hysteresis are termed as supracritical Hopf bifurcations [124, 128]. A number of studies have suggested that irregular vibration and voice bifurcations can also be produced by desynchronization between vibratory modes. This was initially discovered by analyzing pathological voices with nonlinear dynamics methods [129], and later verified using empirical orthogonal functions (also referred to as empirical eigenfunctions) from node displacement in a finite element model [103], spatio-temporal analysis of *in vivo* high-speed videoendoscopy [120], and a study of the medial surface dynamics in a physical rubber model of the vocal folds [130]. Other studies have suggested alternative explanations, ranging from left-right asymmetry [92], high subglottal pressure [93], coexisting limit cycles [123, 124], and the presence of a vocal membrane [131].

A recent study by [132] investigated the open quotient during register transitions for untrained male subjects. Recordings of high-speed video using a rigid endoscope during upward pitch glides (between 110 Hz and 440 Hz) were obtained. Even though the subjects intended to utter a vowel /i/ to keep an open back cavity, the insertion of the rigid endoscope forced the vocal tract into a neutral configuration closer to a vowel /ae/ with a first formant higher than 500 Hz. Thus, these recordings could not exhibit pitch and formant crossings and the bifurcations observed corresponded to source-induced ones.

2.3.2 Acoustically-induced bifurcations

Early excised larynx studies acknowledged that involuntary register transitions were related to tracheal resonances [44, 133]. Further experiments with both excised larynx and artificial vocal folds have confirmed a noticeable influence of the subglottal and supraglottal resonances on the vocal fold dynamics [62, 63, 113, 114, 116].

Numerical models have also served to explore the effect of acoustic coupling on voice bifurcations. Using a two-mass model with a one-tube resonator, bifurcations and instabilities for weak asymmetries were found [125]. Correlation between bifurcations and supraglottal resonances were found when using a lumped mass model a with straight tube extension [126]. Similar behavior was observed using more descriptive source models, subglottal system, and wave propagation schemes [17, 128].

Evidence of source-filter interaction leading to bifurcations has been reported in human subjects, as seen in register transitions in singers [48], voice breaks in normal and pathological cases [118], and particularly in dynamic vocal exercises [49]. Although some correlations between vocal tract resonances were observed in human subject recordings [121], such effects were considered minor with respect to that of the tension control and not explored in depth. In recordings of different vocal gestures [49], the vocal exercise exhibiting most of the instabilities was found to be pitch (F_0) glides at a soft loudness while producing a sustained vowel that has a low first formant frequency (F_1), such as in vowels /i/ and /u/. Although voice bifurcations could include different phenomena such as frequency jumps, deterministic chaos, aphonation, and subharmonics, crossings between F_0 and F_1 in this vocal exercise primarily yielded frequency jumps. The fact that these voice breaks were more evident in male subjects with no vocal training suggested that muscle control and familiarity with unstable regions can overcome the bifurcations.

A previous attempt to explain voice breaks at low-frequency (e.g., at 150 Hz) was based on constructive and destructive interference between subglottal formants and vocal fold movement. It was suggested that maximum amplitude and minimum amplitude can occur at pitch crossing with specific ratios of the first subglottal formant frequency [134]. A modest amount of support of these predictions was evidenced by means of excised canine larynx experiments [62], where the low consistency among different larynges was deemed to confound the results of the experiment. Although it has been further established that the subglottal tract can hinder vocal fold vibration [17,60,63], the idea of instabilities at entrainments lower than the first subglottal

resonance does not completely fit with the current nonlinear theory [17], and thus has not been further explored.

2.4 Inverse filtering of speech signals

In order to estimate the glottal airflow, the modulating effects of the acoustic loads need to be removed via inverse filtering (IF). This is typically performed using source-filter theory principles, i.e., the acoustic loads can be linearly separated from the source and removed via filter estimation techniques. However, current nonlinear principles bring the fundamental separation between source and filter into question.

Early methods made use of adjustable filter banks to model the vocal tract as a filter. As manual intervention provides no physiologically reliable estimates, this approach is not longer considered acceptable. A more widely used scheme is based on linear prediction based on autoregressive (AR) parametric modeling, where an all-pole model and a least squared error estimation approach is used to approximate the vocal tract filter [135]. This method has been dominant in speech processing due to its computational characteristics, error tractability, stability, and spectral estimation properties [34]. Two methods are commonly used in AR modeling, the autocorrelation and covariance methods. The autocorrelation method assures stable solutions, but requires a large number of samples (including several cycles) to yield a reliable solution. However, long windows are prone to incorrect format estimation due to the underlying harmonic structure of speech and the averaging of source-filter interaction effects [34, 135]. In addition, when applied to the assessment of vocal function, this techniques generally requires manual adjustments of the formant frequencies and bandwidths with the intention of removing fluctuations observed during the open and closed phases [24–26]. On the other hand, the covariance method requires a much smaller number of points, allowing estimation even within a single cycle. Even though this method delivers a stable solution only when the system truly follows

an all-pole system, this instability is easily solved without affecting its magnitude response [32, 88, 136–138].

Many researchers have opted for the covariance method since the smaller temporal windows allow for estimation of the vocal tract filter within the closed phased portion of the vocal fold cycle. An estimate in this portion is thought to be a more accurate representation since it discards the changes in vocal tract formants product of source-filter interactions that are more predominant during the open cycle. A number of methods have been proposed for the detection of the closed phase portion of the cycle, including a multi-channel method that uses electroglottography (EGG) [139,140] and a formant frequency modulation scheme that searches for the stable regions of the first formant [31], among others [33, 34].

In spite of the presence of source-filter interactions, it has been suggested that separability between source and filter is still possible [139]. This can be achieved in two ways, by assuming that: 1) the source is independent and that the vocal tract will have different formants and bandwidths during the open and closed phase, or 2) that the vocal tract is time-invariant as in the closed phase and that the glottal source contains the formant frequency and bandwidth changes from the previous case. The latter explains the vowel dependent ripples normally observed during the open phase of the glottal waveform [6]. In addition, linear prediction applied to a whole period (or more) will contain formant frequency and bandwidth errors [139]. This evidently means that the estimation of the properties of the vocal tract (e.g., formant frequencies and bandwidths) obtained via autocorrelation and covariance methods will differ. In fact, closed phase inverse filtering (CPIF) using the covariance method has been shown to provide better estimates of the glottal waveform than those obtained with the autocorrelation method [34, 138]. In particular, better estimates are those that preserve the source-filter interactions observed during the open phase when removing the vocal tract filtering effect. However, it is unclear how the CPIF scheme handles incomplete glottal closure, where subglottal coupling is present and there is no true closed phase. It is possible that CPIF incorrectly suppresses ripples

during the closed phase that are introduced by sub- and supra-glottal coupling, and thus are physiologically relevant.

Another approach used for inverse filtering of glottal waveforms is the parametric autoregressive moving average (ARMA). This method has been shown to remove the influence of pitch and improve the accuracy of the parameter estimation and spectral matching. However, the estimation of its parameters is a nonlinear problem and suboptimal solutions can be found. In addition to being more computationally intensive, stability and convergence problems can occur when the order of the system is not estimated correctly. Moreover, adequate separation of poles and zeros requires a priori knowledge of the system properties [33, 34]. Therefore, a parametric spectral estimate of the vocal tract using ARMA is normally used along with other methods to counterbalance its drawbacks. The most common complement for an ARMA estimate is the use of a source model, such as the "LF model" [7]. That approach has been used to include separate filter estimates for the closed and open phase vocal tract properties [141], an optimization scheme to fit the LF model [142], and an adaptive Kalman filtering scheme [143]. The main problems with inverse filtering using glottal source models are observed in non-modal and pathological cases, where model provide a poor fit to the actual glottal waveform. This limitation severely affects the application of this type of IF schemes.

As an alternative to CPIF, iterative schemes have been designed to remove the influence of the glottal waveform in the frequency domain. One scheme inverse filtered a signal multiple times with different filters, representing for instance, the glottal tilt and vocal tract filter [144]. This method has a higher computational complexity yet shows no significant improvements over the CPIF [34].

Other spectral estimation methods have been tested to inverse filter glottal waveforms. These approaches explore the use of more recent techniques such as cepstrum and bispectrum. Although these techniques have the potential for improvement in non-minimum phase cases, as with nasal sounds, there are a number of inherent limitations. The main drawback of these methods is that they require large amount of

data points to obtain an acceptable variance, thus necessitating the use of several cycles and affecting the accuracy of the estimates (as in the case of the autocorrelation method of linear prediction). In addition cepstral inverse filtering has shown difficulties due to the phase unwrapping constraints associated with the method. Therefore, these more complex methods tend to be used along with other schemes, such as ARMA or glottal models. Even so, these methods have not shown to be more reliable that the CPIF method previously described.

Nonlinear inverse filtering was designed in an attempt to extract the uncoupled glottal airflow in a volume velocity signal [47]. This method differs from all the above in that it attempts to remove not only the vocal tract filtering, but also its interaction with the glottal source. A time-varying Norton equivalent airflow of the glottal source was used along a feedback procedure. This inverse filter is nonlinear in that certain nonlinear parameters were used. However, the method had a number of limitations that have been previously discussed and resulted in a more complex implementation, and thus the scheme has not been adopted widely.

All the aforementioned studies illustrate that CPIF is currently the most reliable method to estimate the glottal airflow from oral airflow or radiated pressure recordings, particularly for applications related to the assessment of normal and pathological voice function. Its ability to describe the "true glottal flow" have not been revised, particularly under an incomplete glottal closure scenario. Furthermore, the scheme cannot be used to estimate the subglottal tract impedance, as it does not fit with the all-pole requirement of the covariance method of linear prediction.

2.5 Clinical assessment of vocal function

The most common voice disorders are chronic or recurring conditions that are likely to result from inappropriate patterns of vocal behavior referred to as vocal "hyperfunction" [25]. Clinical assessment of vocal function is pivotal to evaluate the presence of vocal hyperfunction before and after any vocal intervention (e.g., surgery,

training). This assessment can be categorized in two ways: perceptual and objective. Perceptual assessments are clinical judgements that rely heavily on auditory and visual interpretations of the subject's voice and endoscopic images, respectively. Although this type of assessment has been dominant in clinical practice, there is an increasing interest in the application of objective measures of vocal function [145–147]. Part of the perceptual assessment has migrated toward a new consensus auditory perceptual evaluation of voice (CAPE-V). In general, the goal of the objective assessment is not to replace the perceptual but to provide tools that help to corroborate clinician's subjective impressions and evaluate the effectiveness of surgery and vocal therapy [145].

The objective assessment of vocal function is normally performed in a clinical setting, where the clinician has access to a larger set of tools and methods. However, the true behavior of patients may be better characterized by a longer term assessment through ambulatory monitoring of vocal function, as individuals engage in their normal daily activities. In the following sections these two types of objective assessments are reviewed separately.

2.5.1 Traditional objective assessment of vocal function

The most common objective measures of vocal function obtained in a clinical setting include endoscopic, aerodynamic, acoustic, and electroglottographic assessments. Although videostroboscopy is an endoscopic imaging method that has well-established clinical value, it still relies on subjective visual assessment and thus does not provide objective data. More recently, other endoscopic imaging methods have been explored such as videokymography [148] and high-speed videoendoscopy [147]. High-speed videoendoscopy has received more attention since it not only facilitates the visual inspection of important parameters that are lost in stroboscopy, but also allows for the quantification of selected vocal parameters via digital video processing. Furthermore, more complete kymographic diagrams can be obtained from high-speed videos [149].

High-speed video is expected to supplement or replace stroboscopy in the assessment of vocal function and promises to provide new insights of the relationships between vocal folds tissue motion and sound production [147, 149].

Aerodynamic assessment considers the use of a circumferentially vented pneumotachograph mask [24] (also referred to as CV mask or Rothenberg mask) that captures oral volume velocity (OVV) and intraoral pressure (IOP) within a bandwidth of approximately 0 Hz to 1.2 kHz [24]. In the clinical practice, however, these signals are used to obtain single-value descriptors such as mean airflow, subglottal pressure, glottal resistance, and phonation threshold pressure. A number of parameters can be extracted when studying the AC airflow signal, including AC/DC ratio, open and closed quotient, minimum and peak flow, and maximum flow declination rate (MFDR), among others. The MFDR, defined as the negative slope of the glottal airflow waveform, is of great interest in the assessment of vocal function, as it has been related to vocal fold closing velocity and vocal fold trauma [25] and correlated with the radiated SPL for inertive vocal tract configurations [150]. In addition, it is suggested that AC flow signal assessment of vocal function can detect pathologies where, in some cases, acoustic assessment do not [98,110]. Furthermore, clinical studies have shown that the AC flow parameters are one of the most salient measures to identify vocal hyperfunction [25, 40].

Acoustic assessment of vocal function is widely used in the clinical practice. Typical measures include sound level pressure (SPL), average fundamental frequency (F_0), jitter, shimmer, and harmonic-to-noise ratio. These measures are generally made from sustained vowels and/or a passage reading. The parameters proposed in [12] (e.g., H1-A1 and H1-A3, where H1 represents the amplitude of the first harmonic and A1 and A3 the first and third formant spectral peaks) are sometimes used to corroborate breathiness and other non-modal descriptions. Other similar measures include the harmonic richness factor (HRF) defined as the difference between the amplitude of the fundamental frequency and the sum of all other harmonics [151], and the difference in the amplitude of the first two harmonics (H1-H2) [27]. Jitter and shimmer,

normally referred to as perturbation analysis, rely significantly on the extraction of fundamental frequency, resulting in an estimation problem which causes the measures to become unreliable for a number of pathological conditions. Furthermore, different possible physiological factors may result in similar values of jitter and shimmer. Because of this, it has been suggested that these parameters do not contribute to the better understanding of the laryngeal physiology or clinical diagnosis [152–154]. Nevertheless, these parameters continue to be used in clinical practice as a reference of the voice periodicity and stability. The limitations of perturbation analysis created the need for different signal processing approaches such as cepstral analysis and nonlinear dynamics [147], although they have not yet been adopted clinically.

Electroglottography (EGG) in the clinical assessment of vocal function appears to be the least-used approach. The technique, where two surface electrodes measure the electrical impedance of the neck, provides estimates of the contact surface between the folds [59], provided some necessary compensation mechanisms (e.g., high-pass filtering) are employed [155]. However, unreliable results can be obtained due to numerous factors, such as compressed ventricular folds, mucous between the true folds, and lack of contact in non-modal phonation. Although these conditions highly distort the EGG signal, which prohibits the estimation of vocal fold contact area, the degraded signal continues to provide additional information regarding vocal fold contact. Thus, the EGG continues to be used for certain applications, particularly for those where the onset of the closed phase portion of the cycle needs to be identified (e.g., CPIF).

2.5.2 Ambulatory assessment of vocal function

Although a complete quantitative evaluation of vocal function can be performed in the clinical environment, many patients undergoing voice therapy may need continuous means of monitoring and/or biofeedback, particularly after voice surgery or when patients have less control of their voices. The technique consists of long-term

recordings of normal vocal activities using a small-weight accelerometer on the skin surface overlying the suprasternal notch. Many advantages of this approach have been exploited. The most attractive features are its non-invasive nature, miniature sensor size, low sensitivity to the background noise, and reduced variability of the recorded signal compared with that of a microphone. The development of this continuous assessment has resulted in a relatively new and commercially available system called the Ambulatory Phonation Monitor (APM) currently marketed by KayPENTAX [38]. However, the adoption of this technology into the clinical practice has been slow due to technical limitations of current devices, the small number of parameters extracted from the signal, and the lack of statistically robust studies that evaluate the true diagnostic capabilities of the ambulatory system.

The ambulatory assessment of vocal function makes uses of measurements of skin vibration on soft tissue [36–38]. Analysis of speech sounds recorded via skin acceleration continues to be a topic of current research in applications such as monitoring of singing techniques [156,157], monitoring of vocal fatigue [158], evaluation of nasalization [159,160], improvement of voice intelligibility in noisy conditions [161], screening of respiratory anomalies by means of pectoriloquy or egophony [162–166], and auditory biofeedback for Parkinson disease patients [167].

A number of challenges are inherent to this technique. Extraction clinical parameters requires further understanding of underlying physical phenomena. Some potential difficulties that have been identified include the coupling between vocal tract and subglottal system and the uncertainty of the tissue properties [35]. In addition, the accelerometer specifications can vary when it is placed on soft tissue [168–170], and air-borne components can corrupt the signal [171]. These factors have limited the accuracy of the clinical parameters extracted from the accelerometer signal. Therefore, the use of neck skin acceleration in voice monitoring requires not only a better understanding of a number of related acoustical issues, but also the design of new inverse filtering approaches that take into account these multiple factors.

2.6 Biosensing considerations for the recordings of neck surface acceleration

The assessment of sound on the skin surface provides a non-invasive approach to examining the circulatory, respiratory, gastrointestinal, and phonatory systems. The bioacoustic transducers used during recordings for this purpose are generally light-weight accelerometers and/or air-coupled microphones attached to the skin. Although intended to capture tissue-borne sounds, the sensors can also detect air-borne transmitted sounds generated by the subject (such as speech) and other sources of background noise. This corrupting signal is generally not taken into account in skin surface measurements under the assumption that it is negligible. The sensitivity to air-borne sounds is rarely provided in the manufacturer specifications, and if reported, such information often cannot be related directly to the actual performance of the sensor when it is mounted on a compliant surface such as human skin. Although there are some guidelines and standards to evaluate air-borne and structure-borne sensitivity of sensors [172, 173], these guidelines are neither intended nor fully appropriate for biomedical acoustics applications. Even though there have been efforts to compare the tissue-borne sensitivity of commonly used bioacoustic sensors [174–177] and standardize these measurements [169, 170] for skin applications, the air-borne sensitivity of bioacoustic sensors and its relation with the tissue-borne sensitivity has not been consistently investigated.

During recordings of lung and heart sounds, typical sources of background noise that are transmitted via air-borne pathway are normally the subject's respiration and room noise. These components tend to be sufficiently uncorrelated with the biomedical signals of interest, the latter of which can be retrieved by means of signal processing techniques [178, 179]. In addition, it is common in clinical practice to make these recordings under relatively low levels of background noise, thus minimizing its influence during the recordings. However, skin surface recordings that require the subject to speak impose a larger challenge. Under this scenario, the sub-

ject's radiated voice in the room will generate a larger background noise that can be highly correlated with the skin vibration, making discrimination between air-borne and tissue-borne components difficult to achieve. As described in the previous section, recordings of voiced sounds at the skin surface have been used in a variety of clinical studies. However, little or no attention has been paid to the effect of corruption in the tissue-borne signal due to undesired air-borne components. The adverse effect of air-borne sound may hamper observations that affect clinical monitoring and diagnosis of vocal and respiratory pathologies. The corrupting effect introduced by air-borne transmitted components may help to explain some discrepancies and uncertainties observed in these and other related studies [180, 181]. Therefore, understanding the relative sensitivity of bioacoustic sensors to the tissue-borne and air-borne components becomes a critical point in order to formulate accurate claims based on skin surface measurements.

An ideal sensor would have a high tissue-borne sensitivity and low (or null) response to air-borne transmitted sounds. Actual bioacoustic sensors have frequency dependent sensitivities and limited bandwidths of operation, making the criteria to discriminate between air-borne and tissue-borne components more challenging. Defining procedures to measure the sensitivities for each component and introducing simple guidelines to discriminate between them are the primary goals of this study. Although the study is primarily focused on the evaluation of skin surface measurements of voiced sounds, many of the approaches and findings can be extrapolated to other applications such as lung and heart sounds measurements.

3. NUMERICAL SIMULATIONS OF ACOUSTIC COUPLING DURING COMPLETE AND INCOMPLETE GLOTTAL CLOSURE

This chapter describes investigations via numerical simulations on acoustic coupling during incomplete glottal closure. The overall aim was to extend previous studies of source-filter interactions and their effect on tissue dynamics, energy transfer, and glottal aerodynamics during the presence of incomplete glottal closure. The methods presented in this chapter provide additional insights on nonlinear source-filter coupling and an extension of lumped element models of phonation.

3.1 Methods

Two different methods were used to explore the coupling between vocal fold motion, glottal airflow, and resulting sound waves during incomplete glottal closure: a driven glottal area scenario and a self-sustained lumped mass model. These methods were selected to mimic previous explorations that provided the foundations of nonlinear source-filter coupling [17].

The selected lumped element model of the vocal folds was the so-called "body-cover" model [42], referred to as ST95 in this thesis document. This model was selected over the traditional two-mass model [46] (previously referred to as IF72) due to its more physiologically meaningful nature, enhanced representation of glottal impedance terms [42], better stability of its associated glottal airflow solver [59], and ability to more accurately mimic pathological conditions [98]. Furthermore, the "body-cover" model was adaptable to represent the traditional two-mass model (by

simply increasing the body stiffness) and was shown to be analogous to a more recent "bar-plate" model (when using updated aerodynamic equations) [74].

Sound propagation for the subglottal system and vocal tract was modeled using a wave reflection analog scheme. This time-domain transmission model was selected since it has been widely used in numerical simulations of phonation. The implementation included a radiation impedance [58] and different loss factors for the subglottal and supraglottal tracts [60, 182]. Unless otherwise stated, the vocal tract area functions were estimated from 3D cine-MRI data of an adult male [21]. The subglottal area function was adapted from respiratory system measurements of human cadavers [70] and included the trachea, bronchi, and a resistive termination impedance. The glottal airflow model was obtained according to Titze [59], which was originally designed to describe nonlinear coupling using a wave reflection analog scheme. A constant subglottal (lung) pressure of 0.8 kPa was employed unless otherwise stated. To match the MRI representations, a time step associated to a sampling frequency of 70 kHz was used in the 4th order Runge-Kutta solver.

3.1.1 Including incomplete glottal closure

The primary goal of this research effort was to describe incomplete glottal closure as given by a parallel posterior glottal gap (also referred to as glottal chink). In order to account for anterior-posterior differences, a method that divided the glottis into multiple flow sections and combined them into a single flow channel [108] was followed. This method assumed that the minimum glottal area was the sum of the minimum areas in the flow channels. Interestingly, no self-oscillating models of phonation have incorporated this method to represent a posterior glottal gap. Thus, a gap area (A_{gap}) was added to that of the membranous vocal folds (A_{mem}) to represent the minimum glottal area ($A_g = A_{mem} + A_{gap}$) that also corresponded to the visual area that could be obtained via laryngeal high-speed videoendoscopy. Similar assumptions on the relation between the minimum area of the vocal folds and a posterior gap have been

made [15, 183, 184]. This new minimum area was then introduced into the glottal airflow solver [59] given by:

$$U_g = \frac{A_g c}{k_t}\left\{-\left(\frac{A_g}{A^*}\right) \pm \left[\left(\frac{A_g}{A^*}\right)^2 + \frac{4k_t}{\rho c^2}(p_s^+ - p_e^-)\right]\right\}. \tag{3.1}$$

The proposed method to describe incomplete glottal closure was contrasted with two existing methods that have been used to describe related phenomena using the IF72 model. These two methods included an increased abduction scheme [102] and a pre-collision scheme [85]. The former was based on an increase of the rest position of the vocal folds with an added nonlinear damping term, and was originally designed to mimic consonant-vowel-consonant gestures. The latter made use of elastic changes due to collision forces that were activated prior vocal fold contact and has been used to model the effect of nodules [85, 98].

The presence of a posterior glottal opening could be sufficiently large to reach the oscillating (membranous) portion of the glottis, which has been referred to as a "linked leak" [15,185]. This case was expected to follow the abduction descriptions from [102] combined with a parallel gap representation. In addition, it is well-established that the presence of incomplete glottal closure is associated with an increase in turbulent noise [12, 41]. This component was not included in these initial investigations.

3.1.2 Self-sustained model: general considerations

The original ST95 implementation [42] illustrated the importance of adding physiologically relevant features in reduced order models of phonation. Thus, detailed methods to obtain material properties were described and contrasted with excised larynx data. However, subsequent studies on vocal tract imaging [18] from the same authors elucidated that the driving forces within the model and/or its material properties made the model unable to self-oscillate with these MRI-based descriptions. Thus, updated methods to obtain material properties and driving forces were subse-

quently proposed [74, 97]. However, no consistent review of the factors that affected the original model have been reported.

More recent investigations suggested that the presence of asymmetric flow behavior was considerable and affected the driving forces of the vocal folds [86, 86, 186]. A boundary layer estimation of asymmetric pressures (BLEAP) scheme was developed from a similarity solution of measured data from experiments with scaled up, driven vocal folds [187]. This scheme proposed an asymmetric method to compute driving forces during the divergent portion of the cycle and was designed to work with any lumped mass representation of the vocal folds. However, this new method has not been applied to a self-sustained oscillation or compared with the symmetric aerodynamic forces proposed in [74]. The implementation of this scheme required extending ST95 to an asymmetric configuration, to allow the jet to fluctuate between vocal fold walls randomly.

A simple parametric study was included to compare different driving forces and material properties with the aim of identifying a more accurate representation of actual phonation for complete and incomplete glottal closure conditions. Thus, four driving forces and four sets of material properties were used during simulations of phonation with and without a posterior glottal gap. The driving forces were: 1) the original from the ST95 model, 2) the updated version from [74], 3) the updated version from [74] including the effects of hydrostatic pressure, and 4) the BLEAP scheme [187]. The material properties considered for these parametric observations were: a) a set from the original ST95 model (Case C) matching description from Hirano [96] b) an adaptation of the material properties from IF72 as suggested in [42], c) an updated set of material properties taken from [97], and d) a set obtained using reduction of order methods from a finite element model of the vocal folds [188]. All cases were selected to represent modal phonation (chest register) with normal male fundamental frequencies. The material properties for each case are presented in Table 3.1.

Selected measures of glottal behavior were used to quantify the effects of coupling. The glottal impedance measure Z_g *mean* was computed using mean airflow and area

values in equation (2.3) (equivalent to dividing mean transglottal pressure with mean glottal airflow in this model). Other measures such as AC airflow, MFDR, H1-H2 and HRF (previously defined in chapter 2) provided insights on the degree of acoustic coupling, pulse skewing, and harmonic distortion.

3.1.3 Self-sustained model: driving forces and material properties

The equations describing the asymmetric extension of the body-cover model and the cases under study are presented in this section. For further details on the underlying assumptions and details of the self-oscillating model please consult the original ST95 implementation [42] and its the updated version [74]. A representation of the model is presented in Fig. 3.1.

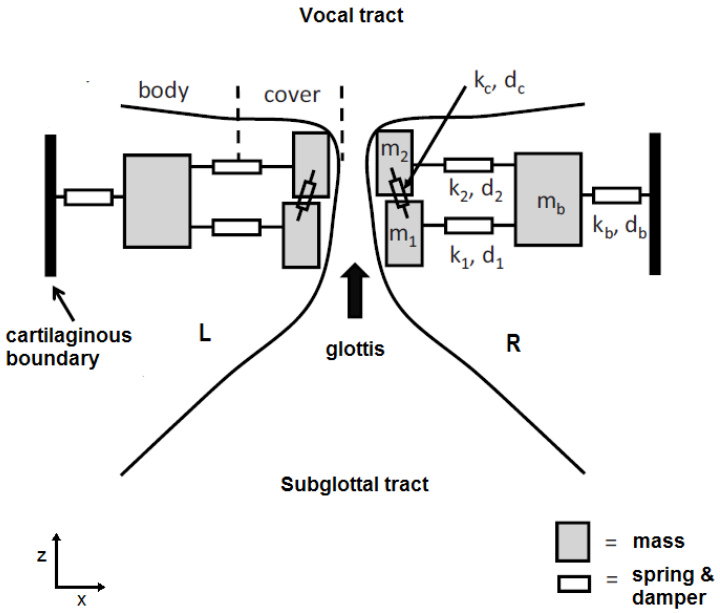


Fig. 3.1. Body-cover model. Adapted from [32].

The equations of motion for the six masses were written in terms of the coupling forces and external driving forces that were exerted on each mass. A compact notation

was used, where the index α indicated left (L) or right (R) vocal folds. Even though asymmetric driving forces were used in some cases (e.g., for the BLEAP driving forces), all material properties were symmetric. Thus, the resulting equations of motion were:

$$F_{1\alpha} = m_1\ddot{x}_{1\alpha} = F_{k1\alpha} + F_{d1\alpha} + F_{kc\alpha} + F_{e1\alpha} + F_{Col1\alpha} \tag{3.2}$$

$$F_{1\alpha} = m_2\ddot{x}_{2\alpha} = F_{k2\alpha} + F_{d2\alpha} - F_{kc\alpha} + F_{e2\alpha} + F_{Col2\alpha} \tag{3.3}$$

$$F_{b\alpha} = m_b\ddot{x}_{b\alpha} = F_{kb\alpha} + F_{db\alpha} - [F_{k1\alpha} + F_{d1\alpha} + F_{k2\alpha} + F_{d2\alpha}], \tag{3.4}$$

where the subscripts $x_{1\alpha}$, $x_{2\alpha}$, and $x_{b\alpha}$ were the (L & R) lower, upper, and body mass displacements, respectively, and the following forces were identified:

$F_{d1\alpha}$, $F_{d2\alpha}$, $F_{db\alpha}$: forces due to damping,
$F_{k1\alpha}$, $F_{k1\alpha}$, $F_{kb\alpha}$: lateral spring forces,
$F_{kc\alpha}$: spring forces due to coupling of m_1 and m_2,
$F_{Col1\alpha}$, $F_{Col2\alpha}$: forces generated only during collision with the opposite vocal fold,
$F_{e1\alpha}$, $F_{e2\alpha}$: external forces generated by the glottal airflow, and
$F_{1\alpha}$, $F_{2\alpha}$, $F_{b\alpha}$: forces of the accelerating masses.

The expressions for these forces were identical to those from ST95 [42] and were not be repeated here. However, the expressions for the external forces during the open phase were presented for the three cases under study to emphasize the nature of each case.

Original driving forces

The original external driving forces in ST95 [42] were symmetric expressions given by $F_{e1\alpha} = F_{e1} = P_1 L_g T_1$ and $F_{e2\alpha} = F_{e2} = P_2 L_g T_2$, where L_g was the vocal fold length, T_1 and T_2 were the thicknesses from the lower and upper masses, and P_1 and P_2 were the pressures applied to the lower and upper masses, given by:

$$P_1 = P_s + p_s - (P_s + p_s - p_e)(a_{min}/a_1)^2 \tag{3.5}$$

$$P_2 = p_e, \tag{3.6}$$

where P_s was the steady component of the subglottal pressure, p_s was the unsteady (acoustic) component of the subglottal pressure, p_e was the unsteady (acoustic) component of the supraglottal pressure, $a_{min} = min(0, a_1, a_2)$ was the minimum glottal area, and $a_1 = (x_{1L} + x_{1R})L_g$ and $a_2 = (x_{2L} + x_{2R})L_g$ were the areas of the lower and upper masses.

Updated and hydrostatic driving forces

The external driving forces in the updated equations for ST95 described in [74] also were symmetric expressions given by $F_{e1\alpha} = F_{e1} = P_1 L_g T_1$ and $F_{e2\alpha} = F_{e2} = P_2 L_g T_2$, where the pressure terms were given by the following set of equations:

$$P_1 = P_s + p_s - P_{kd}(a_2^2/(a_1 a_n)), \tag{3.7}$$

$$P_2 = P_s + p_s - P_{kd}(a_2/a_n), \tag{3.8}$$

for $a_1 \geq a_2$, otherwise

$$P_1 = P_s + p_s - P_{kd}(a_d^2/(a_1 a_n)) \tag{3.9}$$

$$P_2 = P_s + p_s - P_{kd}/T_2(T_1 + T_2 - z_d + (a_d/a_n)(z_d - T_1)), \tag{3.10}$$

for $z_d \geq T_1$, otherwise

$$P_1 = P_s + p_s - P_{kd}/T_1(T_1 - z_d + (a_d/a_1)z_d), \tag{3.11}$$

$$P_2 = P_s + p_s - P_{kd}, \tag{3.12}$$

where P_{kd} was the kinetic pressure, k_e was the pressure recovery coefficient, a_n was the area at the nodal point, a_d was the flow detachment area, and z_d was the separation point. These expressions were computed using:

$$P_{kd} = (P_s + p_s - p_o)/(1 - k_e), \tag{3.13}$$

$$k_e = 2a_d/A_e(1 - a_d/A_e), \tag{3.14}$$

$$a_d = max(0, min(1.2a_1, a_2)), \tag{3.15}$$

$$a_n = a_1 + T_1(a_2 - a_1)/(T_1 + T_2), \text{ and} \tag{3.16}$$

$$z_d = min(T_1 + T_2, max(0, (a_1 - a_d)(T_1 + T_2)/(a_1 - a_2))). \tag{3.17}$$

An additional driving force investigated in this chapter used the same equations but included an hydrostatic pressure term during complete of the vocal folds instead of a no pressure condition. This hydrostatic pressure was given by the average between subglottal and supraglottal pressures, i.e., $P_{hydro} = (P_s + p_s + p_e)/2$.

BLEAP driving forces

The BLEAP scheme [187] is described below in relatively more detail since it has not been applied to a lumped mass model representation before. The scheme used the original equations of ST95 for the converging portion of the glottal cycle (i.e., $a_1 \geq a_2$) and an asymmetric pressure distribution otherwise. The asymmetric pressure was associated to a jet that was randomly attached to either vocal fold and was derived from the distribution of the flow velocity such that:

$$v(z) = z_0 \left(\frac{z + z_{\text{off}}}{z_{\text{off}}} \right)^n, \tag{3.18}$$

where $z_{\text{off}} = 0.15/7.5\ cm$, $n = -0.0148$, and v_0 was the flow velocity at the glottal entry. In these simulations, an acoustic formulation for this entry velocity was proposed. Initially, the velocity at the glottal exit ($z = z_e$) was computed from Eq. (3.18) such that:

$$v(z_e) = z_0 \left(\frac{z_e + z_{\text{off}}}{z_{\text{off}}} \right)^n, \tag{3.19}$$

with the glottal exit given by $z_e = T_1 + T_2$. Applying Bernoulli's equation between the glottal entry and exit and substituting terms yielded

$$v_0 = \sqrt{\frac{2(P_s + p_s - p_e)}{\rho}} \left(\frac{z_{\text{off}}}{z_e + z_{\text{off}}} \right)^n. \tag{3.20}$$

A Bernoulli regime was assumed from the subglottis to compute the pressure at the glottal entry using

$$P_o = P_s + p_s - \frac{1}{2}\rho v_o^2, \tag{3.21}$$

which was then utilized to obtain the pressure distribution across the glottis

$$\begin{aligned} P_B(z) &= P_o + \frac{1}{2}\rho v_o^2 \left[1 - \left(\frac{z + z_{\text{off}}}{z_{\text{off}}} \right)^{2n} \right], \\ &= P_s + p_s - \frac{1}{2}\rho v_o^2 \left(\frac{z + z_{\text{off}}}{z_{\text{off}}} \right)^{2n}, \\ &= P_s + p_s - (P_s + p_s - p_e) \left(\frac{z + z_{\text{off}}}{z_{\text{off}}} \right)^{2n}. \end{aligned} \tag{3.22}$$

Finally, the pressure distribution was integrated over the entire glottis to obtain the external forces that were randomly applied (uniform distribution) during each cycle to either the left or right masses, such that:

$$F_{e1\alpha} = \int_0^{T_1} P_B(z) dz, \text{ and} \tag{3.23}$$

$$F_{e2\alpha} = \int_{T_1}^{T_2} P_B(z) dz. \tag{3.24}$$

In addition, the material properties utilized in the numerical simulations are listed in Table 3.1

Table 3.1
Material properties used for the parametric observations. Parameter values were selected to yield similar fundamental frequencies of oscillation.

Material property	Notation	Units	IF72	Case C	Rules	FEM
lower mass thickness	T_1	cm	0.25	0.15	0.15	0.24
upper mass thickness	T_2	cm	0.05	0.15	0.18	0.20
vocal fold length	L_g	cm	1.40	1.00	1.41	1.25
lower mass	m_1	g	0.125	0.01	0.070	0.071
upper mass	m_2	g	0.025	0.01	0.080	0.086
body mass	m_b	g	0.05	0.05	0.130	0.05
lower spring constant	k_1	N/m	80.0	5.0	56.4	59.1
upper spring constant	k_2	N/m	8.0	3.5	64.4	72.1
coupling stiffness	k_c	N/m	25.0	2.0	5.8	9.1
body stiffness	kb	N/m	100000	100	301.9	100000
lower stiffness (closed phase)	k_{1c}	N/m	240.0	15.0	169.2	177.2
upper stiffness (closed phase)	k_{2c}	N/m	24.0	10.5	193.3	216.4
lower damping ratio	η_1	-	0.1	0.4	0.1	0.1
upper damping ratio	η_2	-	0.6	0.4	0.6	0.6
body damping ratio	η_b	-	0.2	0.2	0.1	0.1
lower damping coeff.	d_1	-	0.02	0.01	0.01	0.01
upper damping coeff.	d_2	-	0.02	0.005	0.09	0.09
body damping coeff.	d_b	-	0.01	0.004	0.04	0.004
lower damping coeff. (closed phase)	d_{1c}	-	0.22	0.02	0.14	0.14
upper damping coeff. (closed phase)	d_{2c}	-	0.05	0.02	0.23	0.25
lower mass rest position	x_{01}	cm	0.018	0.025	0.0001	0.020
upper mass rest position	x_{02}	cm	0.018	0.025	0	0.020
body mass rest position	x_{0b}	cm	0.3	0.3	0.3	0.3
air density	ρ	Kg/m^3	1.13	1.13	1.13	1.13
nonlinear stiffness coeff.	ν	-	100	100	100	100
nonlinear stiffness coeff. (closed phase)	ν_c	-	500	500	500	500

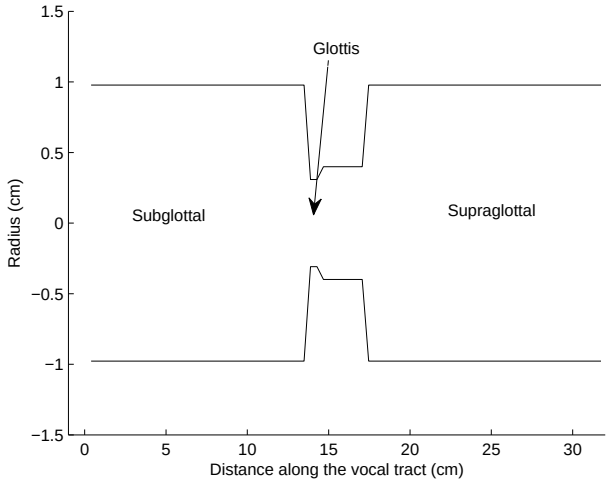
3.2 Results

The results of the investigations using both driven glottal area and self-sustained models are presented in this section. The effects of incomplete glottal closure were initially explored via the driven model. A parametric analysis was performed on the self-sustained model to select the most realistic scenario available for further observations. A selected model configuration was used to assess the coupling differences between vowels, types of incomplete closure, and to provide insights on vocal hyperfunction.

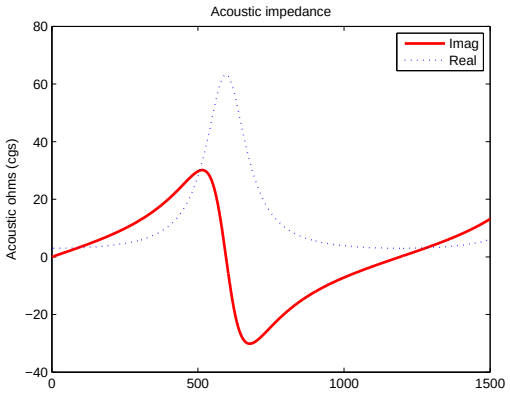
3.2.1 Driven glottal area

In this initial section, the results from previous studies [17] were replicated and extended to provide further insights on the effects of coupling due to the presence of incomplete closure. Thus, the example of Fig. 2 in [17] was used as a reference. An imposed (driven) glottal area was connected to simplified vocal tract and subglottal systems through a wave reflection analog technique, including radiation impedance for the vocal tract and losses for each tract as in [60, 182]. The geometries and impedances of the tracts for this configuration are shown in Fig. 3.2.

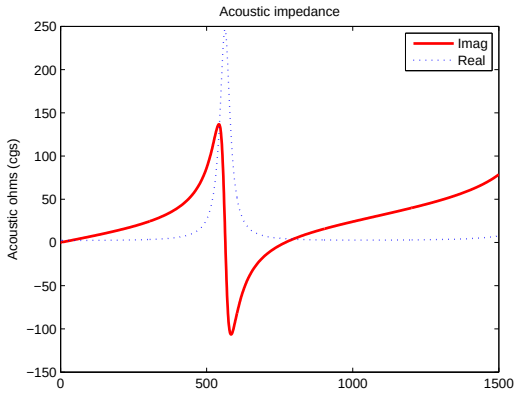
A driven glottal area given by $A_g = min(0, 0.15\cos(2\pi F_0 t) + A_{off})\ cm^2$ was used to simulate three different glottal conditions for two fundamental frequencies (F_0). First, the simulation presented in [17], where a 200 Hz tone and no contact (i.e., $A_{off} = 0.15\ cm^2$) was replicated. Using the same pitch, two additional scenarios with different offsets were simulated: incomplete glottal closure (i.e., $A_{off} = 0.25\ cm^2$) and complete closure (i.e., $A_{off} = 0.05\ cm^2$). The same cases were then repeated using $F_0 = 100\ Hz$. For each case, the glottal airflow was computed using the glottal area and acoustic pressures via equation (3.1). The interactive airflow (U_g) was contrasted with the uncoupled one (U_o) given by:



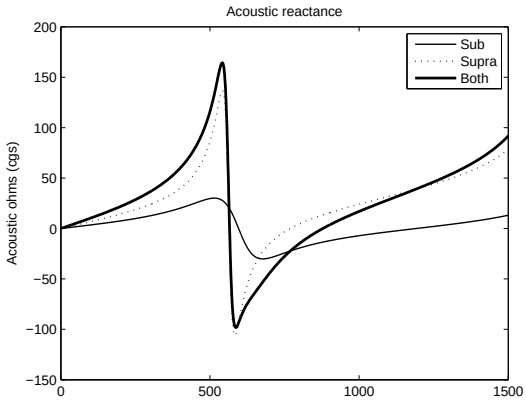
(a) Simplified tract geometries for driven glottal area cases



(b) Subglottal impedance



(c) Supraglottal impedance



(d) Reactance of tracts and their series connection

Fig. 3.2. Tract geometries and impedances used for the driven area observations

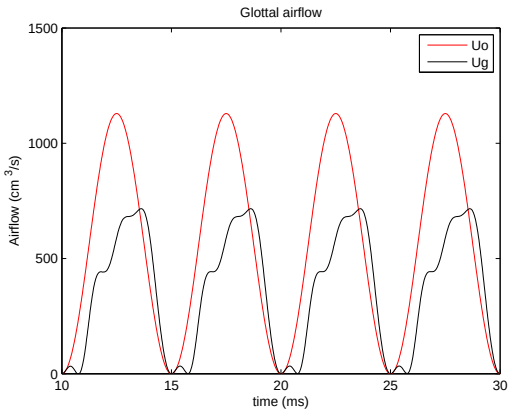
$$U_o(t) = A_g(t) \cdot \sqrt{\frac{2\ P_s}{\rho}}, \tag{3.25}$$

where P_s is the steady subglottal (lung) pressure, which was set to 0.8 kPa in these simulations.

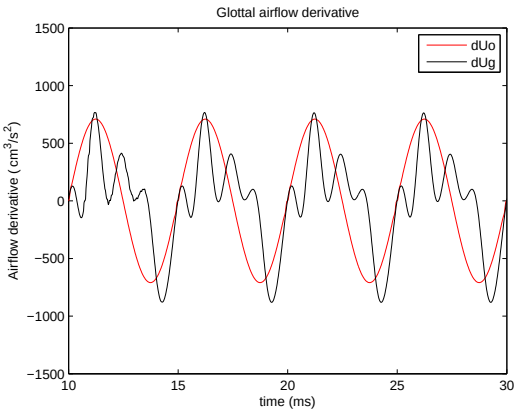
The airflow and its derivative for the selected cases are presented in Fig. 3.3. The reference example from [17] (Fig. 3.3-a) yielded the acoustic coupling altering the airflow with three key features: ripples in its open phase, skewing of the glottal pulse, and an apparent closed phase (where there was none in the glottal area). These features created important "harmonic distortion" with respect to the uncoupled airflow (i.e., new harmonic components not present in U_o). It was noted that the presence of complete glottal closure increased these distortions in the glottal pulse, thus enlarging the ripples and skewing (the latter was also noted in the larger negative peak of the airflow derivative). However, during incomplete glottal closure, the opposite trend was observed. Glottal pulses exhibited less skewing and ripples and oscillated at a higher level and with smaller excursion.

The same cases were then simulated again using a lower fundamental frequency and the results shown in Fig. 3.4. The same trends as before were observed, but in all cases the amount of harmonic distortion was considerably lower, indicating that the coupling phenomena was frequency (F_0) dependent.

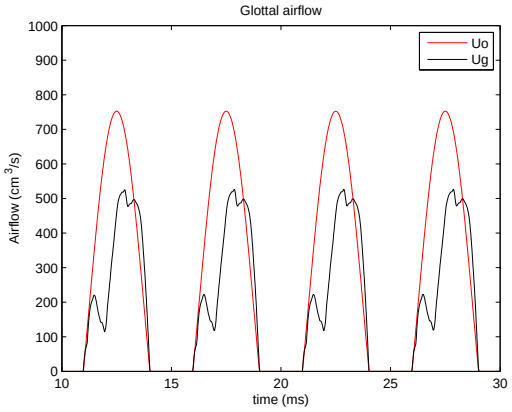
The dependence of acoustic coupling on frequency and glottal closure for a given tract configuration was quantified through glottal measures extracted for each case. Five measures were selected for this case, each one evaluated for the uncoupled glottal airflow (U_o) and interactive glottal airflow (U_g). The amount of coupling and harmonic distortion was evaluated by comparing the measures from the uncoupled and interactive glottal airflows, where AC airflow and harmonic measures were always smaller for all cases and MFDR was more sensitive to the glottal conditions. Incomplete glottal closure exhibited smaller AC airflow, MFDR, and lower mean impedances, elucidating that the nature of the harmonic distortions in this case had



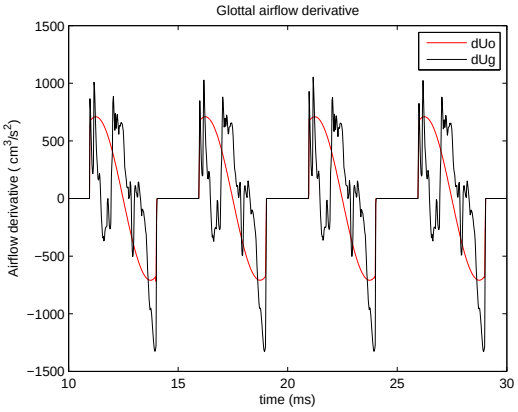
(a) Near closure: airflow



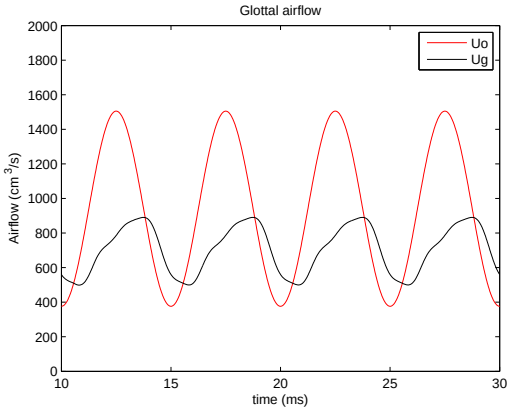
(b) Near closure: airflow derivative



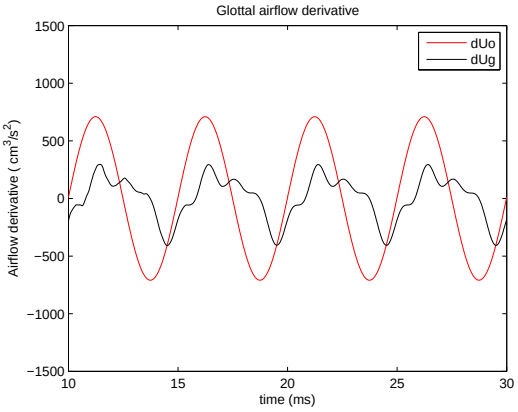
(c) Complete closure: airflow



(d) Complete closure: airflow derivative

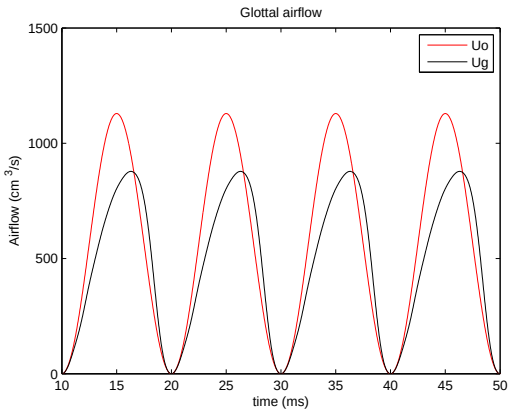


(e) Incomplete closure: airflow

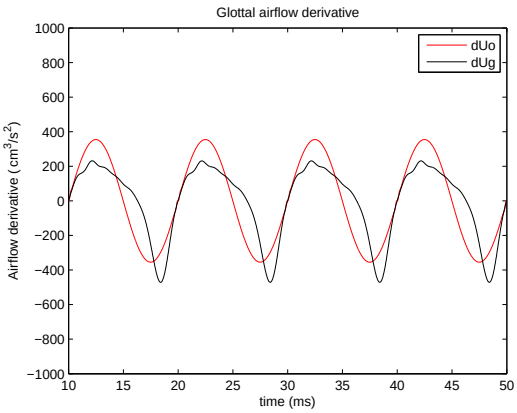


(f) Incomplete closure: airflow derivative

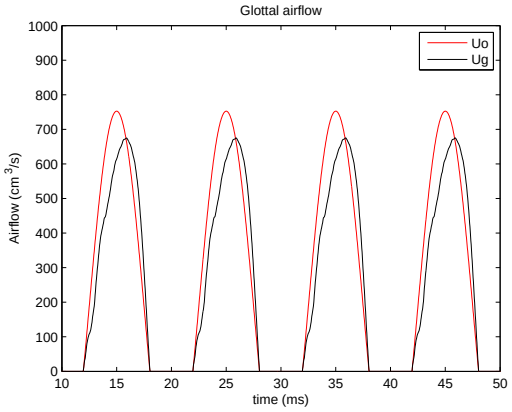
Fig. 3.3. Driven model for $F_0 = 200$ Hz: airflow and its derivative for a highly coupled scenario and its variation with respect to the degree of abduction



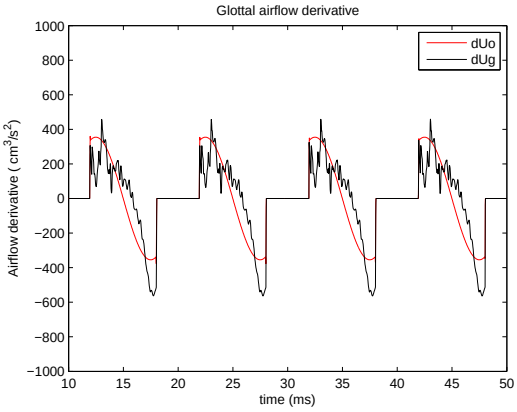
(a) Near closure: airflow



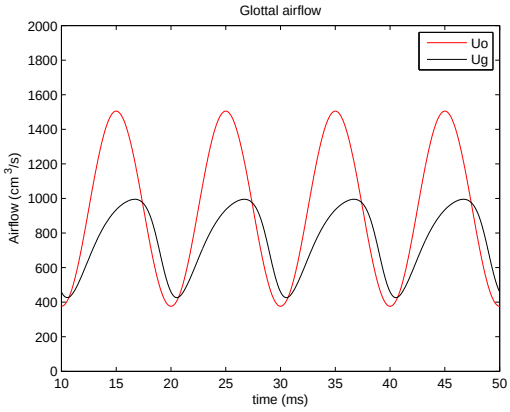
(b) Near closure: airflow derivative



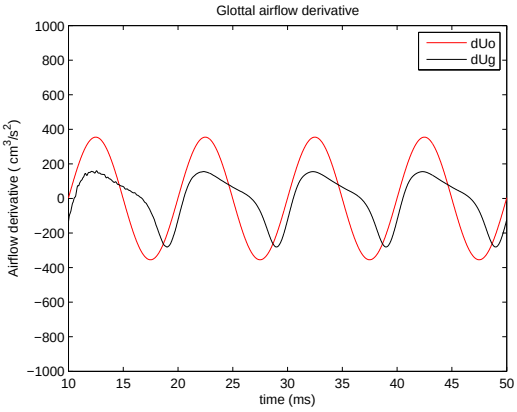
(c) Complete closure: airflow



(d) Complete closure: airflow derivative



(e) Incomplete closure: airflow



(f) Incomplete closure: airflow derivative

Fig. 3.4. Driven model for $F_0 = 100$ Hz: airflow and its derivative for a less coupled scenario and its variation with respect to the degree of abduction

a different origin. In all cases, the harmonic distortion was found to increase for the higher frequency simulations.

The frequency dependence of the acoustic coupling can be explained based on the tract impedances, as suggested in [17]. Harmonics lying on regions of large positive reactance were amplified, and viceversa. The high frequency scenario placed the first few (stronger) harmonics in a region near this change, thus significantly amplifying many harmonic components (and creating noticeable distortion). The lower frequency scenario placed most of the significant harmonics in the inertive yet low amplitude reactance area, thus creating less significant harmonic gain.

The dependance on the degree of glottal closure appeared less obvious since the incomplete glottal closure cases exhibited a stronger coupling (evidenced in the smaller AC airflow), but with less significant harmonic distortion (evidenced in the smaller MFDR). These observations were explained by exploring the nature of equation (3.1). The term A_g/A^*, normally referred to as coupling parameter [17], became larger due to the offset of the glottal area, thus reducing the amplitude of the AC airflow and making it proportional to the transglottal drop term $(p_s^+ - p_e^-)$. However, since no complete closure was achieved, the transglottal pressure term was smooth and relatively proportional to the glottal area, thus reducing the harmonic distortion of the incomplete glottal closure case. These observations were consistent with the fact that a lower mean impedance was expected to exhibit larger acoustic coupling.

Table 3.2
Glottal measures for uncoupled glottal airflow (1) and interactive glottal airflow (2). Notation: Titze: $A_{off} = 0.15\ cm^2$, Titze: $A_{off} = 0.05\ cm^2$, and Titze: $A_{off} = 0.25\ cm^2$.

		$f0$ = 200 Hz			$f0$ = 100 Hz		
Measure	Units	Titze	Low	High	Titze	Low	High
Zg mean	Ω (cgs)	10.3	20.9	7.1	13.9	29.5	7.8
AC airflow 1	cm^3/s	1129	753	1129	1129	753	1129
AC airflow 2	cm^3/s	716	526	391	878	675	569
MFDR 1	cm^3/s^2	709	711	709	355	365	355
MFDR 2	cm^3/s^2	880	1327	407	471	564	281
H1-H2 1	dB	-69.3	-11.7	-69.2	-48.6	-12.2	-48.6
H1-H2 2	dB	-12.1	-8.4	-14.0	-13.8	-12.1	-12.9
HRF 1	dB	-67.1	-11.1	-69.2	-48.5	-11.6	-48.6
HRF 2	dB	-10.5	-6.1	-13.2	-13.3	-10.4	-12.6

3.2.2 Self-sustained model: parametric observations

The self-sustained ST95 vocal fold model was tested with four different versions of driving forces (original, updated, hydro, and BLEAP) and three parameter sets describing material properties (IF72, case C, and rules). All cases were evaluated with and without acoustic coupling and with complete and incomplete glottal closure. The aim was to find those conditions that more closely resembled actual phonation for modal male voices with a normal degree of incomplete closure. A constant steady subglottal pressure of 800 Pa and a gap area of 0.03 cm^2 were used for all cases. The vocal tract was selected for a vowel /e/ from a MRI database [21], and the subglottal tract was based on [70] including trachea, bronchi, and a resistive termination impedance.

Data from recordings on human subject from [39, 189] were used as a reference to identify the desired behavior in the model. Two data sets were summarized and presented in Table 3.3, based on different signal processing schemes, where "1988" referred to [39] and "1993" to [189]. Given that the latter results were considered more accurate, the 1993 measures were linearly extrapolated to match the steady subglottal (lung) pressure used during the simulations (0.8 kPa $\approx$ 7.84 cm H_2O).

Table 3.3
Summary of selected measures for human subject recordings [189] and linear interpolation to match a 0.8 kPa mean subglottal pressure

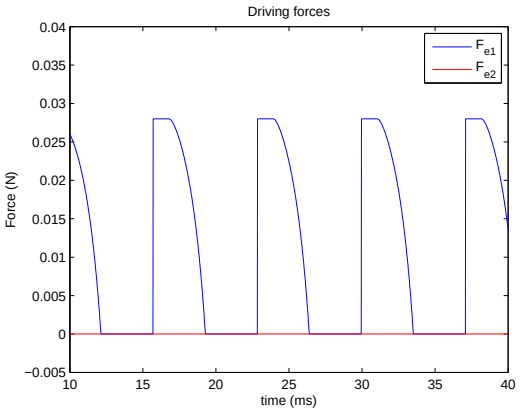
		Male-Normal		Male-Loud		Male-Normal/Loud*
Measure	Units	1988	1993	1988	1993	1993*
F0	Hz	116 ± 12	112.4 ± 11.8	126 ± 14	128.9 ± 24.7	124.3 ± 20.1
MFDR	cm^3/s^2	279.6 ± 90.4	337.2 ± 127.2	481.1 ± 162.6	650 ± 251.2	516.7 ± 193.3
SPL	dB	79.5 ± 3.4	77.8 ± 4	86 ± 4.3	85.6 ± 4.7	83.4 ± 4.2
AC airflow	l/s	0.26 ± 0.07	0.33 ± 0.07	0.38 ± 0.1	0.47 ± 0.14	0.39 ± 0.11
min airflow	l/s	0.12 ± 0.07	0.08 ± 0.05	0.11 ± 0.05	0.08 ± 0.05	0.09 ± 0.06
mean airflow	l/s	0.19 ± 0.07	0.2 ± 0.06	0.2 ± 0.06	0.24 ± 0.09	0.22 ± 0.08
Zg mean	ohm (cgs)	37.7 ± 16.7	32.6 ± 13.7	48.6 ± 19.4	41.4 ± 18.7	40.07 ± 17.98
P_S	cmH_2O	6.3 ± 1.4	5.9 ± 1.1	9 ± 2.4	8.7 ± 2.4	7.84 ± 2.04

Additional criteria to select the desired configuration were based on stability of material properties to changes in the driving forces, incomplete closure, and acoustic coupling conditions. In this case stability referred to consistency in the results from selected glottal measures including F0, MFDR, MADR, SPL, airflow measures (AC, min, mean), Z_g mean, H1-H2, HRF, total forces (average of F_{e1} and F_{e2}), and total energy transfer (mean and rms). SPL was obtained by subtracting 30 dB (based on empirical observations) from lip output to describe a 10 cm position. The energy transfer measures were computed according to [41, 104].

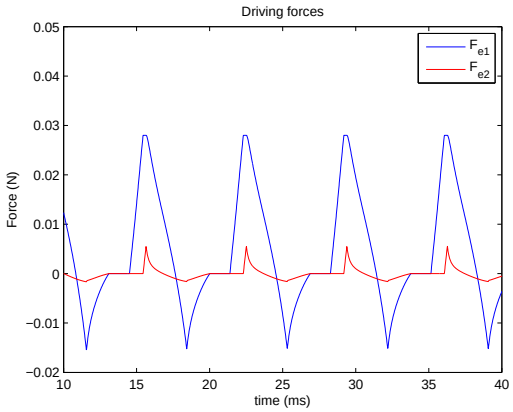
To illustrate the nature of the different driving forces under evaluation, the ST95 model was set using the IF72 parameter set from Table 3.1. The driving forces F_{e1}

and F_{e2} are presented for the scenario with no acoustic coupling (level 1) in Fig. 3.5 and with acoustic coupling (level 2) in Fig. 3.6. The BLEAP cases were shown using dotted lines for the left vocal fold and solid for the right vocal fold. It was noted that: 1) the differences between the forces were more pronounced when no coupling was present, and 2) the acoustically coupled scenario yielded much stronger driving forces. Furthermore, the pressure distributions of the driving forces for convergent and divergent conditions are presented in Fig. 3.7, where the discontinuous nature of the driving forces is evident. Cases with the same pressure distributions are shown together.

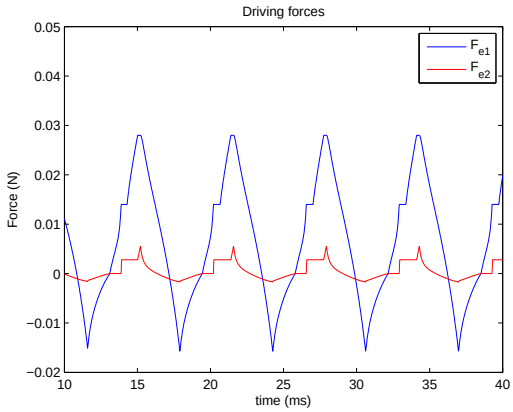
The simulations were classified between parameters sets and shown for "IF72" parameter set in Table 3.4, "Case C" parameter set in Table 3.5, "Rules" parameter set in Table 3.6, and "FEM" parameter set in Table 3.7. General observations indicated that all parameter set exhibited self-sustained oscillations under some conditions. This behavior was not completely expected for Case C, as previous observations had indicated that the parameter set was not able to sustain oscillations with MRI based vocal tracts. The use of a different, less constricted vocal tract geometry made this possible. Nevertheless, Case C showed a very unstable behavior that did not converge in some cases or yielded variable results due to minor variations. Thus, this complete parameter set was subsequently discarded. Although it yielded reasonable behavior under some conditions, the FEM parameter set was also subsequently discarded since it was the least prone to yield self-sustained oscillations and vibrated at a much higher fundamental frequency. On the other hand, IF72 and Rules yielded more stable regimes, yet cases with no acoustic coupling for Rules were less prone to sustain the oscillations (as noted from cases with very low MFDR). Although a no coupling scenario was not physiologically relevant, it related to excised larynx experiments, that have shown that oscillation without the vocal tract was possible [113]. A closer look was taken into the coupling case for the two parameter sets before discarding the Rules set for this reason. Thus, subsequent observations were only limited to Rules and IF72 for the acoustic coupling (level 2) case.



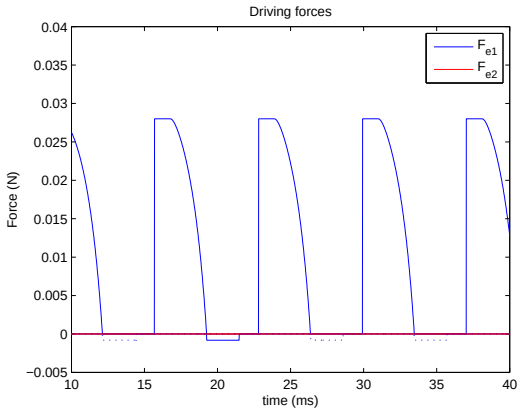
(a) Original



(b) Updated

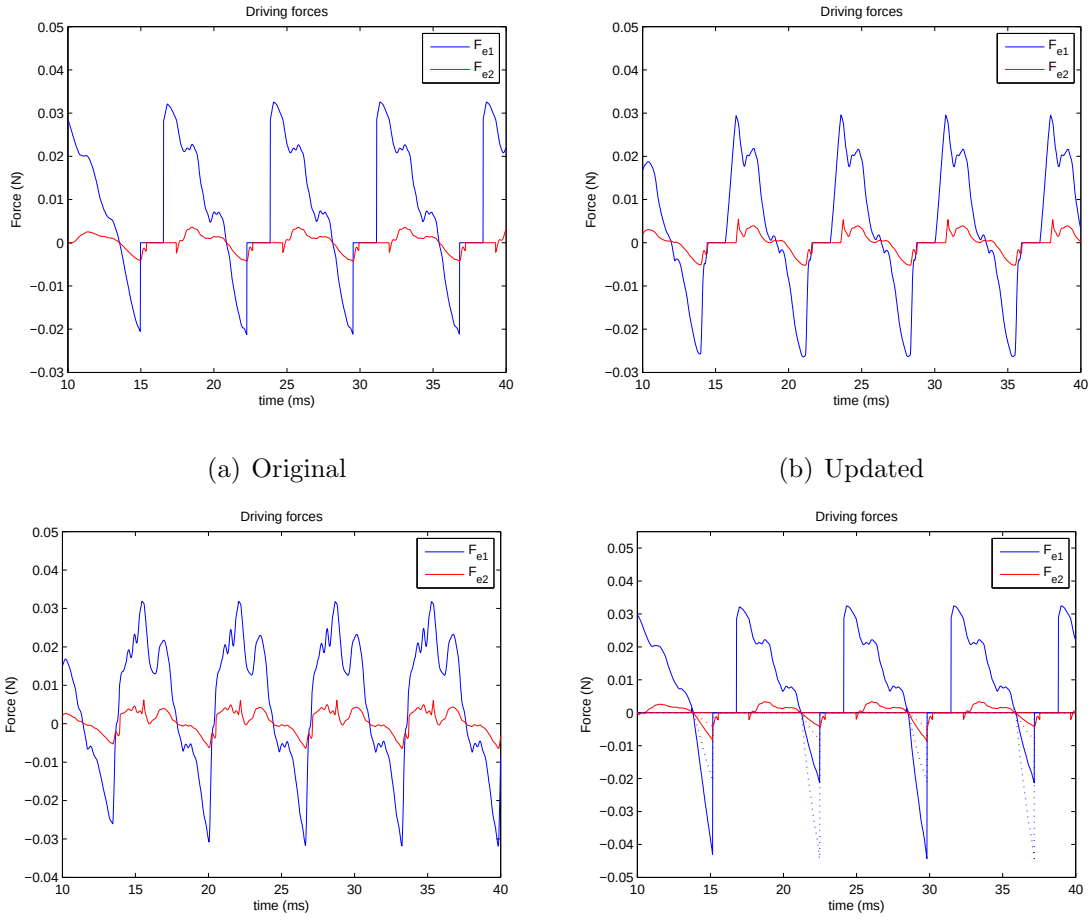


(c) Updated + Hydro



(d) BLEAP

Fig. 3.5. Driving forces for IF72 with no acoustic coupling



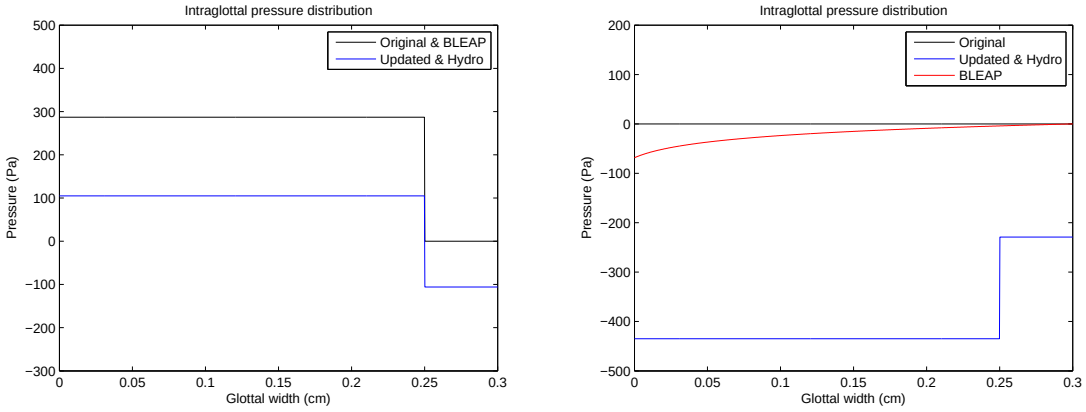
(a) Original

(b) Updated

(c) Updated + Hydro

(d) BLEAP

Fig. 3.6. Driving forces for IF72 with acoustic coupling



(a) Convergent (total angle = 5^o)

(b) Divergent (total angle = -5^o)

Fig. 3.7. Intraglottal pressure distributions for IF72 without acoustic coupling

Table 3.4
Simulations from body-cover model using the parameter set labeled as "IF72" from Table 3.1.

		Acoustic coupling (level 2 of interaction)								No acoustic coupling (level 1 of interaction)							
		Complete closure				Incomplete closure				Complete closure				Incomplete closure			
Measures	Units	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP
F_0	Hz	137	137	154	137	137	137	154	137	137	154	154	137	137	154	154	137
MFDR	cm^3/s^2	467	582	619	520	327	391	419	328	350	385	411	345	296	322	349	289
MADR	cm^2/s	89	105	109	98	79	88	92	76	93	93	100	94	93	93	100	94
SPL	dB	84	85	87	83	79	80	81	78	83	85	86	83	80	81	82	79
AC airflow	l/s	0.44	0.42	0.45	0.42	0.40	0.39	0.41	0.38	0.42	0.39	0.40	0.41	0.40	0.37	0.38	0.38
min airflow	l/s	0	0	0	0	0.09	0.08	0.08	0.09	0	0	0	0	0.08	0.08	0.08	0.08
mean airflow	l/s	0.16	0.16	0.18	0.16	0.26	0.25	0.27	0.26	0.16	0.14	0.15	0.15	0.25	0.23	0.24	0.24
Zg mean	ohm (cgs)	49.2	49.6	44.8	50.0	31.2	32.5	29.8	31.2	50.9	57.1	54.7	52.4	32.3	34.3	33.2	33.2
H1-H2	dB	-11.6	-11.7	-11.1	-12.1	-13.9	-12.9	-14.8	-15.7	-12.3	-7.5	-11.2	-12.8	-13.4	-9.1	-13.9	-14.1
HRF	dB	-10.9	-10.9	-10.2	-11.2	-13.5	-12.2	-13.9	-14.9	-11.6	-7.1	-10.1	-12.0	-13.1	-8.9	-13.0	-13.6
F_T opening	$N \cdot 10^3$	22.6	16.7	19.6	21.7	22.4	14.6	16.3	22.2	22.4	14.6	16.3	22.2	21.4	15.7	17.5	20.6
F_T closing	$N \cdot 10^3$	-4.8	-12.6	-12.9	-7.7	6.2	-6.3	-6.4	1.1	6.2	-6.3	-6.4	1.0	-2.4	-8.8	-9.0	-3.1
F_T collision	$N \cdot 10^3$	14.1	18.3	19.4	16.0	8.0	9.7	10.4	8.4	8.0	9.7	10.4	8.4	11.3	14.4	15.1	12.3
Π open mean	J/s	0.24	0.29	0.30	0.22	0.18	0.22	0.23	0.16	0.27	0.17	0.18	0.27	0.27	0.17	0.18	0.27
Π open rms	J/s	0.35	0.39	0.40	0.33	0.28	0.28	0.29	0.26	0.37	0.24	0.26	0.37	0.37	0.24	0.26	0.37

Table 3.5
Simulations from body-cover model using the parameter set labeled as "Case C" from Table 3.1.

		Acoustic coupling (level 2 of interaction)								No acoustic coupling (level 1 of interaction)							
		Complete closure				Incomplete closure				Complete closure				Incomplete closure			
Measures	Units	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP
F_0	Hz	137	171	-	120	137	154	171	137	120	171	188	137	120	171	188	137
MFDR	cm^3/s^2	1157	1826	-	955	772	1276	1735	696	488	1216	1449	507	407	964	1141	415
MADR	cm^2/s	195	276	-	188	168	245	319	156	140	229	256	142	140	229	256	142
SPL	dB	87	92	-	86	82	87	88	82	85	91	91	84	81	87	86	80
AC airflow	l/s	0.58	0.64	-	0.56	0.58	0.63	0.68	0.49	0.49	0.54	0.54	0.47	0.47	0.54	0.55	0.44
min airflow	l/s	0	0	-	0	0.07	0.04	0.02	0.07	0	0	0	0	0.08	0.05	0.04	0.08
mean airflow	l/s	0.29	0.28	-	0.28	0.37	0.35	0.39	0.34	0.17	0.18	0.22	0.17	0.26	0.28	0.31	0.25
Zg mean	ohm (cgs)	27.8	29.1	-	29.1	21.7	23.1	20.7	23.3	47.6	45.3	35.6	48.4	31.3	28.8	26.1	32.0
H1-H2	dB	-11.7	-10.0	-	-14.7	-14.1	-14.6	-15.0	-16.0	-7.9	-5.7	-12.6	-9.2	-8.9	-9.1	-16.0	-8.9
HRF	dB	-10.2	-7.7	-	-12.4	-12.9	-12.2	-11.2	-14.0	-6.9	-4.0	-8.1	-8.0	-8.4	-7.6	-11.3	-8.3
F_T opening	$N \cdot 10^3$	7.9	7.0	-	8.2	8.7	7.5	9.0	14.1	10.9	7.1	7.4	10.7	10.9	7.1	7.4	10.6
F_T closing	$N \cdot 10^3$	-2.8	-13.4	-	-1.0	-1.5	-9.5	-8.4	4.8	7.0	-5.7	-5.3	3.5	7.0	-5.7	-5.3	3.5
F_T collision	$N \cdot 10^3$	13.8	22.8	-	10.9	10.5	18.6	22.1	5.6	2.5	9.0	7.0	2.8	2.5	9.0	7.0	2.8
Π open mean	J/s	0.38	0.71	-	0.31	0.34	0.60	0.94	0.13	0.25	0.45	0.53	0.24	0.25	0.45	0.53	0.24
Π open rms	J/s	0.71	1.24	-	0.56	0.54	0.94	1.50	0.28	0.62	0.66	0.77	0.61	0.62	0.66	0.77	0.61

Table 3.6
Simulations from body-cover model using the parameter set labeled as "Rules" from Table 3.1

		Acoustic coupling (level 2 of interaction)								No acoustic coupling (level 1 of interaction)							
		Complete closure				Incomplete closure				Complete closure				Incomplete closure			
Measures	Units	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP
F_0	Hz	137	137	154	137	137	137	154	120	137	120	137	137	137	120	137	137
MFDR	cm^3/s^2	156	491	455	257	163	330	340	177	27	310	303	26	26	252	247	25
MADR	cm^2/s	60	100	98	84	73	86	90	67	4	68	67	4	4	68	67	4
SPL	dB	74	82	82	76	72	76	76	71	51	78	77	50	50	75	73	50
AC airflow	l/s	0.19	0.27	0.27	0.21	0.18	0.24	0.25	0.18	0.06	0.21	0.21	0.06	0.06	0.20	0.20	0.06
min airflow	l/s	0	0	0	0	0.09	0.09	0.09	0.09	0.05	0	0	0.05	0.16	0.09	0.09	0.15
mean airflow	l/s	0.09	0.12	0.13	0.10	0.20	0.21	0.22	0.19	0.07	0.10	0.12	0.07	0.17	0.20	0.21	0.17
Zg mean	*ohm* (cgs)	87.8	64.4	62.5	76.3	41.0	37.5	36.2	42.8	107.7	77.9	68.5	108.2	46.2	40.2	37.8	46.8
H1-H2	dB	-23.7	-19.0	-16.3	-21.2	-23.7	-20.0	-16.6	-18.7	-35.3	-19.1	-13.5	-35.1	-35.3	-18.6	-13.6	-35.3
HRF	dB	-21.4	-14.1	-13.7	-17.8	-21.3	-16.4	-14.6	-17.0	-35.3	-12.7	-11.2	-35.1	-35.3	-14.5	-12.2	-35.3
F_T opening	$N \cdot 10^3$	16.6	13.6	12.9	15.8	15.5	13.2	12.8	15.2	15.9	14.3	13.4	15.9	15.9	14.3	13.4	15.9
F_T closing	$N \cdot 10^3$	10.6	1.1	1.5	13.7	9.8	2.8	2.8	13.2	15.9	6.5	7.2	24.0	15.9	6.5	7.2	24.0
F_T collision	$N \cdot 10^3$	15.8	28.0	21.6	21.4	15.7	23.6	19.0	15.1	0.0	16.7	11.1	0.0	0.0	16.7	11.1	0.0
Π open mean	J/s	0.02	0.09	0.08	0.03	0.02	0.06	0.07	0.02	0.00	0.04	0.04	0.00	0.00	0.04	0.04	0.00
Π open rms	J/s	0.12	0.17	0.15	0.15	0.12	0.13	0.13	0.12	0.05	0.10	0.09	0.05	0.05	0.10	0.09	0.05

Table 3.7
Simulations from body-cover model using the parameter set labeled as "FEM" from Table 3.1

		Acoustic coupling (level 2 of interaction)								No acoustic coupling (level 1 of interaction)							
		Complete closure				Incomplete closure				Complete closure				Incomplete closure			
Measures	Units	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP	Original	Updated	Hydro	BLEAP
F_0	Hz	171	171	188	171	171	171	171	171	171	154	171	171	171	154	171	171
MFDR	cm^3/s^2	3	522	560	3	3	345	383	2	2	424	438	2	2	339	350	2
MADR	cm^2/s	0	112	118	0	0	97	104	0	0	95	96	0	0	95	96	0
SPL	dB	21	82	82	21	20	77	76	20	9	82	81	9	9	78	76	9
AC airflow	l/s	0.01	0.18	0.18	0.01	0.01	0.17	0.17	0.01	0	0.18	0.17	0.01	0.01	0.18	0.17	0.01
min airflow	l/s	0.03	0	0.03	0	0.14	0.09	0.14	0.09	0.03	0	0.03	0	0.13	0.09	0.10	0.09
mean airflow	l/s	0.03	0.07	0.07	0.03	0.14	0.16	0.17	0.14	0.03	0.07	0.07	0.03	0.13	0.17	0.17	0.10
Zg mean	ohm (cgs)	235.0	123.0	108.3	228.6	57.8	48.5	46.8	56.8	241.9	116.3	114.3	242.4	59.3	47.8	47.6	79.2
H1-H2	dB	-52.9	-10.1	-9.8	-53.3	-56.9	-10.0	-12.7	-58.5	-56.4	-9.5	-10.9	-54.4	-57.5	-11.6	-13.1	-52.9
HRF	dB	-52.9	-8.0	-6.8	-53.3	-56.7	-8.2	-9.6	-58.4	-56.3	-7.4	-8.3	-54.3	-54.4	-9.5	-10.4	-52.7
F_T opening	$N \cdot 10^3$	22.9	31.1	31.3	23.0	22.1	29.7	30.0	22.4	23.1	32.2	31.7	23.1	23.1	32.2	31.7	23.1
F_T closing	$N \cdot 10^3$	23.8	5.9	8.0	35.7	23.0	7.5	9.3	34.0	24.2	12.5	13.9	36.3	24.2	12.5	13.9	36.3
F_T collision	$N \cdot 10^3$	0	23.8	19.5	0	0	20.7	16.8	0	0	17.5	14.1	0	0	17.5	14.1	0
Π open mean	J/s	0	0.17	0.23	0	0	0.14	0.18	0	0	0.15	0.18	0	0	0.15	0.18	0
Π open rms	J/s	0.06	0.38	0.48	0.06	0.06	0.34	0.40	0.06	0.06	0.41	0.42	0.06	0.06	0.41	0.42	0.06

It was also observed that the presence of incomplete glottal closure reduced the energy transfer from the fluid to the vocal folds (as noted from the smaller difference in F_T between opening and closing and Π mean and rms values) for all cases where coupling was present, indicating again that lower acoustic pressures were present for that case (also noted in the lower SPL values).

Cases with high MFDR and SPL were correlated with those exhibiting larger differences between F_T opening and F_T closing and higher Π mean and rms values. The presence of negative forces for F_T closing (sometimes referred to as the Bernoulli effect) was consistent for the complete closure cases and case-dependent during incomplete closure. The energy transfer was lower than that reported using an M5 geometry [104] due to the smaller integration region used for the ST95 model that only accounted for the glottal width. The mean energy transferred due to incomplete glottal closure was reduced by about 25% for IF72 and between 25% and 5% for Rules. In addition, high correlations between MFDR and collision forces were observed for these cases, matching the behavior noted by [25], where MFDR was correlated with tissue velocity and thus related to vocal hyperfunction and trauma. However, IF72 exhibited a much higher correlation than its Rules counterpart, as shown in Fig 3.8.

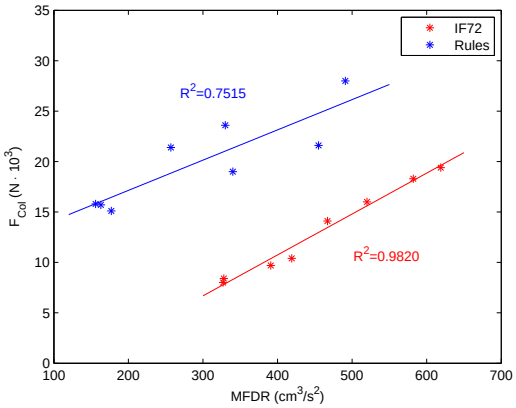


Fig. 3.8. Correlation between MFDR and collision forces for IF72 and Rules with acoustic coupling with complete and incomplete glottal closure

When contrasting incomplete closure data with that from human subject recordings from Table 3.3 certain parameters (AC flow, min flow, mean flow) were observed to be in the range of the mean values, and others (MFDR, SPL) slightly lower. These latter measures were better aligned with the results from the complete closure scenarios. Nevertheless, both IF72 and Rules parameters sets were within one standard deviation from the means in Table 3.3 and thus considered representative of human physiology. However, the IF72 set was preferred since it yielded measures closer to human average data. Within this case, the forces that better resembled the human data were given by the updated equations from [74]. The original driving forces from ST95 yielded reasonable behavior but with measures lower than those from the human recordings. The driving force that included hydrostatic pressure yielded much higher frequencies of oscillation and rather unstable results. The BLEAP forces were reduced due to the presence of incomplete closure and yielded lower measures. Therefore, all subsequent analyses were performed using IF72 parameter set with acoustic coupling and updated equations.

3.2.3 Self-sustained model: effect of vowels

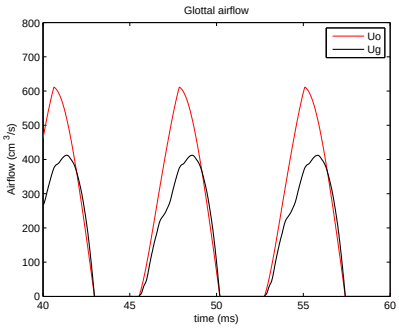
The effects of the acoustic coupling for complete and incomplete glottal closure were explored for four distinct vowels. The vowels were selected in pairs that had similar first format frequencies but larger variations in the second formant frequency (e.g., /a/-/e/ and /i/-/u/). As with the driven glottal area case, comparisons between the interactive airflow (U_g) and the ideal uncoupled airflow source (U_o) were made for selected glottal measures to evaluate the degree of harmonic distortion and coupling. All cases were examined for complete and incomplete glottal ($A_{gap} = 0.03\ cm^2$) conditions. Estimates of SPL were not included in observations since they could be highly affected by the vocal tract rather than the glottal source.

It was observed in Fig. 3.9 and Fig. 3.10 that noticeable differences were present between each vowel, being the first formant the most dominant parameter controlling

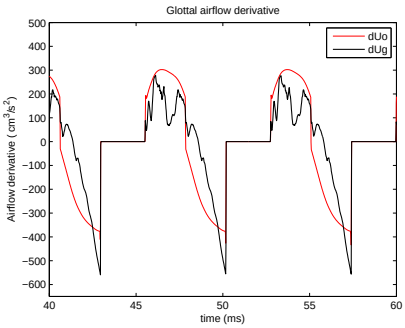
these differences. However, a closer inspection of the glottal measures from Table 3.8 indicated importance differences that could be attributed to the second formant. The degree of coupling was primarily determined by the lower Z_g mean and confirmed by the reduced AC airflow and was more pronounced for vowel /i/. As in the driven area observations, stronger coupling did not necessarily imply stronger pulse skewing since vowel /i/ exhibited the lowest skewing of all vowels (as evidenced by the lowest MFDR). Vowels /a/ and /e/ had very similar responses, showing that minor formant changes for low vowels do not affect significantly the coupling or harmonic distortion in modal speech. However, vowel /u/ had notable differences with vowel /i/ that were observed in its higher MFDR and lower H1-H2. These differences were primarily attributed to frequency differences of their second formants that changed the complete tract impedance curves. The same trends were observed for incomplete glottal closure. The ripples for the closed phased portion of the cycle were noted to be vowel dependent and matched previous observations [14, 15] for vowels /a/ and /e/. Higher vowels (not evaluated in previous studies) exhibited a single long decaying ripple during the closed phase, most likely due to the longer wavelength effect of the tract for these cases. It also appeared that the second formant affected the shape of this ripple, as vowel /u/ had a more pronounced slope in it.

3.2.4 Self-sustained model: types of incomplete closure

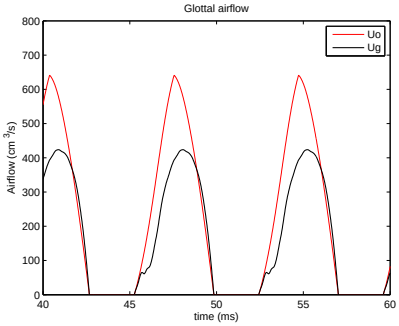
All observations up to this point have been made using a parallel gap description to account for incomplete glottal closure. In this section, this representation was contrasted with two additional schemes: abduction [102] and pre-collision [85]. A vowel /e/ was selected for the vocal tract in these cases. All other configurations were unchanged from before. SPL was included in the measures since all cases used the same vowel and thus SPL variations could be attributed to source differences. The abduction case used a nonlinear constant of 50, which differed from that of [102], given the different underlying modeling assumptions. As in their case, this constant



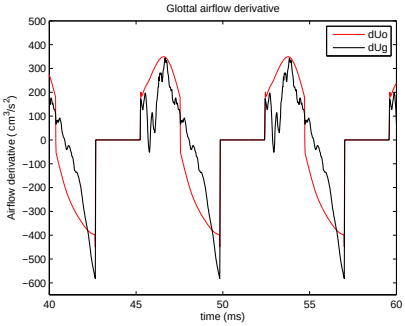
(a) vowel /a/ - airflow



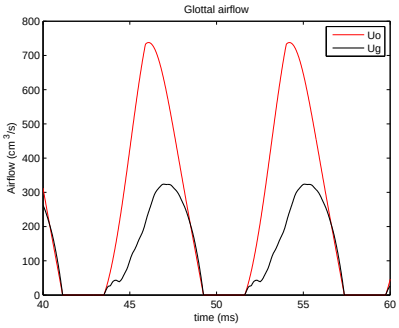
(b) vowel /a/ - airflow derivative



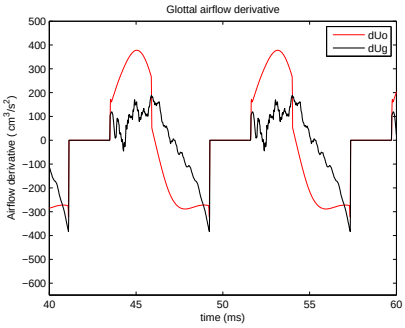
(c) vowel /e/ - airflow



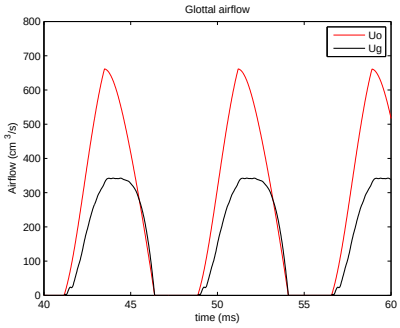
(d) vowel /e/ - airflow derivative



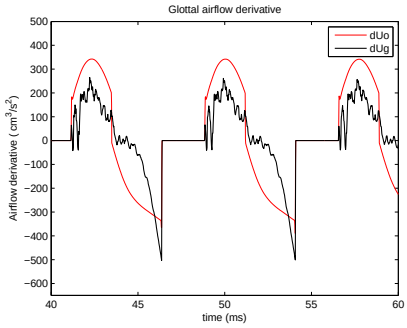
(e) vowel /i/ - airflow



(f) vowel /i/ - airflow derivative

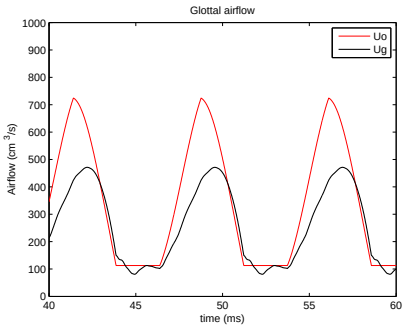


(g) vowel /u/ - airflow

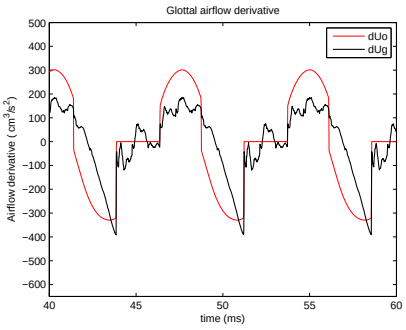


(h) vowel /u/ - airflow derivative

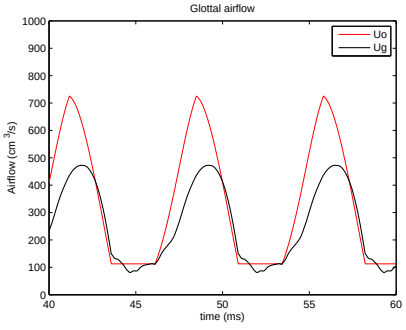
Fig. 3.9. Effect of coupling in vowels for complete closure



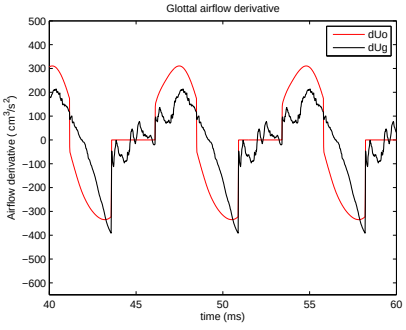
(a) vowel /a/ - airflow



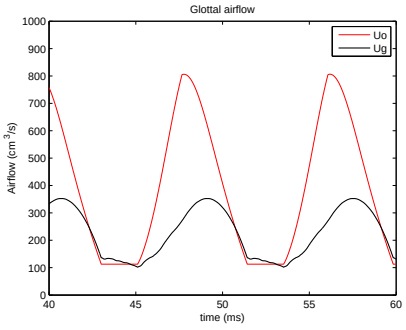
(b) vowel /a/ - airflow derivative



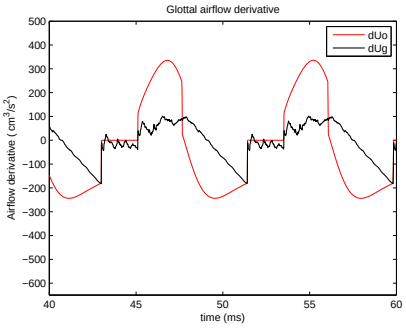
(c) vowel /e/ - airflow



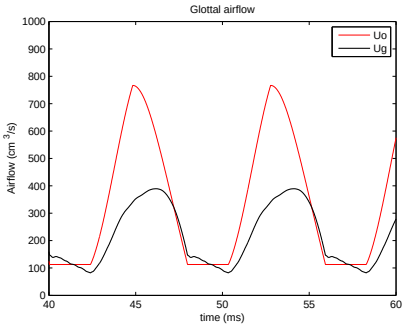
(d) vowel /e/ - airflow derivative



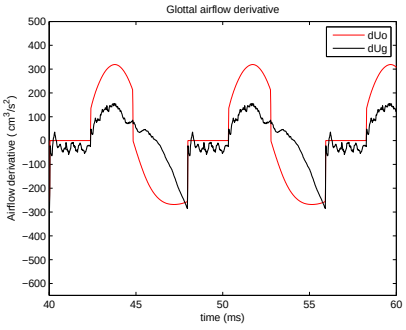
(e) vowel /i/ - airflow



(f) vowel /i/ - airflow derivative



(g) vowel /u/ - airflow



(h) vowel /u/ - airflow derivative

Fig. 3.10. Effect of coupling in vowels for incomplete closure

Table 3.8
Summary of selected glottal measures for different vowels during complete and incomplete glottal closure for ideally uncoupled airflow (1) and interactive airflow (2).

		Complete closure				Incomplete closure			
Measure	Units	/a/	/e/	/i/	/u/	/a/	/e/	/i/	/u/
Zg mean	Ω (cgs)	29.7	28.4	14.7	20.4	20.8	20.7	13.0	15.5
AC flow 1	cm^3/s	611	641	738	661	611	612	693	654
AC flow 2	cm^3/s	412	424	324	343	390	391	251	307
MFDR 1	cm^3/s^2	426	449	322	374	330	335	244	268
MFDR 2	cm^3/s^2	557	582	383	495	391	391	183	285
H1-H2 1	dB	-11.9	-11.4	-14.1	-12.7	-11.9	-11.5	-14.3	-13.6
H1-H2 2	dB	-13.6	-11.4	-12.4	-17.3	-13.6	-12.8	-16.8	-23.1
HRF 1	dB	-11.7	-11.3	-13.7	-12.3	-11.7	-11.3	-13.8	-13.2
HRF 2	dB	-11.5	-10.5	-11.6	-12.5	-12.6	-12.2	-15.3	-15.0

was arbitrarily set to avoid large increases in amplitude due to a change in the rest position of the masses in the lumped model. Two cases were explored for each type of incomplete closure: a normal gap ($A_{gap} = 0.03\ cm^2$) and large gap ($A_{gap} = 0.05\ cm^2$).

It was observed in Fig. 3.11 (normal gap), Fig. 3.12 (large gap), and Table 3.9 (summary of measures) that the parallel gap exhibited a DC offset and maintained the structure of the closed phase ripples for both conditions. Although all cases exhibited energy reduction with respect to the complete closure, the energy reduction between gap conditions was more subtle in the parallel chink case. Pre-collision did not exhibit incomplete closure with the smaller gap but reduced the MFDR. The large case more severely affected the MFDR and amplitude of oscillation for that case. Abduction did show DC offset in both cases, but severely reduced the amplitude of oscillation.

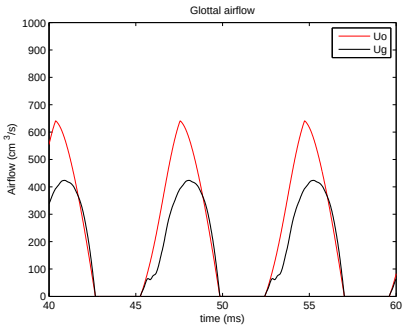
It thus appears that these cases represented different physiological phenomena associated with incomplete closure; where the parallel gap better resembled the effects of a posterior gap or chink, the pre-collision the effects of nodules or any organic

pathological condition that can affect collision between folds, and abduction a breathy voice with abnormal laryngeal muscle control or paralysis.

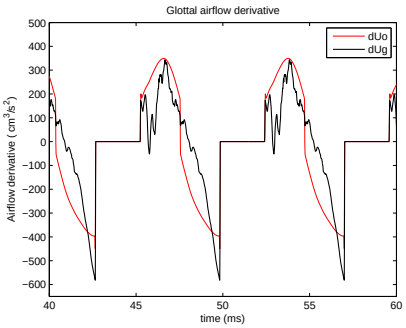
3.2.5 Self-sustained model: modeling hyperfunction

Based on the results of the previous section it was evident that the model could be used to mimic a number of pathological conditions, although further research will be needed to determine the accuracy of each case accordingly. In this section, the ability of the model to represent vocal hyperfunction was assessed using human subject recordings from previous studies [25]. Vocal hyperfunction has been related to the origin of most organic pathologies and it is associated with a given inadequate condition that is incorrectly compensated, thus increasing the risk of tissue damage. In this section, the inadequate condition was modeled as a large parallel gap ($A_{gap} = 0.05\ cm^2$) and the compensation was an increased effort translated as a higher subglottal (lung) pressure. The normal subglottal pressure was selected to be 0.8 kPa for the complete glottal closure and increased to match the SPL of such case due to the presence of the posterior gap. The results of the complete, small gap, large gap, and compensated large gap were examined.

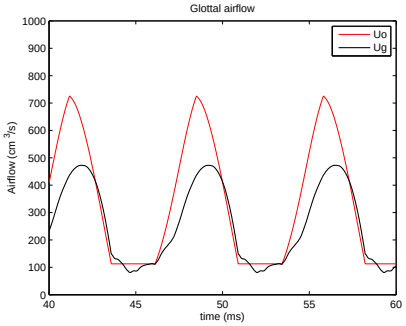
It was found that the compensated effort yielded the same SPL level, but significantly increased the levels of MFDR and AC airflow, as noted in bold font in Table 3.10. These results followed previous observations from pathologic cases in human subject recordings that indicated that the same three measures significantly increased due to the presence of vocal hyperfucntion [25]. It was noted before that an increased MFDR was highly correlated with increased collision forces (see Fig. 3.8), which indicated potential risk of generating an organic pathology. Thus, it appeared that incomplete glottal closure, as associated to a large parallel posterior gap, may be one possible explanation of this type of behavior. Other scenarios with incomplete glottal closure as those studied in the previous section or a combination of them could also lead to this behavior and should be the the topic of further research.



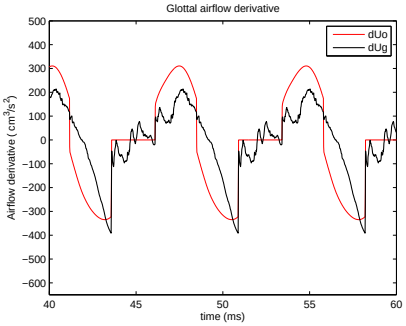
(a) Complete- airflow



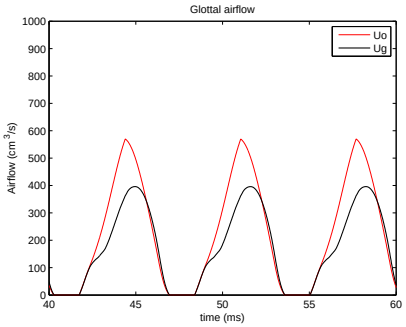
(b) Complete - airflow derivative



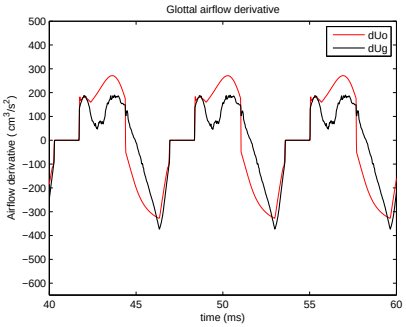
(c) Parallel gap - airflow



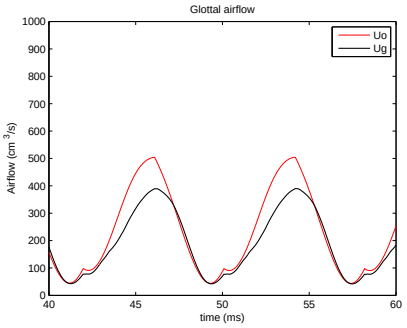
(d) Parallel gap- airflow derivative



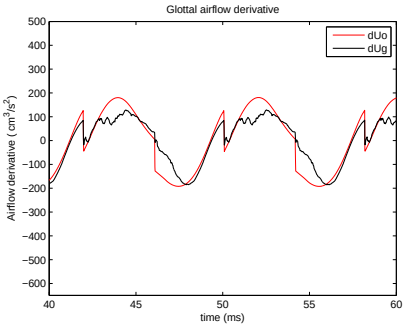
(e) Pre-collision - airflow



(f) Pre-collision - airflow derivative

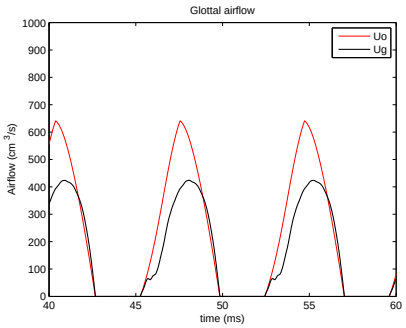


(g) Abduction - airflow

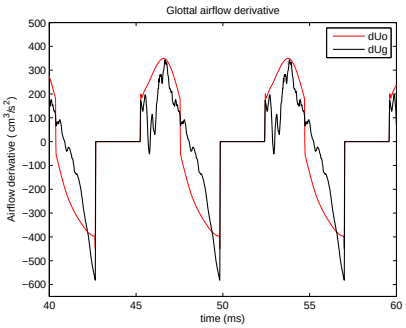


(h) Abduction- airflow derivative

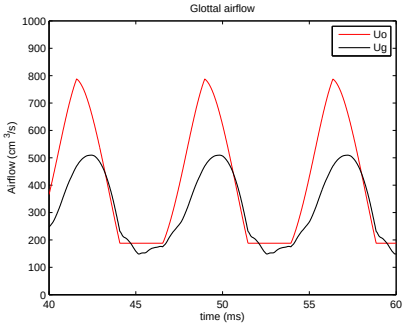
Fig. 3.11. Types of incomplete glottal closure for A_{gap}=0.03 cm^2



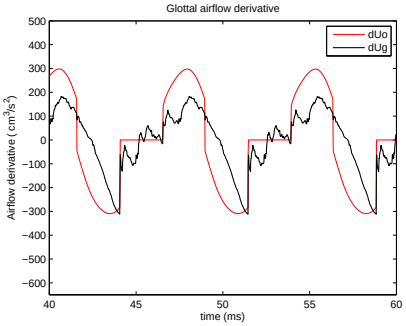
(a) Complete- airflow



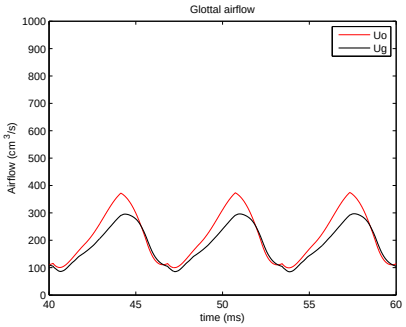
(b) Complete - airflow derivative



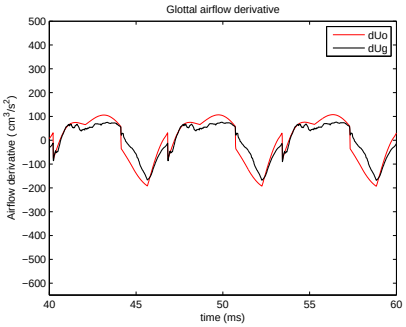
(c) Parallel gap - airflow



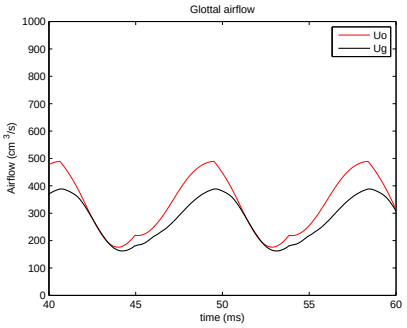
(d) Parallel gap- airflow derivative



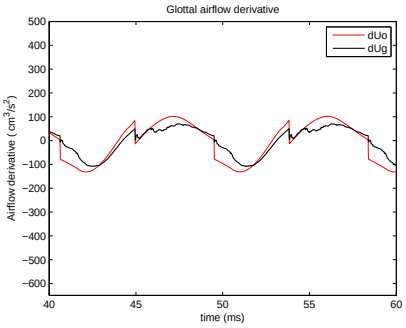
(e) Pre-collision - airflow



(f) Pre-collision - airflow derivative



(g) Abduction - airflow



(h) Abduction- airflow derivative

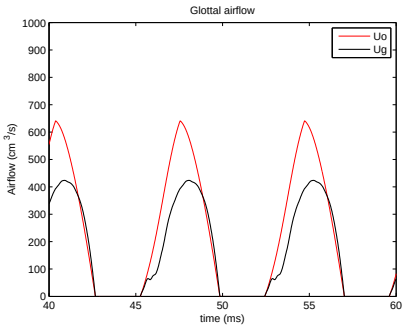
Fig. 3.12. Types of incomplete glottal closure for A_{gap}=0.05 cm^2

Table 3.9
Summary of selected glottal measures for different types of incomplete glottal closure for ideal uncoupled airflow (1) and interactive airflow (2)

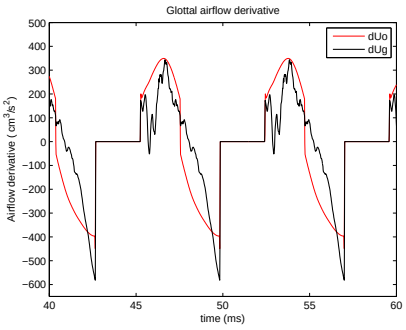
		Complete	Parallel gap		Pre-contact		Abduction	
Measure	Units	0 cm^2	0.03 cm^2	0.05 cm^2	0.03 cm^2	0.05 cm^2	0.03 cm^2	0.05 cm^2
Zg mean	*ohm* (cgs)	28.4	20.7	17.0	29.2	34.5	28.8	23.0
AC flow 1	cm^3/s	641	612	599	570	276	461	313
AC flow 2	cm^3/s	424	391	361	396	213	348	227
MFDR 1	cm^3/s^2	449	335	309	328	194	192	132
MFDR 2	cm^3/s^2	582	391	312	373	169	186	109
H1-H2 1	*dB*	-11.4	-11.5	-12.0	-15.7	-14.4	-18.5	-15.5
H1-H2 2	*dB*	-11.4	-12.8	-14.2	-13.2	-14.2	-16.6	-14.7
HRF 1	*dB*	-11.3	-11.3	-11.8	-15.2	-14.0	-17.8	-15.1
HRF 2	*dB*	-10.5	-12.2	-13.8	-12.5	-13.9	-16.1	-14.5
SPL	*dB*	85.6	80.2	77.9	82.5	73.6	74.7	69.4

Table 3.10
Summary of selected glottal measures for different conditions related to vocal hyperfunction

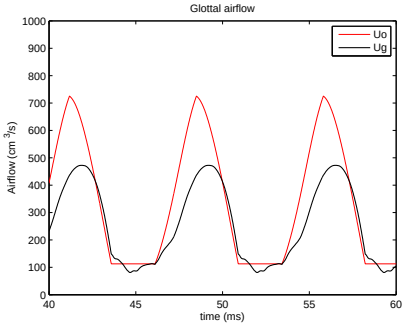
		No gap	Normal gap	Large gap	Hyperfunction
Measure	Units	Complete	0.03 cm^2	0.05 cm^2	0.05 cm^2
SPL	*dB*	85.6	80.2	77.9	85.5
Ps	kPa	0.8	0.8	0.8	**1.32**
Zg mean	*ohm* (cgs)	27.3	20.5	17.0	19.8
AC flow	cm^3/s	424	391	361	**613**
MFDR	cm^3/s^2	581	391	312	**762**
H1-H2	*dB*	-11.4	-12.8	-14.3	-11.1
HRF	*dB*	-10.5	-12.2	-13.8	-10.6



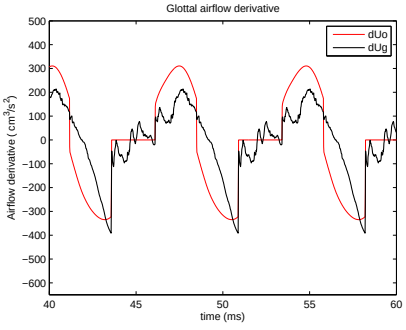
(a) Complete- airflow



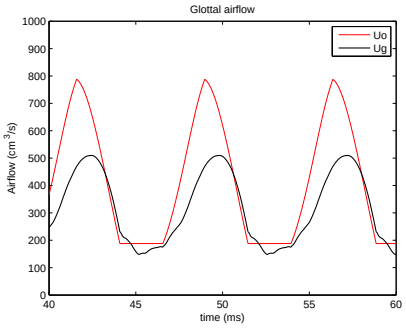
(b) Complete - airflow derivative



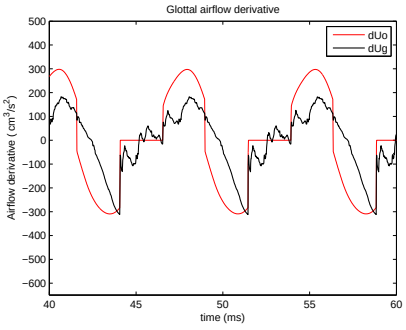
(c) Normal gap - airflow



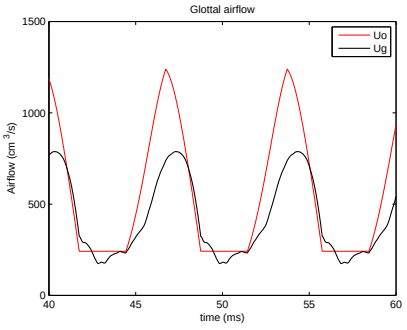
(d) Normal gap- airflow derivative



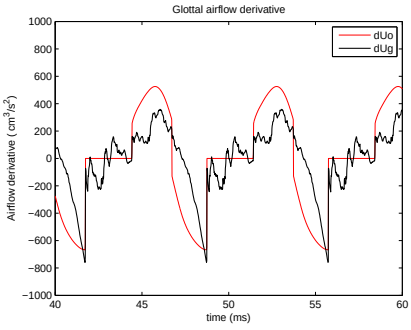
(e) Large gap - airflow



(f) Large gap - airflow derivative



(g) Hyperfunction - airflow



(h) Hyperfunction- airflow derivative

Fig. 3.13. Effect of incomplete glottal closure and its relation with vocal hyperfunction

3.3 Discussion

The initial observations using a driven glottal area provided insights into the factors affecting acoustic coupling. The interaction initially followed the principles described in [17], where the combined tract impedances determined the degree of coupling and harmonic distortion (skewing and ripples in the glottal pulses) with respect to the uncoupled airflow. Although the tract impedances explained the frequency and vowel dependence of the coupling phenomena, the degree of abduction was also shown to affect both components. Complete closure had a larger associated glottal impedance, which indicated less coupling, yet large harmonic distortion. Such a distortion was observed to be a product of the large acoustic pressures generated under this condition. On the other hand, incomplete glottal closure exhibited a larger coupling but smaller harmonic distortion. A strongly coupled condition was shown in previous studies to be proportional to the transglottal pressure term [59]. However, incomplete closure generated smooth acoustic pressures with smaller amplitudes, that reduced the harmonic distortion. These observations were well-aligned with nonlinear coupling theory [17] and linearizations of the sound production system that are discussed in chapter 6, where a small glottal impedance could generate large acoustic coupling.

A parametric analysis identified that the body-cover model (and presumably similar low order lumped element representations) was very sensitive to the parameter set values describing material properties, driving forces, and acoustic coupling. The latter was considered critical as the acoustic models normally used along with lumped element models may not accurately represent the near field conditions at the glottis. Further experimental studies are needed to explore the associated near field acoustic phenomena and validate the representations and configurations used in these numerical models. Nevertheless, the extension of the body-cover model proposed in this chapter was shown to mimic normal male human data in most cases [189]. In particular, the IF72 parameter set exhibited the most stable behavior for various conditions

and driving forces and better matched mean values of selected measures of vocal function from human data [189] using the updated equations from [74] but with no hydrostatic pressure. This behavior may suggest that there is not a compelling need to use the third degree of freedom from the body-cover model. However, the IF72 parameter set was only valid for male modal speech and does not allow for muscle control (i.e., it can only be used to study one glottal configuration). Further research is needed to evaluate and contrast with human data recordings other parameter sets from the the rules system [97] representing diverse glottal conditions.

Simulations using the self-sustained model for different vowels with and without complete closure indicated source-related differences between cases that were primarily given by the vocal tract interaction. For low vowels (/a/, /e/), MFDR was well-correlated with the degree of skewing, harmonic distortion, and collision forces. These vowels exhibited comparable results, whereas high vowels (/i/ and /u/) differed significantly in the degree of skewing (MFDR) and harmonic composition (H1-H2). Differences between these vowels illustrated the importance of considering all the impedance interaction and not just the first format to predict the coupled behavior, i.e., the second formant does alter the source-filter interaction. In all cases, a reduced AC airflow was indicative of the degree of coupling. The same trends were maintained during incomplete glottal closure. In addition, the presence of ripples during the closed portion of the cycle during incomplete glottal closure was consistent with previous observations using a parametric glottal area excitation [14, 15]. These findings suggested that studies focused on the assessment of vocal function need to carefully select the desired vowels, as they can vary the results. Further research is needed to adapt the assessment of vocal function for running speech. In chapter 6, these theoretical observation are contrasted with inverse filtered data on human subjects.

Different types of incomplete glottal closure were evaluated using the self-sustained model. All cases were related to different physiological phenomena and yielded different levels of interaction. It was found that the parallel gap case proposed in this

chapter better represented the normal amount of DC offset in the airflow present in normal speakers, matching the behavior reported in previous studies [39, 189]. In addition, it was found that when the reduction of energy transfer introduced by a large parallel gap was compensated with an increased subglottal pressure, the model mimicked vocal hyperfunction, where MFDR and AC flow became much larger than normal case. This finding was consistent with previous observations [25] for the same parameters. Further efforts considered including the effects of turbulent noise in the interaction.

3.4 Conclusions

The simulations presented in this chapter confirmed the impedance-based nature of the interaction between the glottal airflow and the sub- and supra-glottal tracts described in nonlinear theory of voice production [17]. Analysis of incomplete closure cases elucidated that the degree of coupling may not always be associated with degree of skewing and harmonic distortion in the glottal airflow, as observed during incomplete closure. Lower AC airflow and mean impedance were the main indicators of the degree of coupling and MFDR of the degree of skewing and harmonic distortion. During incomplete glottal closure airflow became proportional to the transglottal pressure drop, which was observed to be smooth. High vowels (particularly /i/) had stronger coupling but less skewing than lower vowels. The model was sensitive to force and parameter set value changes, but a selected parameter set and driving forces closely matched human data for simple observations on normal, hyperfunctional, and pathological cases. The addition of a parallel gap in the body-cover model made it more more realistic and allowed for extending observations from nonlinear source-filter coupling.

4. EFFECTS OF ACOUSTIC COUPLING ON TISSUE INSTABILITIES

The goal of the research presented in this chapter was to explore the effects of acoustic coupling on tissue dynamics *in vivo* with human subjects. Based on nonlinear source-filter coupling theory [17], it was expected that the relation between fundamental frequency of vocal fold vibration and sub- and supra-glottal tracts impedances could visibly affect tissue dynamics by either favoring or hampering the oscillation. Verification of this hypothesis validated in part the nonlinear coupling theory and supported the description of source-filter interactions based on lumped representations of the system. This validation was considered critical to apply acoustic coupling and impedance concepts in inverse filtering of speech signals from chapter 6.

Numerical simulations have shown that source-filter interactions can lead to "voice breaks", i.e., unstable tissue motion and sudden frequency jumps referred to as bifurcations [17, 128]. Supporting evidence that this phenomenon takes place in actual human speech was based on the relation between the frequency where these voice breaks occurred and the formant frequencies of the vocal tract and subglottal system [49, 128]. This evidence led to the hypothesis in this chapter that strong acoustic coupling would introduce additional driving forces that would visibly affect the tissue motion, phenomenon that could be observed *in vivo*. Bifurcations where less significant acoustic-interaction is present were expected to exhibit visible differences in the unstable tissue motion with respect the previous case. It was expected that these differences would be more evident by selecting vocal exercises that maximize and minimize the source-filter coupling. Furthermore, it was hypothesized that adding acoustic interaction in a region where bifurcations can occur naturally (as in register transitions regions) would facilitate their occurrence.

4.1 Methods

The vocal exercises and classification proposed by Titze [49] were followed so that "acoustically-induced" bifurcations could be distinguished and contrasted with "source-induced" ones using pitch glide gestures. It is emphasized that even though it is likely that a combination of components leading to bifurcations contributed in all cases, it was expected that "source" or "acoustic" related factors would be more dominant in certain cases. The adequacy of this classification is discussed at the end of the chapter based on the results obtained.

It was desired to obtain estimates of the complete system behavior for each dynamic vocal task exhibiting bifurcations. Thus, simultaneous recordings describing glottal behavior, flow aerodynamics, and acoustic pressures were obtained. A particular emphasis was put on documenting and analyzing the unstable motion of the vocal folds by means of digital high-speed video and image processing.

4.1.1 Experimental setup

Three types of experimental configurations were used for different purposes. The most complete configuration allowed for simultaneous measurements of laryngeal high-speed videoendoscopy (HSV), radiated acoustic pressure (MIC), neck skin acceleration (ACC), electroglottography (EGG), and oral volume velocity (OVV). This configuration captured HSV with a flexible endoscope, which not only allowed for aerodynamic assessment but also a normal articulation for the participant. A representation of this configuration is illustrated in Fig. 4.1. A similar configuration used HSV with a rigid endoscope, which provided higher image quality and spatio-temporal resolution, but did not allow for aerodynamic assessment and limited the degree of articulation for the participant. Synchronous measurements of ACC, EGG, and MIC were used in this configuration. A much simpler configuration did not include HSV and was intended for screening purposes. Thus, only measurements of ACC, EGG, and MIC were considered in this setup. All recordings were obtained in

an acoustically-treated room at the Center for Laryngeal Surgery & Voice Rehabilitation at the Massachusetts General Hospital.

HSV recordings were acquired using a Phantom v7.3 high-speed video color camera (Vision Research Inc.) and a Phantom v7.1 high-speed video monochromatic camera (Vision Research Inc.). A C-mount lens adapter with adjustable focal length (KayPENTAX) was placed between the image sensor and the corresponding endoscope: a 70^o transoral endoscope (JEDMED) was used for rigid endoscopy and a transnasal fiberscope (model FNL-10RP3, KayPENTAX) for flexible endoscopy. HSV data were recorded at 4000 or 10000 images per second depending upon lighting conditions with maximum integration time and a spatial resolution of 320 horizontal x 480 vertical pixels to capture an approximately 2 cm^2 target area. Images were saved to partitions in a 4 GB on-board memory buffer and transferred to a personal computer's hard drive after the recordings were completed. The buffer restricted the recording time to less than 12 seconds at the lowest desired resolution (4000 images per second). The light source contained a short-arc Xenon lamp rated at 300 watts (KayPENTAX). The fan-cooled housing produced a collimated beam of light with a color temperature of over 6000 K. Three glass infrared (two dichroic, one absorbing) filters blocked infrared light to reduce thermal energy buildup during endoscopy.

The MIC signal was recorded using a head-mounted, high-quality condenser microphone (model MKE104, Sennheiser electronic GmbH & Co. KG) with a cardioid pattern, offering directional sensitivity and a wideband frequency response. The microphone was situated approximately 4 cm from the lips at a 45^o azimuthal offset. The microphone's gain circuitry (model 302 Dual Microphone Preamplifier, Symetrix, Inc.) offered a low-noise, low-distortion preconditioning.

The ACC signal was obtained using a light-weight accelerometer (model BU-7135, Knowles) housed in a 1-inch diameter silicone disk. The accelerometer was preamplified with a custom-made preamplifier [36] and was attached to the suprasternal notch ($\sim$5 cm below the glottis) to obtain indirect estimates of the subglottal pressure. This accelerometer at this location was essentially unaffected by sound radiated

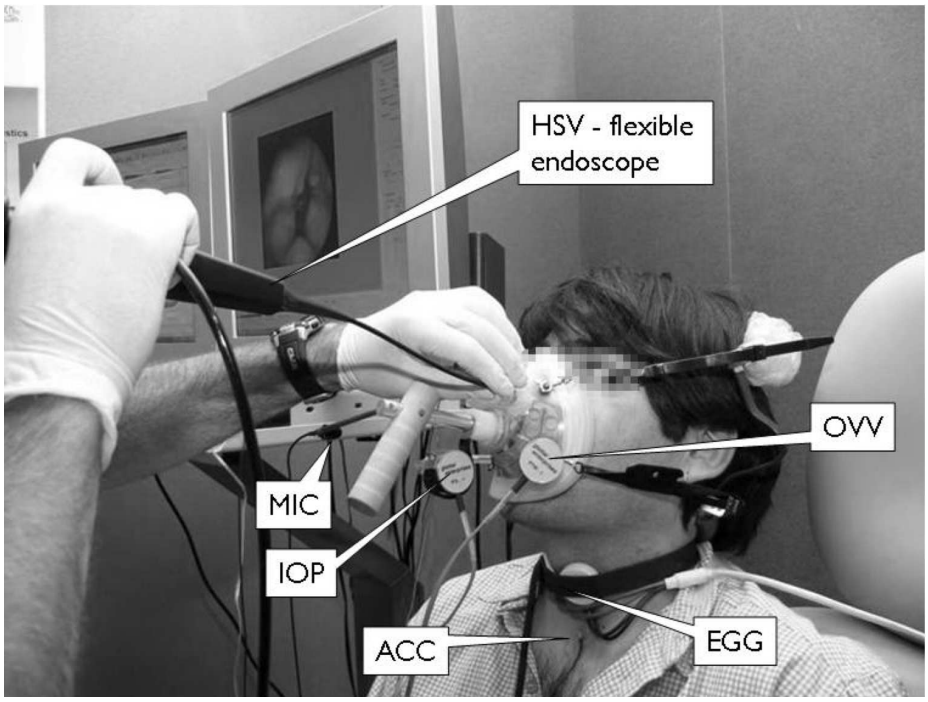


Fig. 4.1. High-speed video measurement and data acquisition system. Flexible endoscopy through a modified CV mask is shown. Real time data visualization is displayed for convenience of the clinician and system operator.

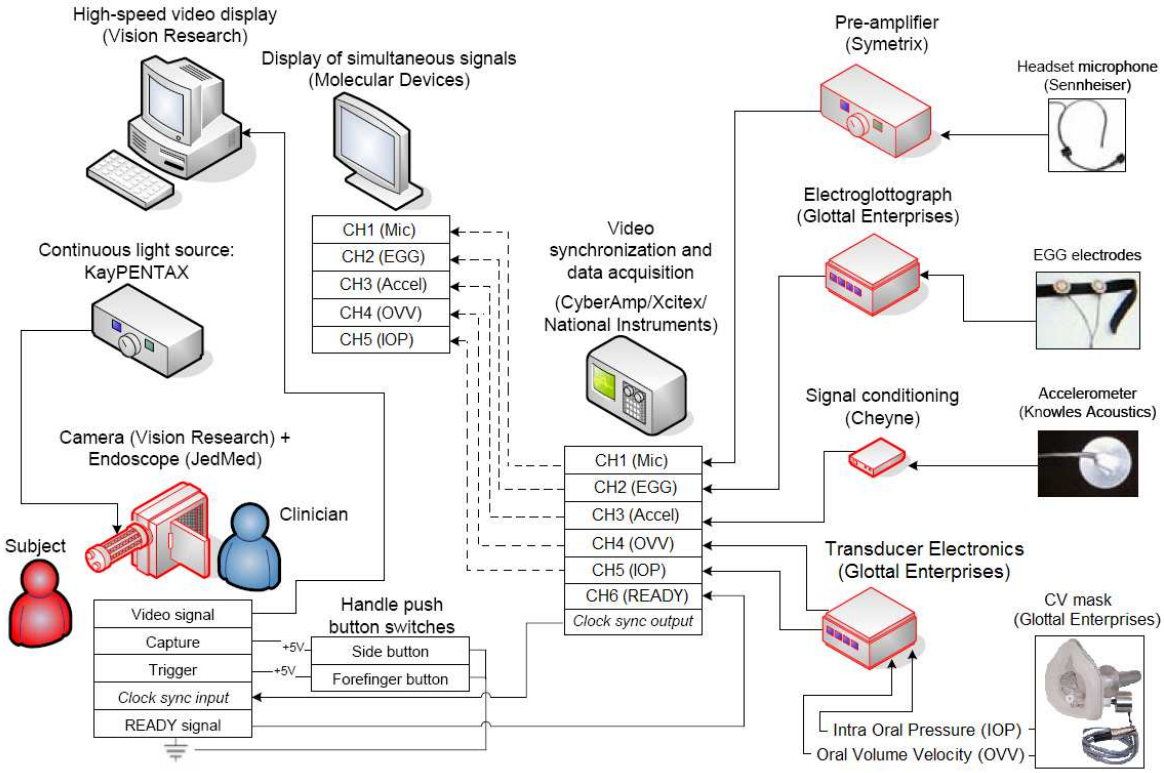


Fig. 4.2. High speed video and data acquisition system setup used at the Laryngeal Surgery & Voice Rehabilitation Center at the Massachusetts General Hospital. Figure modified from [149].

from the subject's mouth (air-borne corrupting components), even with loud vocal intention [171].

An EGG signal was used to provide estimates of glottal contact. The EGG electrodes (model EL-2, Glottal Enterprises) were attached to the neck without interfering with the accelerometer placed at the suprasternal notch. The EGG electrodes were connected to a signal conditioner (model EG2-PC, Glottal Enterprises) which contained both gain and high-pass filter stages.

Simultaneous measurements of OVV required modifying the standard circumferentially-vented (CV) mask (model MA-1L, Glottal Enterprises) to allow for adequate placement of the flexible endoscope with sufficient mobility while maintaining a proper seal [190]. The CV mask was also modified so it could be self-supported around the subject's head and could hold the OVV sensor (model PT-series, Glottal Enterprises), an intraoral pressure (IOP) sensor (not analyzed in these experiments), and the MIC sensor. An electronics unit (model MS-100A2, Glottal Enterprises) provided signal conditioning and gain circuitry for the OVV sensor prior to digitizing. Fig. 4.1 displays the modified CV mask along with other sensors used during the recordings.

Although all signals were calibrated to obtain physically meaningful units, only normalized values are presented in this study for the sake of simplicity. All analog signals were passed through additional signal conditioning and gain circuitry (CyberAmp model 380, Danaher Corp.) with anti-aliasing low-pass filters set with a 3-dB cutoff frequency of 30 kHz and later digitized at a 120 kHz sampling rate, 16-bit quantization, and a $\pm$ 10 V dynamic range by a PCI digital acquisition board (6259 M series, National Instruments).

Time synchronization of the HSV data and the digitized signals was critical for enabling correlations among them and synchronous representations. The hardware clock division and data acquisition settings were controlled by MiDAS DA software (Xcitex Corporation). Alignment of the HSV data and the other signals was accomplished by recording an analog signal from the camera that precisely indicated the

time of the last recorded image. To compensate for the larynx-to-microphone acoustic propagation time, the microphone signal was shifted by 600 μs (17 cm vocal tract length plus 4 cm lip-to-microphone distance), the OVV signal by 500 μs (17 cm vocal tract), and the ACC signal by 125 μs (5 cm distance from the glottis), all into the past relative to the HSV data. Time delays caused by circuitry (Model MS-100A2, Glottal Enterprises) required an additional 100 μs shift into the past for the OVV signal.

4.1.2 Subject selection and protocol

It has been proposed that the strength of both source- and acoustically-induced instabilities can be minimized by vocal training and repeated exposure to unstable regions [49]. Thus, only a reduced percentage of subjects could potentially exhibit both types of voice breaks. In addition, the proposed experimental platform considered complex vocal tasks with relatively invasive sensors employed. These conditions imposed an additional challenge for the subjects to exhibit the desired instabilities.

The most complete experimental setup described was initially tested on eight normal adult subjects uttering simple vocal tasks. Only three of these subjects (two male and one female, the latter with vocal training) completed the more complex protocol required to yield vocal instabilities. Although all three subjects exhibited some type of vocal instability, only one male was able to consistently produce both source-induced and acoustically-induced voice breaks that were clearly observable in the tissue motion. The other two subjects exhibited the following behaviors: 1) the male subject exhibited only a source-induced frequency jump, 2) the female subject only exhibited one minor acoustically-induced instability observable as a subharmonic in the microphone signal that was not observable at the tissue level. Thus, these two cases were discarded to focus on a case study that more clearly illustrated both nonlinear phenomena. The selected subject was an adult male subject with no vocal training and no history of vocal pathology.

Instabilities occurring when F_0 was located within the bandwidth of F_1 (sub or supra) were labeled as acoustically-induced breaks, whereas those occurring when F_0 was outside of this frequency range were labeled as source-induced breaks [49]. To maximize the likelihood of these events, two different vowels were elicited at soft loudness levels: a close front unrounded vowel /i/ (where F_0-F_1 crossings are more likely to occur) and a near-open front unrounded vowel /ae/ (where F_0-F_1 crossings are less likely to occur). Vowel /ae/ is produced naturally when trying to utter a vowel /i/ while a rigid endoscope is in place. Both vowels were uttered as upward and downward pitch glides limited by the subject's vocal range and endoscopic procedure, with no reference tones used.

Data collected from human subjects for this research was approved by institutional review boards at the Massachusetts General Hospital and the Massachusetts Institute of Technology.

4.1.3 Video and data processing

Data and video was processed to perform qualitative observations and quantitative analysis. Six measures obtained from HSV post-processing were used. The main considerations used in this processing are discussed in this section.

Vocal fold tissue motion was measured by tracking the medio-lateral motion of the left and right vocal fold edges closest to the glottal midline (see Fig. 4.3). The tracked edges collapsed the three-dimensional motion of the vocal folds that constituted the mucosal wave to a two-dimensional space that ignored inferior-superior phase differences. Semiautomatic algorithms generated glottal contours, glottal area (Ag), digital kymograms, and phonovibrograms to extract vibratory patterns and different glottal measures. Digital kymograms (DKG) were obtained from three selected cross sections representing the anterior-posterior (AP) glottal axis, as shown in Fig 4.3.

HSV-based measures depended on accurate extraction of vocal fold tissue motion from the time-varying glottal contour. All frames were cropped and rotated such

that the glottal midline was oriented vertically. Glottal area and glottal contours were obtained using threshold-based edge detection. It was found that alternative methods of image segmentation, such as texture analysis [191], watershed transformations [192], and Canny edge detection [193] were not robust to the many variations that occurred in the images obtained, including errant shadowing, arytenoid hooding, and mucus reflections.

Four quantitative HSV-based measures of glottal behavior were computed before and after the voice breaks for two selected DKGs across the AP glottal axis (middle and posterior) where no artifacts in the edge detection were present. The four selected HSV-measures were: open quotient (OQ; ratio between open phase duration and period), speed quotient (SQ; ratio between opening and closing phase durations), left-right (LR) amplitude asymmetry (AA; ratio between amplitude difference and total amplitudes), and LR phase asymmetry (PA; ratio of the time difference between the maximum lateral displacements of the left and right vocal folds and the open phase duration). These measures have been used to study soft, normal, and loud voice [39], register transitions [132], and normal and pathological cases [194–196].

Phonovibrograms (PVG) [197] are spatio-temporal plots constructed from lateral displacement waveforms of the vocal folds. The color scheme proposed by [197] was simplified into a grayscale representation. Empirical Orthogonal Function (EOF) analysis was performed following the decomposition described by [120] before and after the voice breaks. Both PVG and EOF were obtained from the glottal area contour for each time step, encompassing no less than 30 cross sections for each vocal fold. Two quantitative measures are extracted from the decomposition: the relative weights and entropy measure, both calculated before and after the break for each vowel. The relative weights of the EOF depicted the contribution of different empirical modes of vibration and the information entropy measure (referred as S_{tot} following the notation from [120]) represented the spatial irregularity and broadness of the mode distribution.

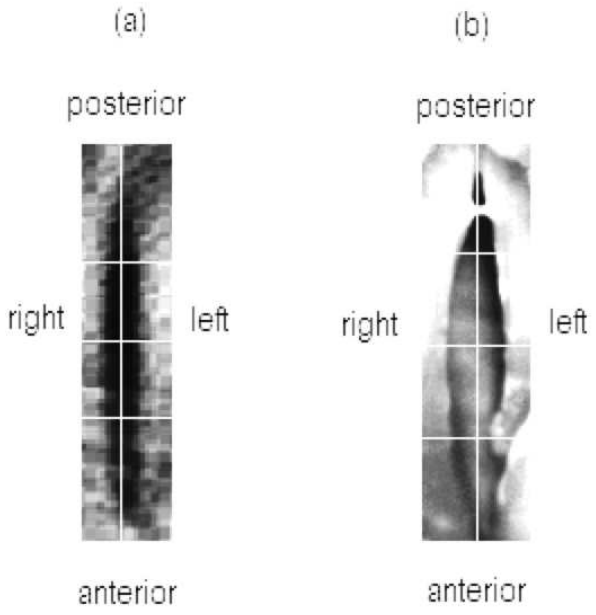


Fig. 4.3. Endoscopic view obtained with with (a) flexible endoscope and (b) rigid endoscope. White horizontal lines indicate the locations of the three selected DKGs. White vertical line indicates the glottal midline.

Center frequencies and bandwidths of the supraglottal and subglottal resonances were computed from the MIC and ACC signals, respectively. The covariance method of linear prediction was used to estimate the pole distributions within the closed phased portion of the vocal fold cycle. The closed phase was determined using the derivative of the electroglottogram (dEEG) [140]. A 50 ms separation from the break point was taken into account to ensure some stability in the signal.

Spectral analysis was also included to match representations used in previous studies dealing with register changes and acoustic interaction [49, 148]. Thus, spectrographic analysis used a Hamming window of 30 ms duration with 8192 FFT points and 90 % overlap for a dynamic range of 60 dB.

4.2 Results

4.2.1 Initial screening

A summary of all vocal tasks that yielded some type of voice instability for the subject in this case study is presented in Table 4.1. Three types of instabilities were observed: pitch jumps, pitch fluctuations, and aphonic segments. Pitch jumps were found to be the most frequent and the most easily repeatable instability and can also be related to bifurcations. For those instabilities labeled with F_0-F_1 crossings (sub and supra), the pitch was observed to have sudden changes before and after the unstable zones, matching the observations in [49].

The primary interest of this investigation is on bifurcations, for which the focus is placed on the frequency jump instabilities from Table 4.1. For vowel /ae/, bifurcations were more easily observed in the ascending pitch glides and only one instance exhibited a bifurcation in the descending pitch glide. Vowel /i/ exhibited the inverse pattern, i.e., the most repeatable bifurcations were on the descending pitch glide and only once a bifurcation was observed in the ascending glide. The average and standard deviations for the fundamental frequency before and after the bifurcations for these cases is summarized in Table 4.2. It can be observed that a more consistent behavior was present on the onset of the bifurcation for both vowels and that hysteresis was observed in both cases. This last observation is less well supported since certain gestures needed to describe hysteresis were only observed once. The fact that the subject was less prone to exhibit instabilities for different conditions may be associated to his familiarity with certain gestures or an effect of the acoustic coupling.

For the subsequent analyses in this chapter, the main focus is on the gestures that were more consistent, i.e., the descending pitch glide for vowel /i/, and ascending pitch glide with a vowel /ae/. These two cases also allow for comparing the presence or lack of F_0-F_1 crossings, regardless of the pitch glide direction. The selected HSV recordings within these cases (denoted by ** in Table 4.1) were within the expected ranges with respect to other experimental configurations, thus ruling out possible effects of the

Table 4.1

Pitch glides exhibiting voice instabilities in the case study. Instances with HSV are denoted by (*) and (**), where the latter were used for post-processing. The labeling for F_0-F_1 crossings was defined when the pitch was: within the bandwidth of the first vocal tract formant (labeled as "supra"), the first subglottal resonance (labeled as "sub"), or outside of them (labeled as "no").

Vowel	pitch glide	F_1 (Hz)	F_2 (Hz)	F_1sub (Hz)	F_2sub (Hz)	F_0 before (Hz)	F_0 after (Hz)	voice break	F_0-F_1 cross.
/ae/	up	697	1229	513	1427	420	399	aphonic	sub
/ae/	down	654	1179	555	1406	498	456	aphonic	sub
/ae/	down	647	1172	527	1484	239	130	jump	no
/ae/	up	718	1208	569	1399	151	279	jump	no
/ae/	up	661	1413	576	1243	172	307	jump	no
/ae/*	up	583	1442	491	1271	158	286	jump	no
/ae/*	up	619	1392	669	1541	172	314	jump	no
/ae/**	up	551	1343	495	1363	159	325	jump	no
/i/	up	335	2491	555	1300	335	293	aphonic	supra
/i/	up	357	2370	498	1335	442	420	aphonic	sub
/i/	down	350	2356	477	1413	201	116	jump	no
/i/	up	328	2604	491	1371	335	442	jump	supra
/i/	down	286	2498	513	1447	293	158	jump	supra
/i/	up	335	2398	484	1342	137	236	jump	no
/i/	down	350	2342	569	1484	513	413	aphonic	sub
/i/	down	321	2363	562	1420	293	165	jump	supra
/i/*	down	327	2254	543	1435	342	307	dip	supra
/i/**	down	327	2254	549	1274	305	190	jump	supra

Table 4.2
Pitch glides exhibiting bifurcations in the case study.

Vowel	pitch glide	F_0 before		F_0 after	
		mean (Hz)	stdv (Hz)	mean (Hz)	stdv (Hz)
/ae/	up	164	9	304	22
/ae/	down	239	-	130	-
/i/	up	137	-	236	-
/i/	down	274	8	158	19

CV mask and endoscope on the unstable behavior. These two recordings described transitions between chest and falsetto registers and were analyzed in detail in the following sections.

4.2.2 Spectrographic observations

Spectrographic and temporal representations of 500 ms around the voice breaks for both MIC and ACC signals are presented for both vowels under consideration in Fig. 4.4 and Fig. 4.5. It can be seen in Fig. 4.4 that vowel /i/ exhibited no transitional changes were observable before or after the break, i.e., both signals suddenly jumped from one vibratory pattern to another with a short, less periodic region immediately after the break that produced inter-harmonic noise (as seen between 325 ms and 425 ms). Contrasting these observations, Fig. 4.5 shows that vowel /ae/ exhibited a gradual change in the harmonic composition before the break, where the second harmonic component (noted as ripple in the temporal representations) was increasing in amplitude (as seen between 150 ms and 250 ms) up to the point of the voice break. This harmonic component became the fundamental frequency after the bifurcation.

4.2.3 High-speed videoendoscopy sequences

A series of HSV sequences spanning a 30 ms window around the bifurcation point is presented for each vowel. The sequence for vowel /i/ is displayed in Fig. 4.6 and has a time span of 10 ms per row. A few cycles before and after the break are observed in the first and last row respectively, whereas the transition between the two registers is depicted in the second row. Differences between the vibratory patterns before and after the break were observed. Before the break, the glottis opened and closed uniformly along the AP direction. After the break, a posterior opening with shorter duration, higher degree of skewing and asymmetry, and reduced amplitude was observed. In addition, the transition between these two modes had a distinct feature toward the end of the second row, where a much larger glottal

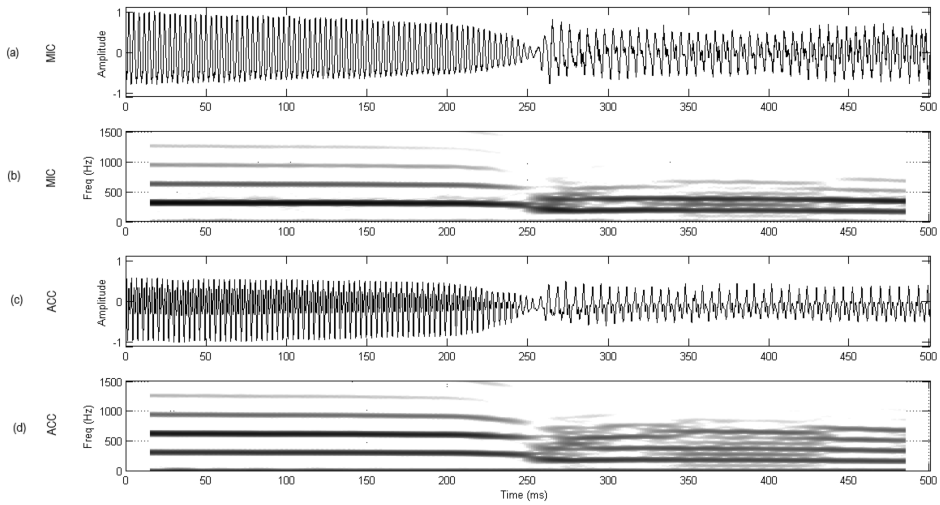
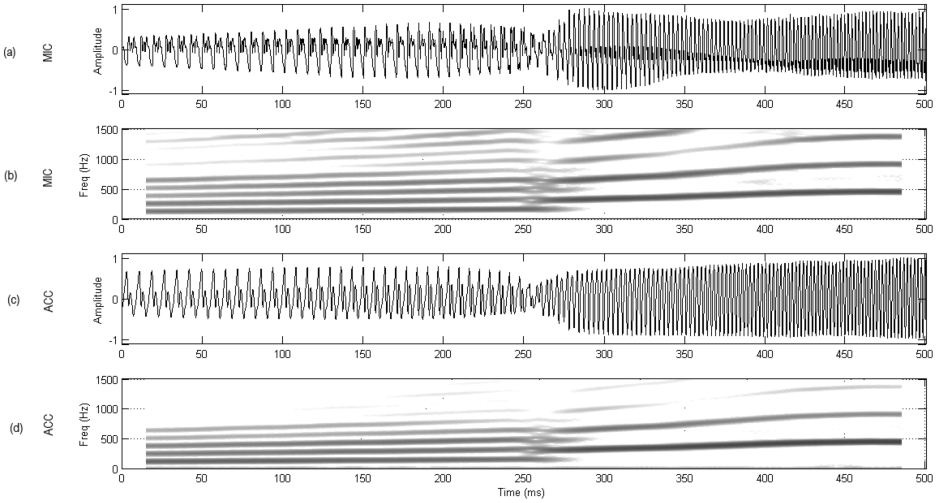


Fig. 4.4. Downward pitch glide for vowel /i/ using the flexible endoscopy setup: a) normalized microphone signal, b) microphone spectrogram, c) normalized accelerometer signal, d) accelerometer spectrogram.



hp]

Fig. 4.5. Upward pitch glide for vowel /ae/ using using the rigid endoscopy setup: a) normalized microphone signal, b) microphone spectrogram, c) normalized accelerometer signal, d) accelerometer spectrogram.

excursion was observed right before the beginning of the chest register. This event appeared consistent with previous observations from Fig. 4.4, where the transition between registers was abrupt. Furthermore, the interval between this marked pulse and the one before exhibited incomplete closure with phase asymmetry in the lateral displacement observed as parallel LR tissue motion.

The HSV sequence for vowel /ae/ is presented in Fig. 4.7, displaying 10 ms per row. A few cycles before and after the break are shown in the first and last row, and the transition between them is depicted in the second row. In contrast with vowel /i/, no significant differences between the vibratory patterns were observed before and after the break. The glottis did not exhibit AP differences in excursion, opening, or closing times. In addition, a smooth transition was observed between the two registers, matching the observations from the MIC and ACC signals for this vowel.

HSV recordings by the same subject during modal speech and sustained pitch exhibited similar differences in the AP direction between the same two vowels. Thus, the more complex behavior observed for vowel /i/ in chest register may be introduced by either differences in laryngeal configuration or by acoustic coupling effects due to the much lower first formant present in that case.

In addition, direct observation of the complete laryngeal view in the HSV depicted a clear displacement of the arytenoid cartilage before and after the voice break in vowel /ae/, movement that was not observed for vowel /i/.

4.2.4 Synchronous spatio-temporal observations

Fig. 4.8 presents the set of synchronous plots for vowel /i/. As in Fig. 4.4, a clear difference was observed in the signal structure for the MIC and ACC signals before and after the break (occurring at around 70 ms). The dEGG signal was weak before the break, nonexistent during it, and very strong after it. This indicated the nature of the collision forces at the glottis and the lack of contact during the break. In addition, multiple contact points were observed during the beginning of the closed phase during

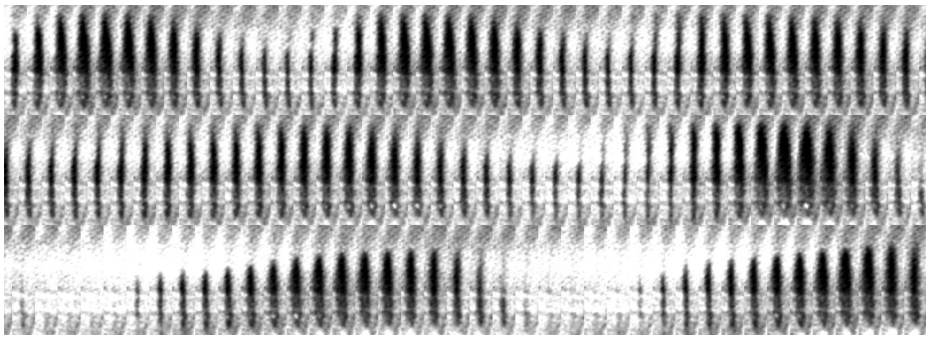


Fig. 4.6. Snapshot sequence of voice break for vowel /i/. Time is represented from left to right and spans 10 ms per row with a 250 μs period between subsequent frames. Laryngeal view is cropped for viewing purposes.

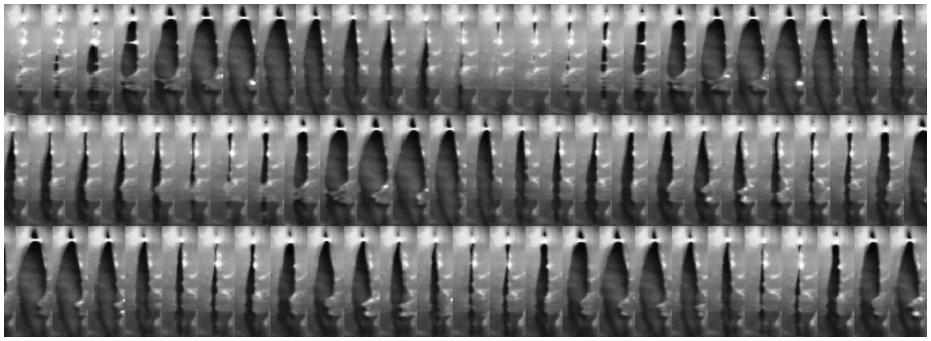


Fig. 4.7. Snapshot sequence of voice break for vowel /ae/. Time is represented from left to right and spans 10 ms per row with a 400 μs period between subsequent frames. Laryngeal view is cropped for viewing purposes.

the 50 ms following the break. This pattern was correlated with the high-frequency ripples observed in the MIC and OVV signals at the same time. Since no mucus was observed in the HSV, this high-frequency behavior was suspected to be product of the tissue motion in that region. The no-contact region observed in the dEGG was also observed as a clear DC drift in the OVV signal in the same region. The glottal area waveform (Ag) illustrated that the frequency changed after the break, as well as the area properties of shape, skewing, and closed and open phase durations.

The DKGs from Fig. 4.8, exhibited significant changes in the oscillatory behavior before and after the break, as well as in the AP direction. Before the break, all three DKGs exhibited excursions of comparable amplitudes with an opening time similar to the closing time. However, after the break, the DKGs had different lateral displacement amplitudes and shapes. The posterior DKG differed from the other two DKGs in that its lateral displacement waveforms had a round shape with smaller amplitude. The anterior and middle DKGs had longer opening and shorter closing portions, which explained the skewing of Ag. In addition, the break portion consistently exhibited incomplete closure and LR phase asymmetry (parallel wall motion), which yielded a large peak in Ag.

The PVG in Fig. 4.8 further elucidated the differences before and after the break in the vibratory patterns of the vocal folds. Before the break, symmetric behavior was observed between the left and right vocal folds and along the AP direction, where the entire glottal edge opened simultaneously. The break exhibited LR asymmetries and a constant opening that ended in an abrupt closure around 75 ms. After this point, an AP difference was observed in the oscillation, where the anterior ends exhibited most of the lateral excursion. The slightly skewed pulses indicated that glottal opening and closure did not occur at the same time along the AP axis. In addition, the regions with maximal excursion (brighter regions) deviated toward the right (in time) with respect to the pulses before the break. This tissue motion indicated abrupt glottal closure that produced the skewing of the glottal area waveform and was hypothesized to yield the high-frequency ripples observed in MIC and OVV signals after the break.

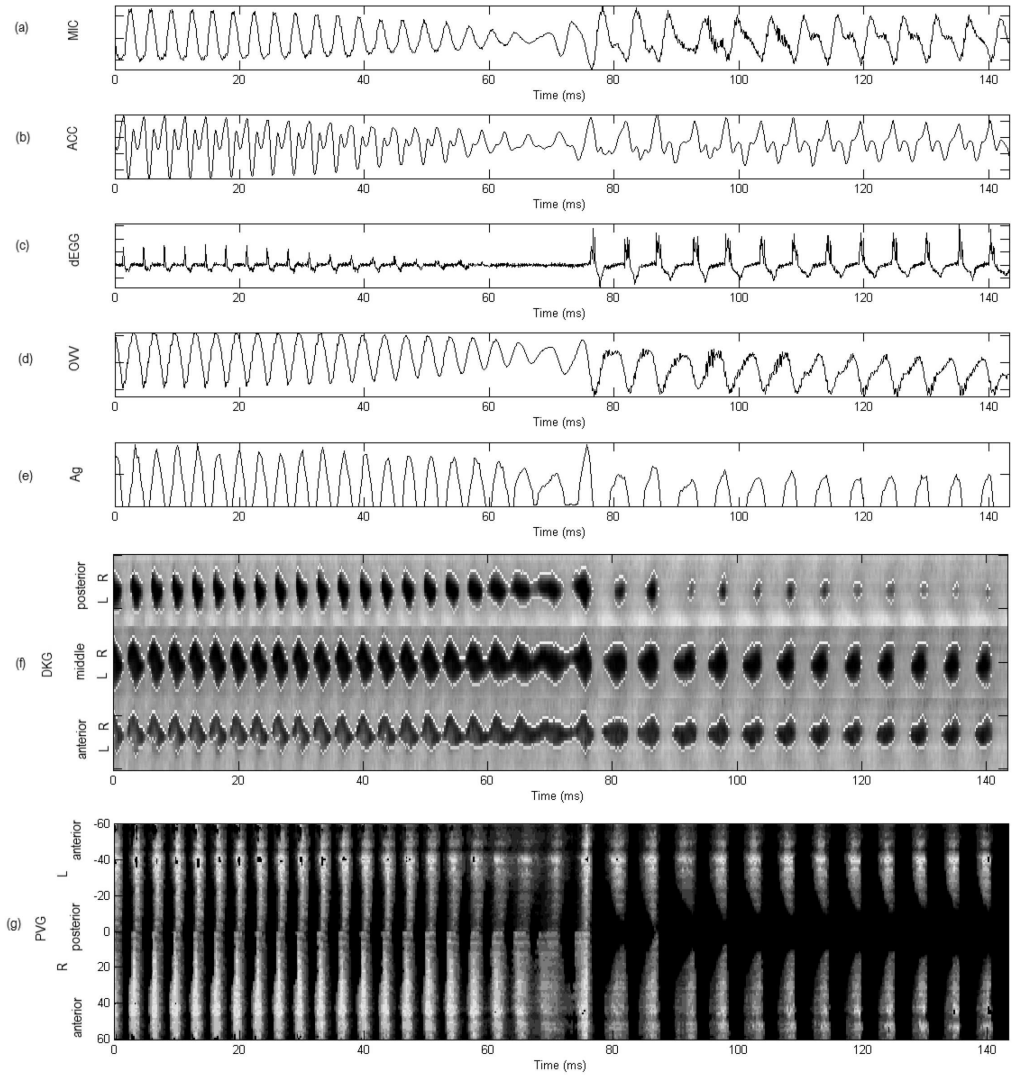


Fig. 4.8. Synchronous representation for vowel /i/: a) microphone, b) accelerometer, c) derivative of electroglottograph, d) oral volume velocity, e) glottal area, f) anterior, middle, and posterior kymograms, g) phonovibrogram. All signals normalized. The normalized phonovibrogram grayscale indicates maximum amplitude in white.

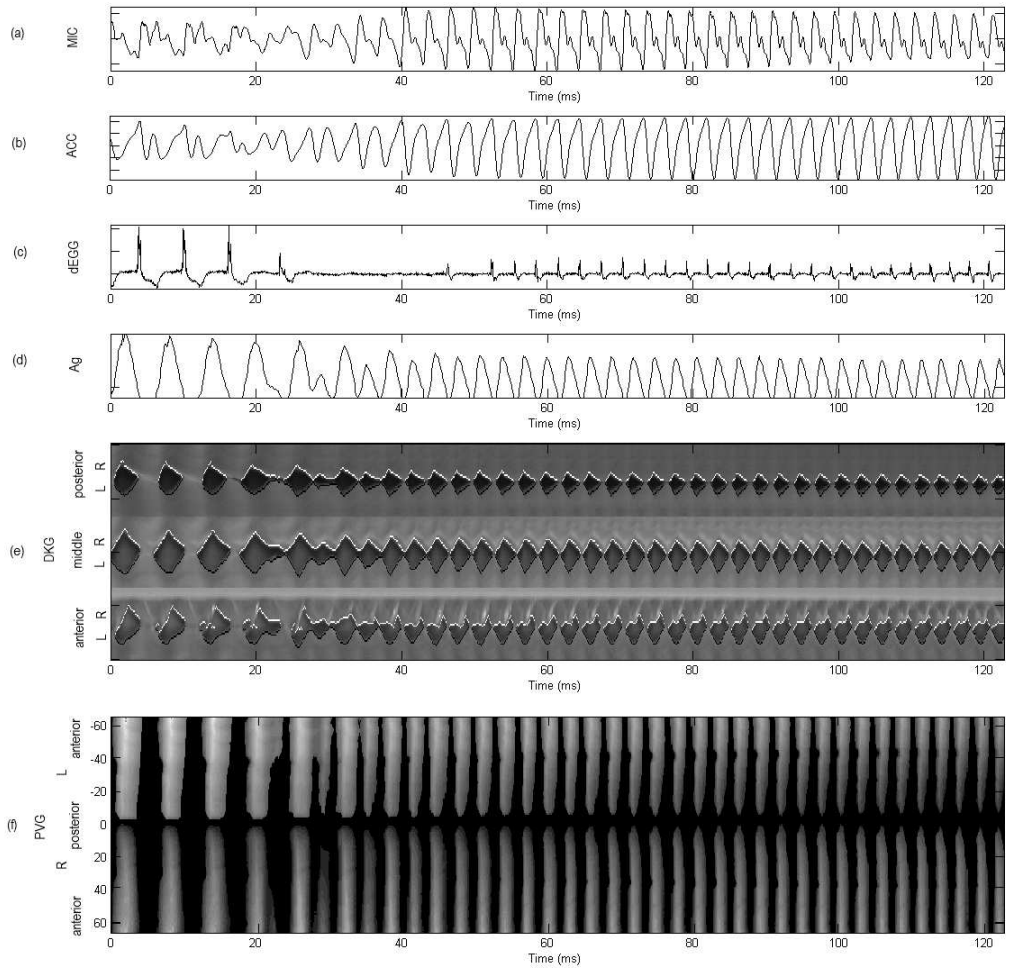


Fig. 4.9. Synchronous representation for vowel /ae/: a) microphone, b) accelerometer, c) derivative of electroglottograph, d) glottal area, e) anterior, middle, and posterior kymograms, f) phonovibrogram. All signals normalized. The phonovibrogram grayscale indicates maximum amplitude in white.

A different scenario is observed for the synchronous plots of vowel /ae/ in Fig. 4.9. The MIC and ACC signals exhibited a more stable behavior before and after the break, and a much smoother transition between the two registers. Similar type of transitions were observed by [132] for source-induced register jumps. As expected, the dEGG indicated that the contact in the chest register was stronger than in the falsetto register. The glottal area waveform illustrated how a higher harmonic component was increasingly appearing during the break, joining both oscillatory regimes smoothly.

The spatio-temporal plots in Fig. 4.9 show a much simpler structure compared with vowel /i/, exhibiting AP uniformity and LR symmetry before and after the break. Both DKGs and PVG illustrated that an additional harmonic pulse was smoothly introduced before the voice break, anticipating the second vibratory pattern. This claim disregarded the mucus that distorted the edge detection of the left vocal fold (see Fig. 4.7), primarily in the anterior DKG. However, the edge detection of the right vocal fold was less prone to these artifacts and was thus preferred in the analysis.

4.2.5 High-speed videoendoscopy based measures

Table 4.3 presents the four selected HSV-based measures of glottal behavior, each one computed for the chest and falsetto registers and both vowels.

A reduction in OQ in the chest register was observed for both vowels as the closed portion gets larger in this case. This expected behavior is in agreement with [132]. Even though comparable differences were observed in OQ for both vowels and registers, a shorter OQ was obtained in the chest registers in the posterior end of vowel /i/, illustrating the different AP behavior between the two vowels.

Similarly, a SQ increases due to the reduction of the closing phase (i.e., glottal area skewing to the right). A greater change in the SQ was observed for vowel /i/ in the chest register. AP differences are shown in this vowel since the posterior end had a more symmetric shape (SQ closer to 100 %). Vowel /ae/ shown less significant changes and rather maintains its SQ for both registers. This is due to the minor

Table 4.3
HSV measures taken from two DKGs representing middle and posterior tissue motion during chest and falsetto registers.
Index notation: (f)= falsetto register, (c)= chest register

HSV-based measures from AP DKGs	Symbol	Vowel /i/ middle mean ±stdv (%)	Vowel /i/ posterior mean ± stdv (%)	Vowel /ae/ middle mean ± stdv (%)	Vowel /ae/ posterior mean ± stdv (%)
Open quotient (f)	OQf	83.3 ± 2.9	61.3 ± 2.2	91.0 ± 1.7	78.0 ± 3.7
Open quotient (c)	OQc	53.1 ± 5.3	35.3 ± 15.5	71.8 ± 1.8	54.5 ± 0.6
Speed quotient (f)	SQf	84.0 ± 11.3	53.9 ± 16.7	66.3 ± 6.1	58.0 ± 7.1
Speed quotient (c)	SQc	189.9 ± 40.7	119.6 ± 59.9	93.8 ± 8.8	48.9 ± 4.9
Amplitude asymmetry (f)	AAf	10.1 ± 7.4	18.3 ± 6.3	2.7 ± 2.8	7.1 ± 4.1
Amplitude asymmetry (c)	AAc	-4.0 ± 6.2	-7.6 ± 18.8	-19.2 ± 7.0	-22.7 ± 6.4
Phase asymmetry (f)	PAf	14.0 ± 2.2	14.3 ± 2.8	3.5 ± 1.8	0.8 ± 0.9
Phase asymmetry (c)	PAc	6.8 ± 3.6	5.0 ± 2.5	0.4 ± 0.6	4.5 ± 1.7

changes that the glottal area waveform and DKGs exhibited between the two registers for this vowel.

The asymmetry measures (AA and PA) were useful to identify differences between LR sides that were not obvious by simple observation of the spatio-temporal plots. Both measures of LR asymmetry were within the normal range for both vowels before and after the break of 14.9 % for the magnitude of PA and 15.8 % for the magnitude of AA, both obtained from the middle DKG [196]. Comparable changes in polarity were observed in AA in both vowels between registers, indicating that the left vocal fold had a larger displacement in the falsetto register. Differences between the registers were more noticeable in the posterior DKGs in both vowels, although larger AP differences were observed for vowel /i/. In addition, PA was uniformly low along the AP direction and also exhibited larger changes for vowel /i/.

4.2.6 Empirical Orthogonal Function decomposition

The spatio-temporal plots of each vowel were divided between falsetto and chest registers and separated by left and right vocal folds for the EOF analysis. Any possible artifact (e.g., mucus or edge detection artifacts) was disregarded to improve the estimates of the mode decomposition. The portions that are analyzed in this section are illustrated in Fig. 4.10 for vowel /i/ and Fig. 4.11 for vowel /ae/.

The most dominant EOF weights for the two vowels, for each register and separated by left and right vocal folds are shown in Fig. 4.12. The decomposition showed that the first modes of dominated the pattern of vibration in all cases. However, vowel /i/ had a slower decay and broader distribution, thus needing more modes of vibration to decompose the oscillatory patterns. A broader weight distribution was observed in the chest register of vowel /i/ compared with the chest register of vowel /ae/. In addition, an apparent increase in the distribution broadness was noted after the voice break in vowel /i/.

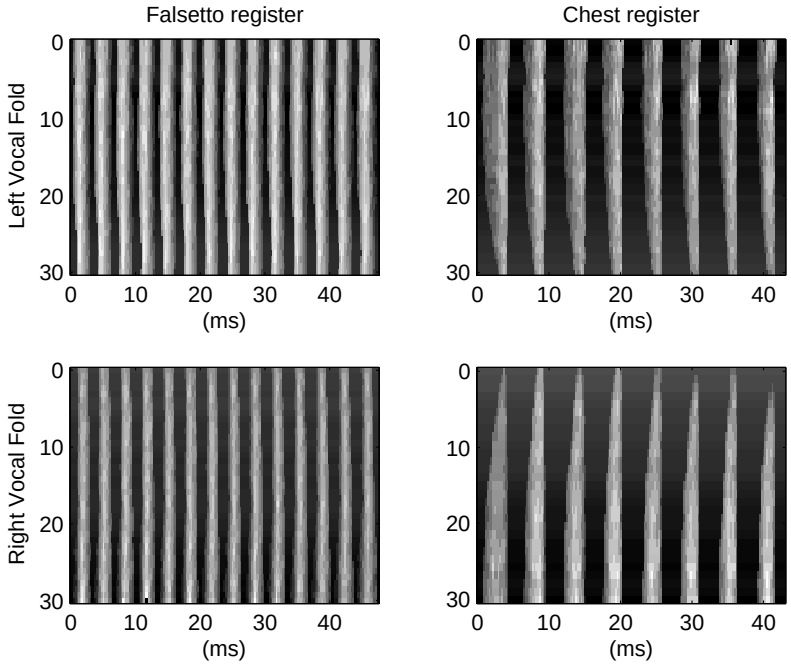


Fig. 4.10. Selected portions of spatio-temporal plots for EOF analysis for falsetto and chest register for vowel /i/.

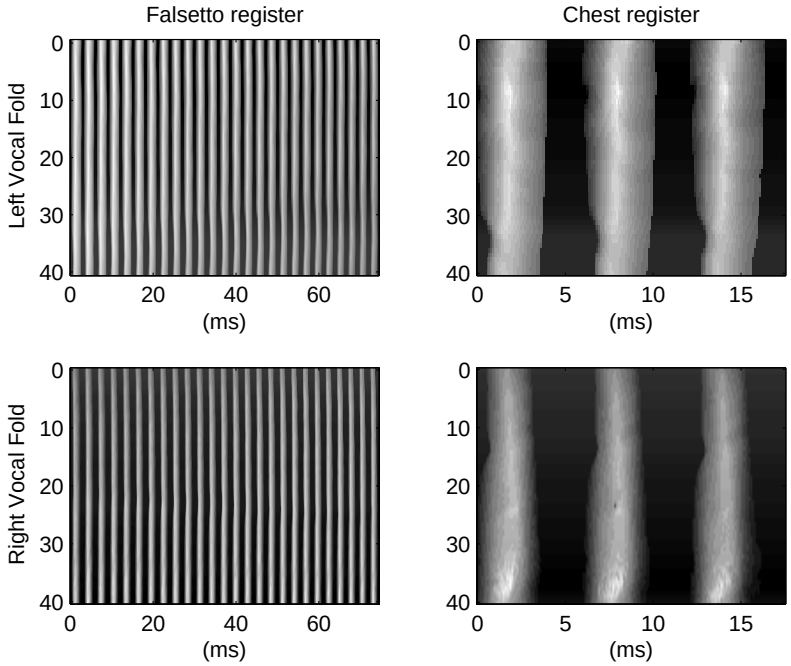


Fig. 4.11. Selected portions of spatio-temporal plots for EOF analysis for falsetto and chest register for vowel /ae/.

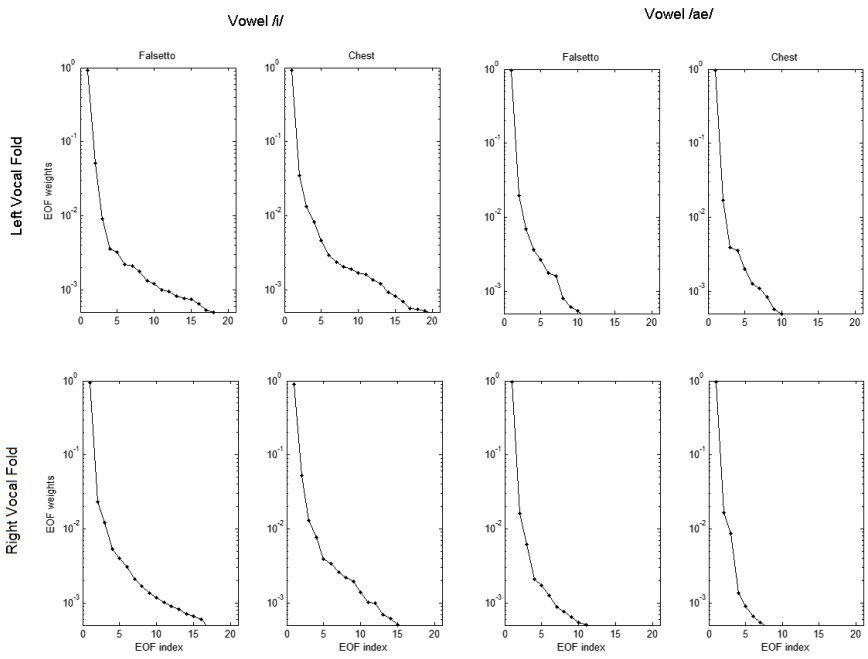


Fig. 4.12. First 21 values of the relative EOF weights for the two vowels before and after the break. First and second rows indicate left and right vocal fold, respectively.

To better describe the EOF weight relations, the cumulative sum of the first few most dominant relative EOF weights is presented in Table 4.4. As suggested by [120], when the total cumulative surpasses 97%, a sufficient precision is obtained in the reconstruction of the vibratory pattern. The values above this threshold are highlighted in Table 4.4 to emphasize the number of modes needed in each case.

This analysis shows the same trends observed in Fig. 4.12. Furthermore, it is noted that vowel /ae/ has a behavior that fits with the distribution of EOF weights for normal cases reported in previous studies [120], where the first two modes capture most of the vocal fold dynamics. On the other hand, a larger number of modes where needed to compose the patterns of vibration for vowel /i/, particularly for the chest register, where more than five modes are needed to meet the 97% threshold. This behavior is expected to be a consequence of the AP asymmetry and more irregular tissue vibration observed before and after the acoustically-induced break observed for vowel /i/. These observations help to quantify previous observation on the complexity of the acoustically-induced case.

4.3 Discussion

The aim of these experiments was to compare voice breaks occurring with and without F_0-F_1 crossings. A comprehensive set of measurements was performed as a case study of an adult male with no history of vocal pathology. The subject exhibited consistent behavior for two desired vocal gestures: a descending pitch glide of a vowel /i/ and an ascending pitch glide of a vowel /ae/. Given that for vowel /i/ there was a clear F_0-F_1 (vocal tract) crossing, such a break was labeled as acoustically induced, whereas that of vowel /ae/ with no F_0-F_1 crossing was considered source induced. The most consistent unstable behavior for the vowel gestures was found to be during jumps in the fundamental frequency that were associated with register transitions.

The initial observations indicated that the source-induced case had observable arytenoid displacement that can be associated with active tension control, contrasting

Table 4.4
Cumulative sum of the first five relative EOF weights for each vowel both before and after the voice breaks. Values above a 97% threshold are underlined to define the number of modes needed for the reconstruction.

	Vowel /i/				Vowel /ae/			
EOF	left vocal fold		right vocal fold		left vocal fold		right vocal fold	
index	falsetto	chest	falsetto	chest	falsetto	chest	falsetto	chest
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1	91.4	91.6	93.6	90.4	95.8	96.7	96.3	96.6
2	96.6	95.0	95.9	95.7	97.7	98.4	98.0	98.2
3	97.5	96.4	97.1	96.9	98.4	98.8	98.6	99.1
4	97.8	97.2	97.7	97.7	98.8	99.1	98.8	99.2
5	98.1	97.7	98.1	98.1	99.1	99.3	99.0	99.3
S_{tot}	0.19	0.21	0.17	0.21	0.11	0.09	0.11	0.09

the acoustically-induced case where no significant arytenoid movement was observed. Transitional changes were observable before the break for the source-induced case, contrasting the acoustically-induced case where the signals suddenly jumped from one register to another with no observable transitions.

These results link the source-induced case with gradual changes in vocal fold tension, which is in agreement with previous studies where smooth changes in tension triggered jumps to a higher mode of vibration, particularly when the oscillation was near coexisting limit cycles [103, 123, 124, 129, 130]. Thus, voice breaks occurring at non F_0-F_1 crossings appeared to better match these source-induced factors and not a destructive interference with sub-harmonic ratios of the subglottal resonances [62,81].

Observable differences at the tissue level can be seen when comparing the acoustically-induced and source-induced breaks. Different analyses showed differences in the tissue motion before and after the voice breaks. Asymmetries between the vocal folds were only present shortly before the breaks, particularly in the acoustically-induced case, where phase asymmetry was observed as parallel lateral motion that was followed by a strong and abrupt closure. In contrast, the source-induced break showed a relatively smoother transition between registers and matched the general behavior observed in [132]. The source-induced break had a more symmetric behavior before and after the break, in contrast with the acoustically-induced case, where anterior-posterior differences were evident. However, it is unclear if these anterior-posterior differences were introduced by the coupling effect or by a particular laryngeal configuration. It is possible that these anterior-posterior differences are associated to the laryngeal configuration for vowel /i/, but were suppressed by the stronger source-filter coupling before the bifurcation. Further research is needed to verify this explanation.

The differences observed between cases labeled as source-induced and acoustically-induced bifurcations support the hypothesis that acoustic coupling can introduce visual differences in tissue motion. The presence of strong acoustic coupling does appear to facilitate register transitions, as the frequency jumps occurred earlier (i.e., at higher frequencies during descending pitch glide and vice versa) when strong coupling

was present. Further investigations with a larger pool of subjects are needed to better support these findings. Nevertheless, these initial observations support nonlinear source-filter coupling theory [17] and its principles where the acoustic coupling was described based on impedance representations.

The results of these experiments are consistent with the use of F_0-F_1 crossings to label bifurcations as "source-induced" and "acoustically-induced", as proposed by Titze [49]. Further investigations will need to test the robustness of this classification since in many instances bifurcations can be observed in ranges where it difficult to establish if they occurred within the formant bandwidth. An alternative classification scheme might be obtained by investigating the hysteresis of the bifurcation and utilizing the distinction between supercritical Hopf bifurcations (smooth transitions) and subcritical Hopf bifurcations (amplitude jumps with hysteresis) [128]. The results of the experiments in this chapter illustrate that both designated "source-induced" and "acoustically-induced" cases exhibited hysteresis and some degree of amplitude differences before and after the breaks, for they would classify as subcritical Hopf bifurcations. This finding is in agreement with previous numerical simulations with and without acoustic interaction [128]. However, it is important to note that a rigorous analysis of the hysteresis was not possible in our experiments since the subject tended to exhibit bifurcations in only one of the pitch glide directions for each vowel. This behavior may be related to the subject's ability to compensate the instabilities in one direction more than in other for certain vowels, a laryngeal configuration that affects the bifurcation for each vowel, the effects of the source-filter coupling, or a combination of these factors. This tendency was also observed in some cases in previous studies [49]. Furthermore, the "source-induced" case presented a smoother transition, which would have suggested the presence of a supercritical Hopf bifurcations if no hysteresis was present. Thus, it appears difficult to attain a controlled hysteresis analysis in human subjects recordings that involve bifurcations during pitch glides and different vowels. Indeed, this may make difficult the use of sub- and supra-critical Hopf bifurcation classification in some cases. In summary, further investigations will need

to evaluate the most adequate classification scheme for the added effect of source-filter coupling in voice bifurcations.

It is noteworthy to comment on the difficulties associated with the subject recruitment in these experiments. As noted by [49], only a reduced percentage of the subjects were able to achieve the desired voice breaks, even for a simple scheme that did not include endoscopy. This finding, along with the more complex experimental setup conditions (including the need to attain full glottal exposure) imposed a challenge for the subjects to accomplish the vocal tasks and exhibit the desired instabilities. Similar challenges were observed in [148], in which only a single untrained subject was able to accomplish the desired task. Low yield in subject pools appear to be intrinsic to experiments where participants are expected to produce complex vocal tasks with relatively invasive sensors employed. Although expanding the current efforts on the effects of acoustic coupling on tissue dynamics is planned, subject recruitment is expected to continue being a practical limitation. This issue also questions the applicability of pitch glide maneuvers as part of routine clinical assessment of vocal function, at least when it includes simultaneous observations of laryngeal dynamics.

4.4 Conclusions

This chapter introduced a comprehensive analysis of vocal fold tissue motion and related measurements during acoustically-induced and source-induced unstable oscillations, aiming to further explore the theory of nonlinear coupling in phonation proposed by [17]. Simultaneous recordings were used, including flexible and rigid laryngeal high-speed videoendoscopy, skin acceleration, oral volume velocity, electroglottography, and radiated acoustic pressure for different vocal gestures. Instabilities were labeled as acoustically-induced when F_0-F_1 crossings where observed, and conversely, source-induced when not. The high-speed video recordings analyzed in this paper are believed to be the first fully documented *in vivo* visualizations of acoustically-induced instabilities.

The results of this chapter showed that differences can be observed between the two types of voice instabilities through laryngeal high-speed videoendoscopy. At the tissue level, acoustically-induced vocal fold instabilities appeared to be more abrupt, exhibited left-right phase asymmetry observed as parallel wall motion, and yielded anterior-posterior differences after the break. Source-induced instabilities showed a smoother transition between oscillatory modes. It appears that acoustic coupling affects the tissue motion before the break, suppressing anterior-posterior differences that are associated to the laryngeal configuration. The results also suggest that strong acoustic interaction can facilitate register transitions by adding an additional acoustic loading effect near transitional zones. Both types of breaks exhibited hysteresis and some degree of amplitude changes after the breaks which would link them to subcritical Hopf bifurcations. However, a rigorous hysteresis analysis was not possible as the subject tended to exhibit voice breaks in one of the pitch glide directions more than the other. Nevertheless, these results are in agreement with previous studies and support nonlinear source-filter coupling theory and descriptions of acoustic coupling in term of lumped impedances. Future numerical and experimental studies are needed to corroborate the observations in this case study.

The effect of the tract impedances on glottal phonatory stability deserves further attention. These tract effects could be even more significant in pathological voices. A strong understanding of the underlying three way interactions between flow, tissue, and sound is key for the further development of related vocal diagnosis, training, and treatment.

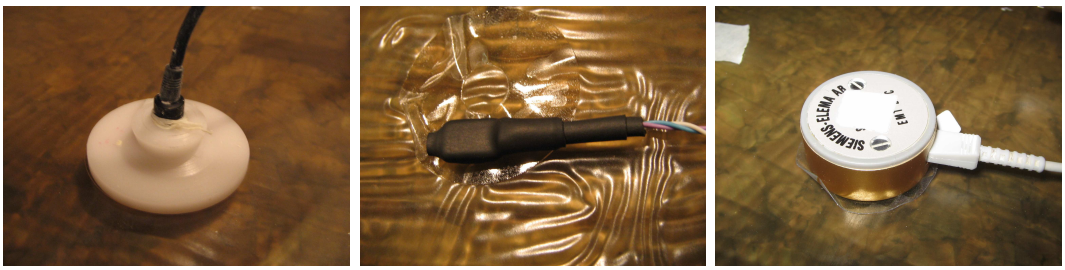
5. BIOSENSING CONSIDERATIONS FOR SKIN SURFACE ACCELERATION MEASUREMENTS

In order to implement an inverse filtering technique that makes use of neck surface acceleration to retrieve glottal airflow, a careful inspection of biosensing associated with recordings of skin surface vibration is needed. This chapter describes methods, results, and discussion of two properties of interest in bioacoustic sensors used on the skin surface: tissue-borne and air-borne sensitivities. Technical specifications from such sensors are expected to change when placed on human skin, and background noise can corrupt recordings made with them. Thus, the sensitivity of commonly used bioacoustic sensors to air-borne sounds was evaluated and compared with their sensitivity to tissue-borne body sounds. To delineate the sensitivity to each pathway, the sensors were first tested *in vitro* and then on human subjects. These sensitivities and their relationships are expected to define the adequacy of bioacoustic sensors for the desired inverse filtering application. The results supported the selection of a specific bioacoustic sensor for further investigations. Additional insights to those reported in [171] are provided.

5.1 Methods

The sensitivity of three commonly used bioacoustic sensors to air-borne sounds was evaluated and compared with their sensitivity to tissue-borne body sounds. The selected sensors were an air-coupled microphone (Sony ECM-77B), a light-weight accelerometer (Siemens EMT25C) and a very light-weight accelerometer (Knowles BU-7135). Herein these sensors will be referred to as air-coupled microphone, Siemens, and Knowles. The sensors were amplified with a fixed gain using a Mackie mixer

1604-VLZ for the air-coupled microphone and two custom made pre-amplifiers based on [35] and [174] for the accelerometers. The total sensors weight including cable and coating was 6.3 g for the air-coupled microphone, 15.8 g for the Siemens, and 0.8 g for the Knowles. These sensors have been used in a large number of applications related to the skin surface measurements of lung, cardiac and voiced sounds (see for example [37,38,170,198]. Additional sensors such as a Polytec laser vibrometer (OFV3000 and OFV511), a B&K hand-held analyzer (Type 2250), a B&K microphone (Type 4191), and PCB accelerometer (A353B17 and 483BO8) were used as reference sensors at different points in this study. Two different testing schemes were required to evaluate the air-borne and tissue-borne sensitivities separately. The methods and materials required for each scheme are described in the following subsections.



(a) (b) (c)

Fig. 5.1. Bioacoustic sensors under evaluation: (a) Air-coupled microphone Sony ECM-77B, (b) Knowles accelerometer BU-7135, (c) Siemens accelerometer EMT25C.

5.1.1 Tissue-borne sensitivity experiments

The procedure to evaluate the tissue-borne sensitivity of the sensors was achieved based on previous studies [169, 170]. The notion of a Bioacoustic Transducer Tester (BATT) was further explored and used to quantify tissue-borne sensitivity. A BATT is an acoustic enclosure that only allows radiation of sound through a compliant

artificial compound (Akton of 1/4") that resembles human skin. The original BATT (herein referred to as BATT 1.0) was contrasted with two larger platforms (referred to as BATT 2.0 and 2.1). These new BATT designs are scaled versions (2:1 and 4:1 respectively) of the original BATT 1.0, with larger inner loudspeakers (3" and 6" of diameter respectively). For further details on the BATT design see [169,170]. Larger testing platforms were observed to be better suited to excite low frequencies and allowed the effect of coverings that surround the sensor (e.g., a passive ear protector covering an air-coupled microphone) to be evaluated. Note that for the amplitude of vibration used in this study, the BATTs can also radiate air-borne sound. However, the radiated sound was found to be at least 25 dB(A) below the level estimated for voiced sounds, and thus it was not expected to affect or corrupt the tissue-borne measurements.

The BATTs were excited with both voiced sounds and broadband pink noise, while the sensors were attached to the test surface of the BATTs using double sided tape (3M 2181). An amplifier (Denon PMA-920) was used to drive the loudspeaker. A pre-whitening scheme was used to compensate for the overall transfer function of each BATT. This scheme is based on established principles [172], assuring that the BATT test surface vibrates with a flat response (in a logarithmic scale since it based on pink noise) between 80 Hz and 8 kHz with a 2dB deviation between neighboring 1/3 octave bands. This approach was convenient since it was set to match the criteria later used to calibrate the frequency response of air-borne components. The pre-whitening was performed using pink noise and a 1/3 octave equalizer (Behringer DEQ2496) and a reference sensor located on the center of the test surface. The manufacturer specifications for transverse shock sensitivity of very light-weight sensors are not expected to vary when placed on a more compliant surface such as that of the BATTs (1/4" of Akton, see [169,170] for more details). Based on this principle and given its miniature size (∼6x8x2 mm), very light weight, and flat and extended frequency response (2Hz-6kHz within 5dB), the Knowles accelerometer was selected as reference sensor for this calibration. It will be shown later in this paper that this

sensor was also the least sensitive to air-borne sounds, which reaffirms its selection as reference sensor in this case. The amplitude of vibration of the BATT test platforms was adjusted to resemble as closely as possible that of the skin vibration at the sternal notch (a.k.a. jugular notch) during normal speech production of sustained vowels. This amplitude calibration was performed using the reference sensor after the aforementioned pre-whitening scheme.

All signals were filtered with an 8 pole Butterworth low-pass filter (Krohn-Hite 3384) and digitalized using a 16-bits data acquisition system (NI BNC-210) with a 96 kHz sampling frequency. Data post-processing was performed using one-third octave band analysis according to [199]. Note also that the logarithmic nature of this representation allows for a more clear appreciation of low frequency phenomena.

All measurements were performed inside a sound proof chamber model (IAC 102871). The background noise was measured according to [200] using a B&K hand-held analyzer (Type 2250) obtaining an equivalent sound pressure level (Leq Slow) of 20.9 dB(A) or equivalently 74.1 dB(Z) (the Z notation was introduced by B&K and incorporates no weighting factor between 12Hz and 20kHz).

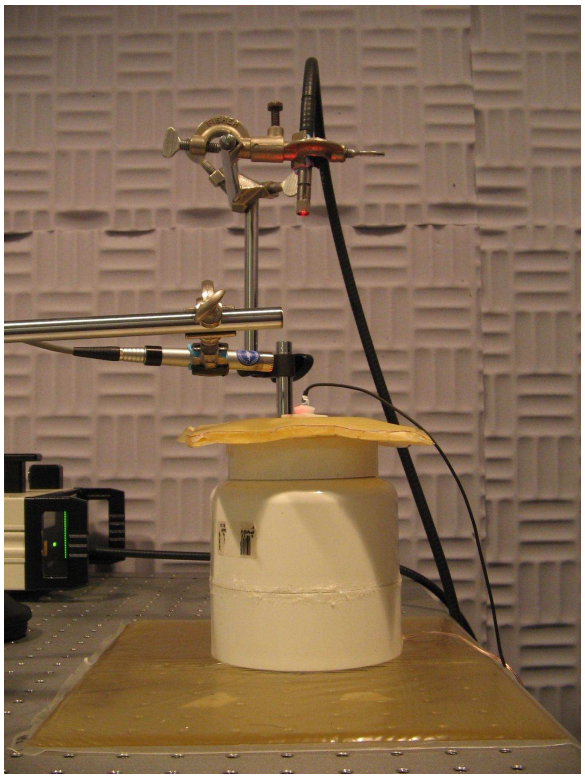


Fig. 5.2. Bioacoustic Transducer Tester (BATT) and laser vibrometer setup for the tissue-borne sensitivity experiments

5.1.2 Air-borne sensitivity experiments

The air-borne sensitivity of bioacoustic sensors is a function of the material properties of the surface where they are attached. For the case of sensors placed on human skin, the body structure and skin compliance will create a combination that is not easily reproducible by artificial skin and mounting conditions. Even though certain artificial compounds (e.g., 1/4" of Akton) have been shown to have an acoustically comparable behavior with that of human skin [169,169], the human body will provide a large and unique mounting configuration that is not well represented by the compounds themselves. Such mounting configuration is critical during the evaluation of the sensor's air-borne sensitivity since it dictates the extent of acoustic energy that is conveyed into transverse vibration of the test bed. Some artificial configurations can facilitate this transmission, thus distorting the air-borne sensitivity experiments. This phenomenon was evident in preliminary in vitro tests, illustrating the difficulties associated with the design of artificial conditions that accurately mimic the skin/body mounting conditions for the sensors. The differences between artificial and human skin mounting conditions on the air-borne sensitivity are discussed in the results section.

On the other hand, the variability given by different human subjects is expected to be small. This assumption is given by the fact that during the recordings each subject remains still in a controlled position without uttering speech and holding his/her breath at functional residual capacity. Therefore, factors such as height, weight (thus Body Mass Index or BMI) and body structure are expected to have a minor effect on the local mounting conditions of the sensor. Given the low expected variations across subjects, the air-borne sensitivity of the bioacoustic sensor was tested using five human subjects. The actual subject variability is discussed in the results section. A relatively large set of measurements was first performed using one healthy young male reference subject, and a selected set of tests were performed using the additional subjects (two healthy young males and two healthy young females). All

subjects were in the 20-30 age range with BMI observed between 23.2 and 26.2. The sensors were attached to the skin surfaces using double sided tape (3M 2181) on three selected locations on the subject's body as illustrated in Fig. 5.3. These locations are in accordance with the aforementioned applications of skin surface measurements.

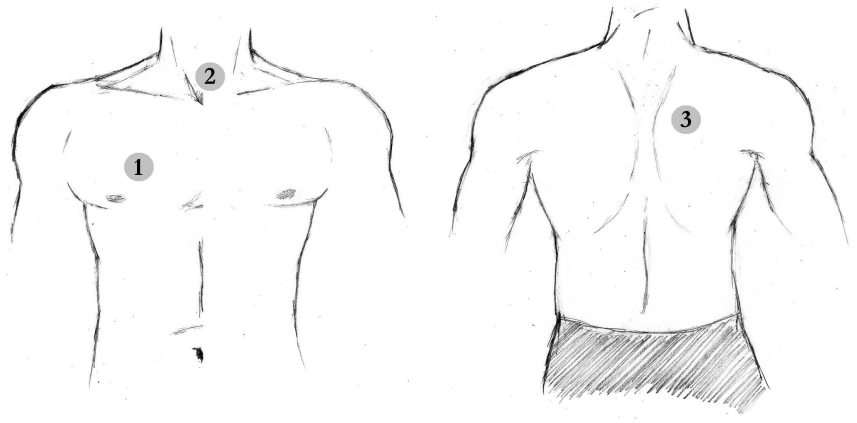


Fig. 5.3. Location of the sensors for the air-borne sensitivity tests in human subjects. (1): second intercostal space over the anterior right upper lobe of the lung, (2): the suprasternal notch, (3): posterior right upper lobe of the lung

The protocol used during these measurements is based on the principles established in [173] and [201], where a loudspeaker (Mackie HR 824) was placed 26" from the anterior chest to achieve high levels of acoustic pressure (60 dB over the noise floor). The background noise conditions were the same as previously described, for which an equivalent sound pressure level (SPL) of 100 dB(A), or equivalently 105 dB(Z), was used. Although this level provides sufficient signal-to-noise ratio (SNR) to minimize the effects of background noise, it is not associated with normal speech levels. Measurements of sound pressure levels during speech production are discussed in details in [37, 202]. These studies showed that typical SPL measurements at 30 cm from the mouth are generally 65 dB(Z) for soft voice, 75 dB(Z) for normal voice, and 85 dB(Z) for loud voice, all with an approximate $\pm$ 5 dB deviation for most of the cases. Observations made in our laboratory suggest that the SPL measured

at 30 cm from the mouth compared with near-skin locations is ∼1 dB(Z) lower for the sternal notch, ∼3 dB(Z) lower for the anterior right upper lobe, and ∼12 dB(Z) lower for the posterior left upper lobe. Previously, the amplitude calibration for the tissue-borne sensitivity was obtained when the reference subject uttered a sequence of sustained vowels at normal/loud loudness, observing a level of 86.6 dB(Z), or equivalently 81.7 dB(A), at 30 cm from the mouth. Therefore, it was considered that the SPL surrounding the sternal notch was approximately 85 dB(Z), for which the levels associated to air-borne sensitivities were adjusted to match this value. Consequently, broadband pink noise was presented at two intensities, 105 dB(Z) and 85 dB(Z), to allow a large SNR and match speech levels, respectively. In addition, the air-borne sensitivity of the sensors to voiced sounds was evaluated with a calibrated level of 85 dB(Z) during the protocol.

The coloring effect of the room resonances was compensated for using a pre-whitening scheme. This compensation was performed using the 1/3 octave equalizer (Behringer DEQ2496) and a near-skin microphone as reference. Such near-skin microphone was selected to assure that the calibration was performed as closely as possible to the sensor location. For this purpose, a Sony ECM-77B was attached to the skin using double sided tape and a small plastic stand, in order to make it face the loudspeaker and keep a distance of 1" from the skin. This approach was based on [173, 200] and assures that the sound field surrounding the skin sensors have a flat response between 80 Hz and 8 kHz 1/3 octave bands with a 2dB deviation between neighboring bands. This procedure was repeated for each subject to account for possible variations of the sound field due to individual body differences. The signal conditioning, digitalization, post processing and analysis was performed as described in the previous section.

5.1.3 Relationship between sensitivities

The air-borne and tissue-borne sensitivities of biacoustic sensors are functions of the specific gains that the corresponding preamplifiers provide. In many cases, these gains are variable and defined by the user. However, the relationship between these two types of sensitivities remains constant independently of the gain. The only non-linearity in this relationship is a result of the noise floor and distortion point of each sensor (normally referred to as the headroom). To generalize the results of this study, the concept of tissue-to-air ratio (TAR) is introduced. This relationship is considered similar to a signal-to-noise ratio (SNR), where the "noise" term is associated to the air-borne component and the "signal" is associated with the tissue-borne one. The TAR can be represented as a frequency dependent plot given by the difference (in dB) between the tissue-borne sensitivity and the air-borne sensitivity. In this study, the TAR is primarily associated with sensitivities to voiced sounds at the sternal notch, i.e., the amplitudes of skin vibration and sound pressure were set to resemble those observed during speech at that location. The translation of these sensitivity and TAR curves for other applications (e.g., lung sounds) is discussed in the application notes section.

In addition, several variables that could affect the tissue-borne and air-borne sensitivities of the sensors were investigated. These variables include the effects of subject variability, sensor location on the body, sensor attachment, passive covering of a sensor, and applied pressure on the skin surrounding the sensor. For simplicity, each passive protector or cover (e.g., earmuffs, PVC caps) is referred to as a BioAcoustic Insulator (BAI). From an initial set of ten BAIs, three were selected via initial screening as having the most attenuation. These were two earmuffs with ratings NRR 30 and NRR 33, and a 3" diameter PVC cap. These BAIs were modified by adhering additional circumferential cushioning and acoustic filling material (acousta-stuff). These conditions showed the highest performance for the selected BAIs. Each BAI

was evaluated using three different cover attachment forces: No force, medium force (∼10N), and strong force (∼30N).

5.2 Results

The tissue-borne and air-borne sensitivities are presented in solid and dotted lines respectively for each sensor in Fig. 5.4-a. The results for air-borne sensitivity were obtained based on the mean values of all subjects measured at position 1 in Fig. 5.3. The difference between the two sensitivity curves is the TAR relation, which is presented for each sensor in Fig. 5.4-b. Based on these results, the Knowles accelerometer shows a higher TAR curve, only superseded by the Siemens accelerometer for frequencies below 300 Hz. The air-coupled microphone was shown to have the smallest TAR curve, i.e., it is easily corrupted by air-borne transmitted signals when attached to the skin. This is a result of the high sensitivity of the air-coupled microphone to both tissue-borne and air-borne sounds.

Comparisons between tissue-borne sensitivity and transverse vibration sensitivity (included in the sensor technical specifications) are possible, although deviations are expected given the differences between the testing conditions. The primary difference between the two types of sensitivities is observed in low frequencies, where the compliance of the artificial compound can alter the response of the sensor. This is more noticeable for more massive sensors, such as the Siemens accelerometer, where a stronger response below 250 Hz is observed in the tissue-borne sensitivity. Higher frequencies show the same essential behavior in the response of the sensor (4 dB variations up to 1.25 kHz, where it shows a -15dB per octave roll-off). Variations between tissue-borne and transverse sensitivity are expected to be less significant for the Knowles accelerometer due to its relatively light weight. No comparisons of this type can be made for the air-coupled microphone, since no specifications of transverse sensitivity are available for this sensor. In addition, no comparisons with previous studies or technical datasheets can be performed for the air-borne sensitivity (doted

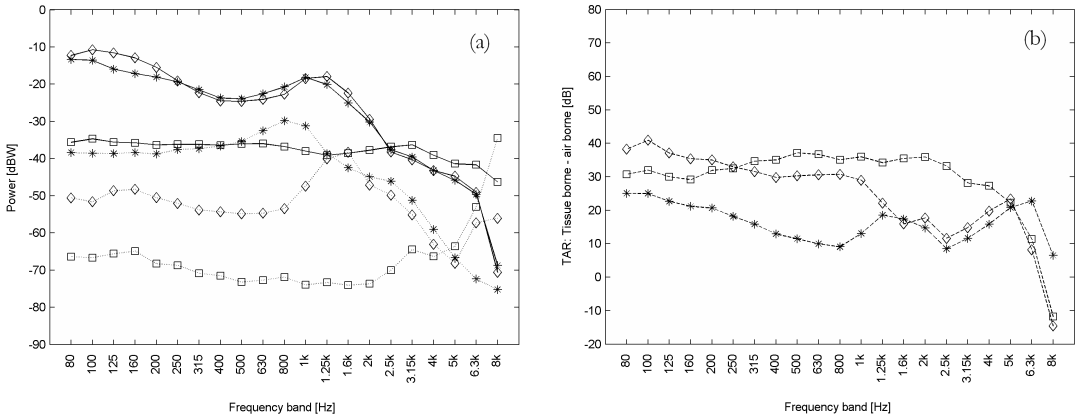


Fig. 5.4. Tissue-borne and air-borne sensitivities of each sensor using pink noise source. Mean air-borne sensitivity values observed on position 1 in Fig. 5.3. (a) Tissue-borne and air-borne sensitivities are represented by solid lines and dotted lines, respectively. Air-coupled microphone: $*$, Siemens EMT25C: $\diamond$, Knowles BU-7135: $\square$. (b) Tissue-borne to Air-borne Ratio (TAR) for each sensor based on se sensitivity curves. Air-coupled microphone: $*$, Siemens EMT25C: $\diamond$, Knowles BU-7135: $\square$.

lines in Fig. 5.4-a) of the evaluated sensors since this type of response is typically neglected or not provided.

On the other hand, previous studies [169,170] reported the tissue-borne sensitivity of the air-coupled microphone and Siemens accelerometer (same make and model). The tissue-borne sensitivities plotted as solid lines in Fig. 5.4-a are in agreement with the reported data, showing only minor differences in the relative amplitudes of the peaks around 125-200 Hz and 1000-1250Hz. These differences are expected to be given by the different calibration schemes. Not incorporated in previous studies, the compensation for the BATT transfer function on the surface allows a flat broadband excitation on the Akton. Therefore, it is believed that the current results minimize non-desired effects introduced by the testing platform and better describe the tissue-borne response of the sensors.

5.2.1 Acoustic protection of sensors with BAIs

Results showing the air-borne sensitivity of each sensor after protecting them with the BAIs are presented in Fig. 5.5. The ability of the BAIs to reduce their air-borne sensitivity depends upon each sensor. For the air-coupled microphone the BAIs exhibit the best results, being able to significantly reduce its effective air-borne sensitivity. Among the three BAIs, the PVC 3" cap yielded the most noise protection. It should be noted that for both accelerometers, the BAIs can either reduce or amplify particular frequency components. This observation makes the BAIs generally less suitable for air-borne protection in accelerometers. However, if low frequency amplification is desired or at least acceptable for a particular application, their usage could be justified.

The presence of a BAI could potentially affect the tissue-borne sensitivity of the sensors. Tests performed in the BATT 2.0 and 2.1 showed that for accelerometers (Siemens and Knowles) the presence of the BAIs did not affect the measured spectra, whereas for the air-coupled microphone a reduction in the sensitivity of low frequencies was observed (-10dB below 315Hz). This effect was proportional to the applied force. Contrasting these observations, measurements of "background body noise" on human subjects (i.e., heart sounds) were amplified at low frequencies due to the presence of the BAIs, with this effect again proportional with the applied force for all sensors. This is illustrated in Fig. 5.6, where the presence of the 3" diameter PVC cap with medium force raised the noise floor of the air-coupled microphone, as noted in the average observed in position 1 from Fig. 5.3. These differences observed between BATTs and human subjects are attributed to the fact that the BATTs do not represent the changes observed in the chest wall when force is applied. In other words, the BATTs do not account for the musculoskeletal structures that will affect the transmission of body sound when pressure is applied in the skin surface.

In all, the changes introduced by the BAIs were shown to be favorable only for the air-coupled microphone. Therefore, sensitivity curves and TARs for this sensor were

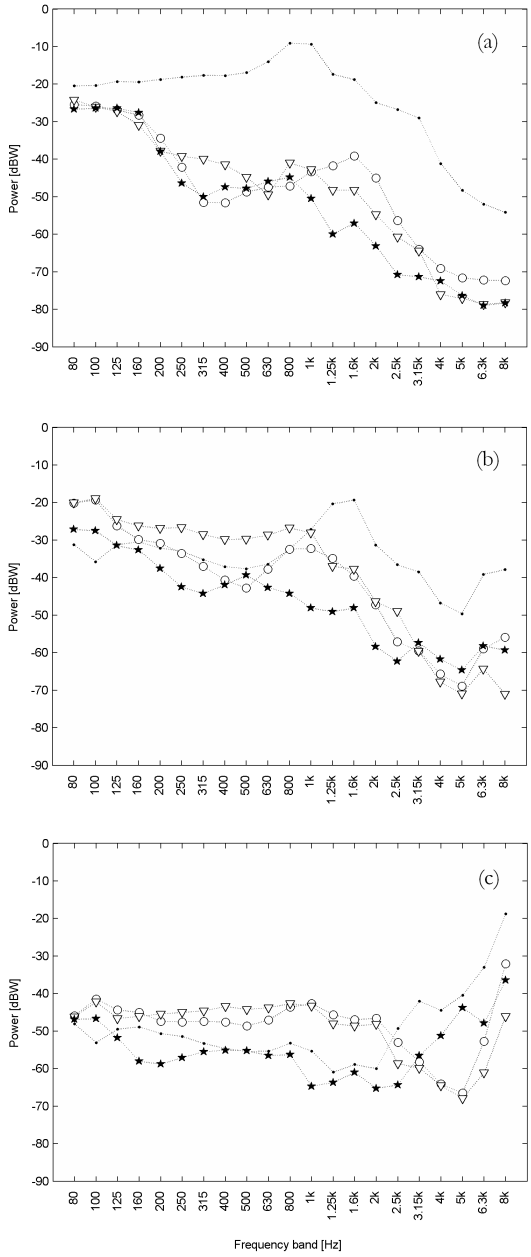


Fig. 5.5. Effect on the air-borne sensitivity of different BAIs applied with medium force for each sensor: a) Air-coupled microphone, b) Siemens EMT25C, c) Knowles BU-7135. No protection: •, NRR30: ∘, NRR33: ∇, PVC30: ⋆

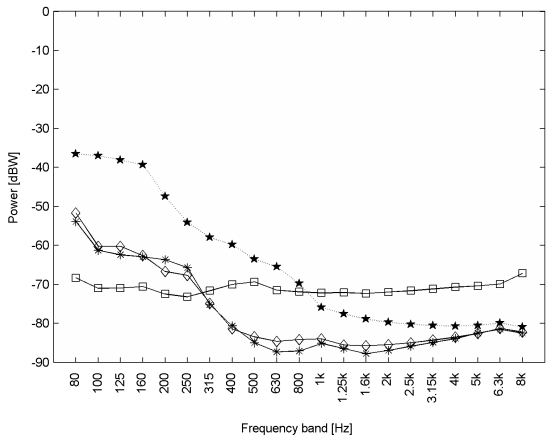


Fig. 5.6. Mean noise floor observed on position 1 in Fig. 5.3: Air-coupled microphone, protected with a PVC 3" BAI: (dotted line) $\star$, and with no BAI protection: (solid line) *. Siemens accelerometer with no BAI protection: (solid line) $\diamond$, and Knowles accelerometer with no BAI protection: (solid line) $\square$.

recomputed for the best observed conditions (PVC 3" protected) and are presented in Fig. 5.7. Note that the increased noise floor in low frequencies hampers the TAR curves in low frequency. However, a significantly enhanced TAR curve is observed for frequencies above 315 Hz compared with those for the air-coupled microphone in Fig. 2.

5.2.2 Effect of sensor placement and mounting conditions

To evaluate the dependence of the air-borne sensitivity of the sensors upon the surface type where they are attached, such sensitivity was also measured when the sensors were placed on non-tissue surfaces. This is shown for each sensor on Fig. 5.9, where three test surfaces were used: Akton of 1/4", steel block of 6", and no surface (free field). The air-borne sensitivity for these test surfaces is contrasted with that of the chest position for convenience. It can be observed that for the air-borne sensitivity, no surface behaves similar to human skin for all sensors. This effect is attributed to the

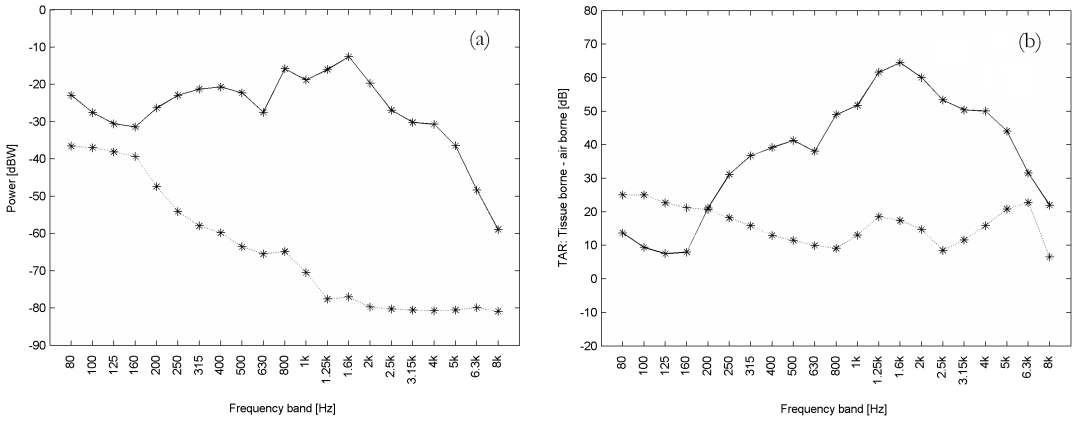


Fig. 5.7. Changes in the response of the air-coupled microphone due to the PVC 3" BAI applied with medium force: (a) Tissue-borne and air-borne sensitivities measured using pink noise. Tissue-borne and air-borne sensitivities are represented by a solid line and a dotted line, respectively. (b) TAR curves. The new and original TARs are represented by a solid line and a dotted line, respectively.

Fig. 5.8. Experimental setup for *in vitro* air-borne sensitivity experiments

unique mounting conditions of the sensors when placed on human skin. Therefore, the tests of air-borne sensitivity made on human subjects are fundamental for the purpose of this study. The variation in sensitivity when the sensors were placed on the anterior chest (position 1 in Fig. 5.3), the sternal notch (position 2 in Fig. 5.3) and posterior RUL (position 3 in Fig. 5.3) was evaluated. The possible changes in the sound field between locations were also taken into account in this analysis. The differences between the body structure and tissue properties at the different locations were found to have small variability (<3dB) for all frequency bands of interest, excepting the 2.5 kHz and 3.15 kHz bands, where a slightly larger deviation was observed (∼5dB). The main factor affecting the dispersion of the data is the subject posture, which explains the larger difference at mid-high frequencies. In all, this response illustrates that differences in skin mounting conditions, such as subcutaneous fat thickness, have only a minor effect on the air-borne sensitivity of the sensors. The fact that the air-borne sensitivity remains relatively constant for different body location is convenient for future studies since the observations made for each sensor in this study can be extrapolated to different body locations.

The variability between subjects was also evaluated. Factors that primarily affect the dispersion of the data are subject posture and body configuration. The air-borne sensitivity across subjects showed deviations smaller than 3 dB for all the frequency bands. Thus, the general air-borne sensitivity trends of the sensors were observed to be consistent across subjects.

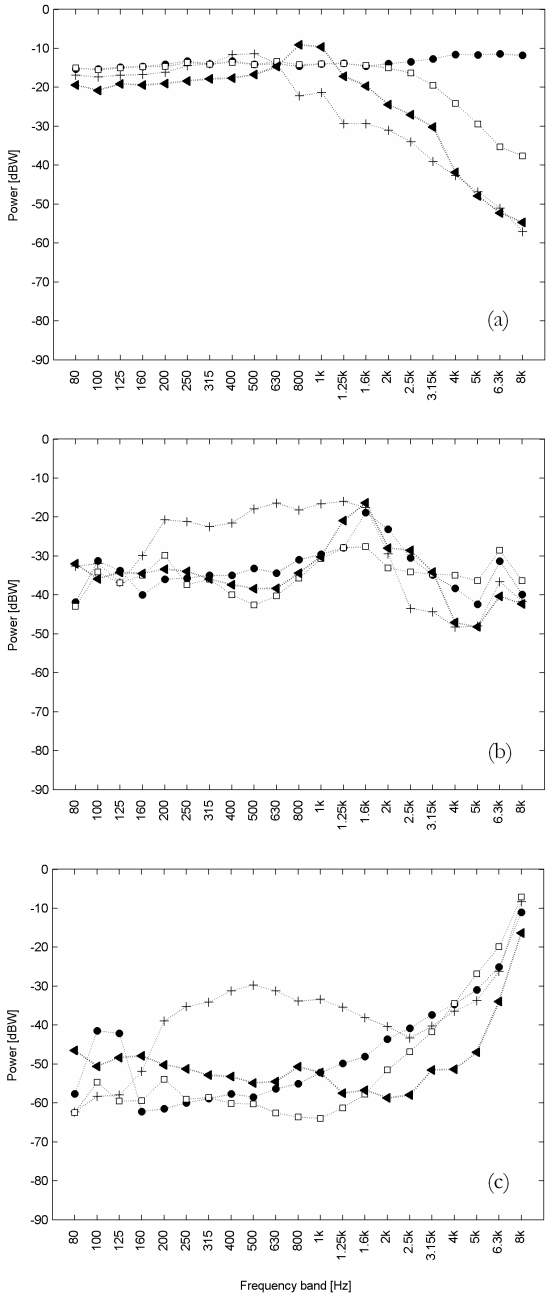
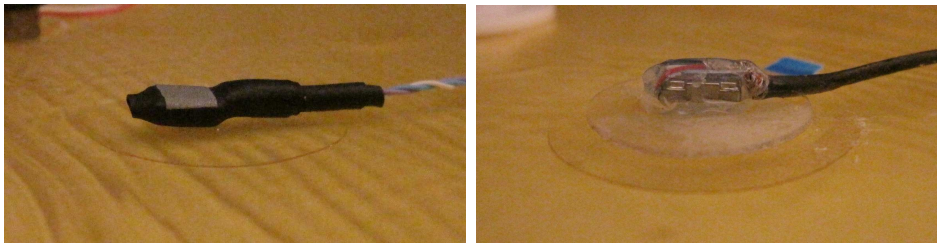


Fig. 5.9. Effect of different surfaces on the air-borne sensitivity of: a) Air-coupled microphone, b) Siemens EMT25C, c) Knowles BU-7135. Akton: +, Steel: □, Posterior Chest (RUL): ◀, Free field: •

5.3 Application notes

5.3.1 Knowles accelerometer mounting

The results from previous sections evaluated the air-borne and tissue-borne sensitivities of the Knowles BU-7135 accelerometer covered with a plastic coating. This configuration (referred to as BARLAB configuration) was contrasted with a different mounting condition (referred to as MGH configuration), where the accelerometer was covered with a thicker layer of silicone and attached to a silicone disk. This latter configuration is the one currently available for the ambulatory monitoring system from KayPENTAX and described in previous studies [35, 36, 38]. The difference between these two configuration can be appreciated from Fig. 5.10.



(a) (b)

Fig. 5.10. Bioacoustic sensors under evaluation: (a) BARLAB configuration, (b) MGH configuration

The difference of air-borne and tissue-borne sensitivities between the two configurations are shown in Fig. 5.11 and indicated that these mounting conditions did affect their response. The MGH configuration was less sensitive to air-borne transmitted sounds and more sensitive to tissue-borne components. These relations were frequency-dependent, having more than 6 dB of difference in certain frequency bands. These variations were associated to the thicker layer of silicone underneath the MGH configuration, which attenuated the air-borne pathway. On the other hand, the disk appeared to enhance the sensor attachment and increase its effective surface, thus

favoring the tissue-borne path between 500 Hz and 3 kHz. This variability product of the mounting conditions is an important design feature that should receive attention to optimize the sensor performance.

In addition, it was observed that the mounting conditions affected the DC output of the sensor, most likely due to differences in static pressure surrounding the sensor between configurations. However, since this DC component is removed from the accelerometer, such a variation was not further explored.

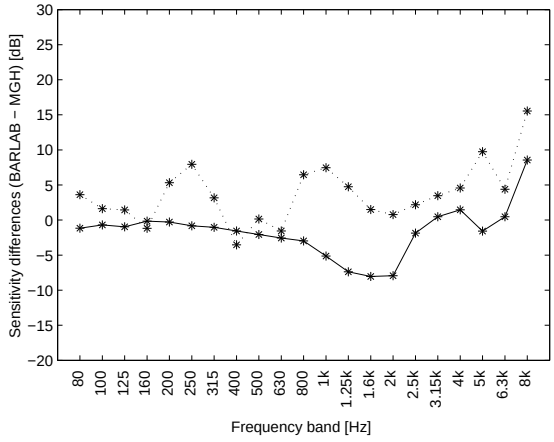


Fig. 5.11. Differences between air-borne (dotted line) and tissue-borne (solid line) sensitivities for the BARLAB and MGH configurations for the Knowles accelerometer.

5.3.2 Using the sensitivities to discriminate components

In order to illustrate how the air-borne and tissue-borne sensitivities can be used to discriminate if the signal detected by a bioacoustic sensor is primarily a product of tissue or air-borne components, the following example was considered. A recording of voiced sounds on the skin surface (sustained vowel /a/) over the anterior right lobe of the lung was performed using the air-coupled microphone protected with the 3" PVC cap with a medium force holding it against the skin. The one third octave analysis of this recorded signal is presented in Fig. 5.12 along with the air-borne

sensitivity of the sensor for this condition. It can be observed from this figure that for frequencies below 1.6 kHz the signal is at least 10 dB stronger than the air-borne sensitivity. This implies that the sensor was primarily detecting tissue-borne sounds in that range. To further investigate this phenomenon, an estimate of the expected tissue-borne component is included in Fig. 5.12. This estimation was performed using the calibrated BATT excited with a sustained vowel, reducing the amplitude of vibration to mimic chest vibration instead of that of the sternal notch. The resulting estimated signal is sufficiently close to the actual measured signal up to 1.6 kHz, where the effect of the noise floor becomes significant. This illustrates that the air-borne sensitivity curve would represent a threshold that allows the discrimination of components. Any response of the sensor above this curve would be largely due to tissue-borne components. The tissue-borne sensitivity of the sensor could be used as shown a priori to create estimates of tissue-borne components that can be contrasted with the actual measured signal.

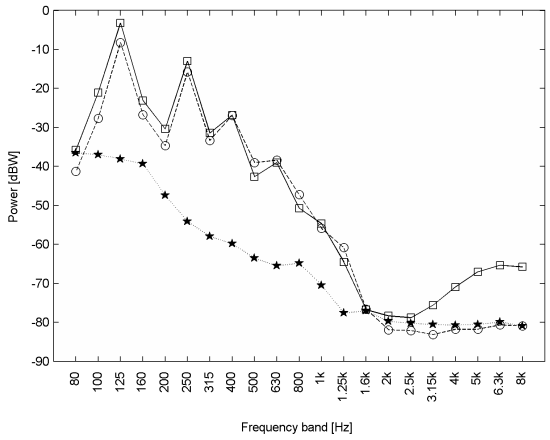


Fig. 5.12. Application of air-borne and tissue-borne sensitivities during the discrimination between components. Sensor: Air-coupled microphone protected with PVC 3" on the chest. Measured signal during vowel /a/: □, Estimated signal for vowel /a/ for tissue-borne components only: ○, Air-borne sensitivity for such condition: ⋆.

5.3.3 Translating the sensitivities and TAR curves for other applications

The example presented in the previous section introduces the idea of translation of the sensitivity curves. This translation is possible, as long as it considers the overlapping effect of noise floors that will bound the translation. The sensitivity curves from Fig. 5.4-a and Fig. 5.9-a could be translated (i.e., shifted up or down by a certain offset). The noise floor that needs to be considered for each sensor is presented in Fig. 5.6. In order to achieve this translation, it is necessary to know the offset in amplitude of vibration and sound level pressure between the conditions presented in this study and those of interest to the reader. This requires comparing both the amplitude of vibration between the sternal notch and the new location and condition of interest. At the same time, it is necessary to compare the SPL surrounding the sternal notch (estimated at 85 dB(Z)) with that surrounding the new location and condition of interest. These comparisons will yield the offset needed to translate both sensitivities curves, and a new TAR curve could be computed using these curves. Alternatively, it is possible to repeat the set of procedures performed in this study, adjusting the levels of amplitude for air-borne and tissue-borne excitation to describe a different desired condition.

5.4 Conclusions

The results of this study indicate that air-borne components can clearly affect skin surface measurements of body sounds, and that the selection and understanding of the sensor behavior for the given recording conditions becomes critical. Air-borne corruption of skin surface measurements of voiced sounds was observed to be substantial. The proposed sensitivity and TAR curves aid the selection of bioacoustic sensors for different audible acoustic applications involving skin surface measurements. However, the construction of these curves depends on the desired application. Although the sensitivities presented in this study are based on skin radiation of voiced sounds at the sternal notch, the proposed methods could serve as a guide for other appli-

cations. The gain of a bioacoustic sensor plays a significant role compromising the absolute tissue-borne and air-borne sensitivities, but does not affect the TAR. The headroom (noise floor vs. distortion point) is the limiting factor of the linearity of this relation. The selection of the sensors is also highly related to the application. Based on the sensitivity and TAR curves with no BAI protection, the best sensor appears to be the Knowles accelerometer, particularly when used in in the MGH configuration. However, this sensor is less sensitive to tissue-borne sounds, for which it is better suited to relatively high amplitude vibration cases (e.g., recordings of voiced sounds at the sternal notch). The TAR curve for the air-coupled microphone is significantly enhanced using BAIs, particularly in mid-high frequencies. However, handling the air-coupled sensor along with a BAI can be more cumbersome than the other sensors and it is not suitable for all body locations. For low frequency applications (below 200 Hz) the best performance was achieved by the Siemens accelerometer.

6. COUPLED AND UNCOUPLED IMPEDANCE-BASED INVERSE FILTERING FROM ORAL AIRFLOW AND NECK SURFACE ACCELERATION

This chapter describes the methods and results of investigations on the effect of acoustic coupling on inverse filtering of oral airflow and neck surface acceleration. Coupling between the subglottal, supraglottal, and glottal systems was investigated for complete and incomplete glottal closure using synthetic speech and human subject recordings. The proposed methods in this chapter make use of several findings from previous research components. Observations using synthetic speech from chapter 3 are contrasted with human subject recordings for incomplete glottal closure configurations for further understanding of inverse filtering methods and acoustic coupling. Observations of *in vivo* vocal fold tissue motion in chapter 4 supported the use of impedance representations to describe source-filter interactions. Such representations are the foundation of the inverse filtering scheme proposed in this chapter. Finally, the bioacoustic sensor exhibiting the most adequate behavior for speech recordings on the skin underlying the suprasternal notch from chapter 5 was selected for the proposed inverse filtering from neck surface acceleration.

6.1 Methods

6.1.1 Impedance-based inverse filtering: general considerations

A model-based inverse filtering scheme that accounts for acoustic coupling between tracts and source-filter interactions was proposed, implemented, and evaluated. A coupled dipole model as shown in Fig. 6.1 is the foundation for the desired inverse filtering scheme, given its direct application to subglottal coupling [11,91]. This simple

but powerful model has four basic components: an ideal airflow source (U_o) and three different lumped terms representing the subglottal (Z_{sub}), glottal (Z_g), and supraglottal (Z_{supra}) acoustic impedances. Inverse filtering based on this representation is referred to as "impedance-based inverse filtering" (IBIF).

Each of the system impedances was estimated from experimental data to obtain subject-specific values. In order to estimate the tract impedances, models of acoustic transmission were applied. This approach provided further insights into the system behavior when compared to a perturbation (black box) model [35]. The proposed transmission model was based on a series of concatenated T-equivalent segments of lumped acoustic elements that related acoustic pressure to volume velocity. A representation of this type of section is depicted in Fig. 6.2. Yielding walls included cartilage and soft tissue for the subglottal system. A radiation impedance was used to account for skin neck properties, as well as accelerometer loading. The remaining lumped acoustic elements in Fig. 6.2 described the standard acoustical representations for losses, elasticity, and inertia.

The transmission line models for the subglottal and supraglottal tracts yielded the driving point impedances as well as transfer functions for any desired location within the tracts. These terms only depended on the tract configuration and its inherent physical properties. Thus, the transfer function between the airflow at the mouth (U_m) and the airflow entering the vocal tract (U_{supra}) was given by

$$T_u = \frac{U_m}{U_{supra}}, \tag{6.1}$$

where T_u was a linear representation of the vocal tract transfer function that only depended on the vocal tract (i.e., uncoupled from the glottis and subglottal impedances) and was computed using the tract geometry through the aforementioned transmission line scheme. Estimation of the airflow entering the vocal tract required inverting this transfer function (i.e., $U_{supra} = U_m/T_u$), yielding airflow estimates that were re-

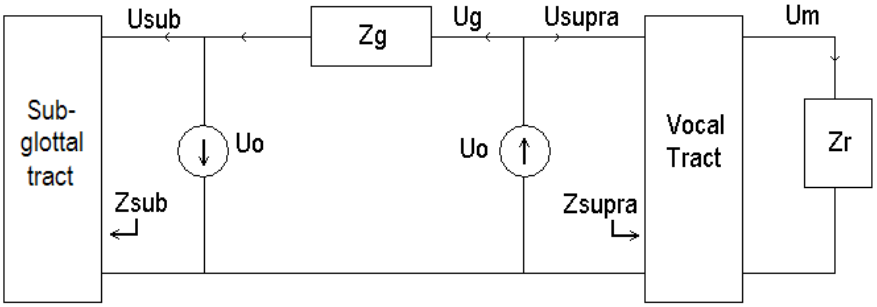


Fig. 6.1. Representation of a dipole source using two ideal airflow sources. Coupling between the subglottal and supraglottal tracts was obtained through a linearized glottal impedance. The glottal impedance in this representation was investigated as a time-invariant and also a time-varying term.

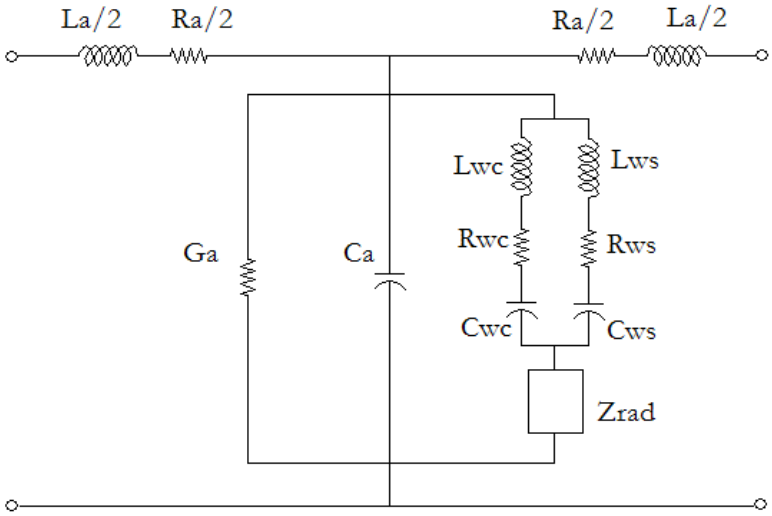


Fig. 6.2. Representation of the T network used for the subglottal and supraglottal models. The acoustic elements L_a, R_a, G_a, and C_a were associated with air inertance, air viscous and heat conduction resistances, and air compliance. Yielding walls parameters were divided in soft tissue (L_{ws}, R_{ws}, C_{ws}) and cartilage (L_{wc}, R_{wc}, C_{wc}) whenever needed for the subglottal tract. The radiation impedance Z_{rad} was attributed to the accelerometer loading.

ferred to as "uncoupled" glottal airflow. This process was performed in the frequency domain by means of the fast Fourier transform (FFT) and its inverse.

When coupling through the glottis was considered, a coupled transfer function between the oral airflow (U_m) and the ideal glottal airflow (U_o) was defined as:

$$T_c = \frac{U_m}{U_o} = \frac{U_m}{U_{supra}} \cdot \frac{U_{supra}}{U_o} = T_u \cdot T_p, \tag{6.2}$$

where

$$T_p = \frac{U_{supra}}{U_o}, \tag{6.3}$$

was a perturbation transfer function obtained by solving the model circuit in Fig. 6.1 via the superposition theorem which yielded:

$$T_p = \frac{U_{supra}}{U_o} = \frac{Z_g}{Z_g + Z_{sub} + Z_{supra}}. \tag{6.4}$$

This decomposition was originally proposed by [91]. An illustration of how the glottal coupling affected the transfer functions is shown in Fig. 6.3.

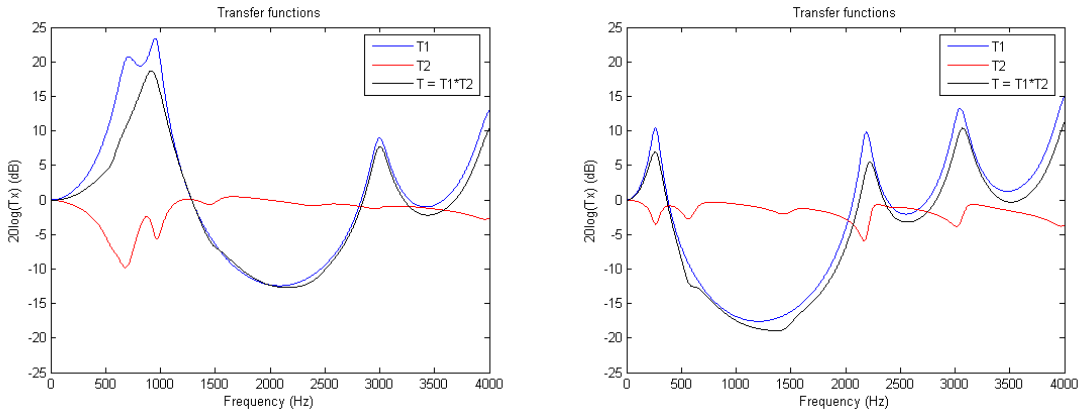
In addition, other relations that were derived from the model circuit in Fig. 6.1 include:

$$U_{sub} = -U_{supra}, \tag{6.5}$$

and

$$U_g = U_o \cdot \frac{Z_{sub} + Z_{supra}}{Z_g + Z_{sub} + Z_{supra}}. \tag{6.6}$$

In order to account for the acceleration signal measured on the neck surface overlying the sternal notch, the circuit in Fig. 6.1 was expanded, resulting in the representation of in Fig. 6.5. The subglottal impedance illustrated in Fig. 6.1 was decomposed into an initial subglottal section Sub_1 that represented the portion of the



(a) Vowel /a/ (b) Vowel /i/

Fig. 6.3. Uncoupled (T1), perturbation function (T2), and coupled vocal tract transfer function (T) from oral airflow to the glottis for a vowel /a/ and /i/.

extrathoracic trachea above the accelerometer. Similarly, Sub_2 and its driving input impedance Z_{sub2} represented the subglottal system below the accelerometer. Consequently, the section where the accelerometer was placed was represented by the contracted T-network between these two. The volume velocity U_{skin} flowing through Z_{skin} was expressed as:

$$U_{skin} = U_{sub1}\frac{Z_{sub2}}{Z_{sub2} + Z_{skin}}, \tag{6.7}$$

where Z_{skin} was constituted by the mechanical impedance of the skin Z_m in series with the radiation impedance Z_{rad} constituted by the accelerometer loading, thus

$$Z_{skin} = Z_m + Z_{rad}, \tag{6.8}$$

$$Z_m = R_m + j(\omega M_m - \frac{K_m}{\omega}), \tag{6.9}$$

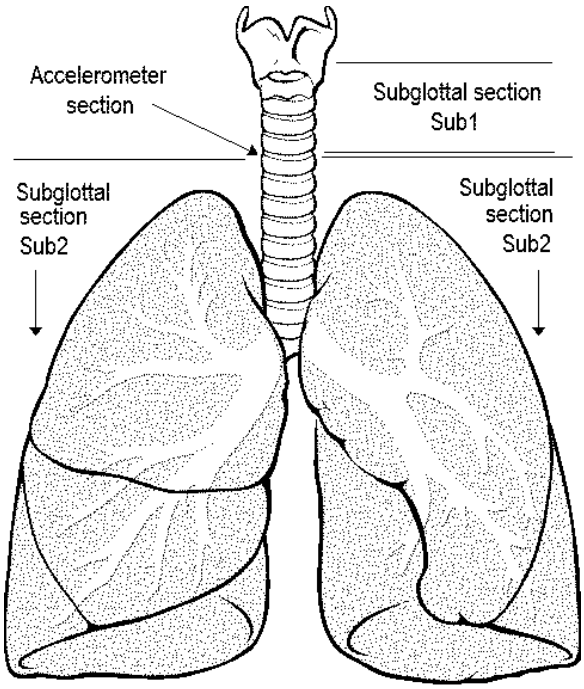


Fig. 6.4. Representation of the subglottal system. The accelerometer is placed on the surface overlying the suprasternal notch at approximately 5 cm below the glottis. The tract above and below such location are defined as Sub_1 and Sub_2. Figure adapted from [203].

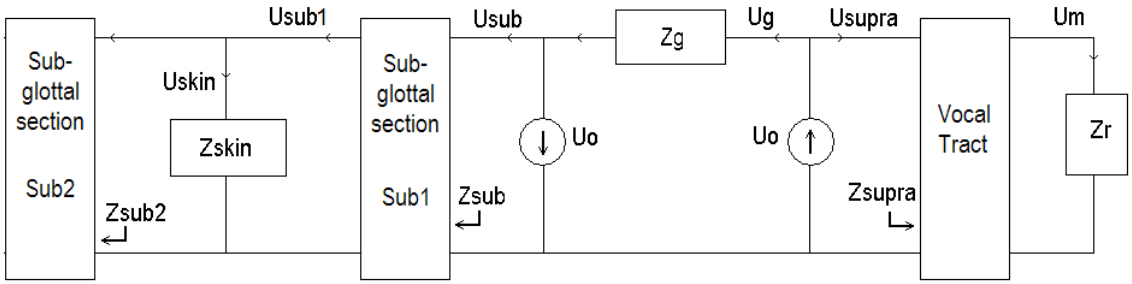


Fig. 6.5. Representation of a dipole model using two ideal airflow sources. The subglottal impedance from Fig. 6.1 was decomposed to include skin neck acceleration. See Fig.6.4 for details.

and

$$Z_{rad} = \frac{j\omega M_{acc}}{A_{acc}}. \tag{6.10}$$

The skin volume velocity was differentiated to obtain the neck surface acceleration signal $\dot{U}_{skin}$. Therefore, the uncoupled transfer function between the supraglottal volume velocity and the acceleration signal, referred to as T_{uskin}, was expressed as:

$$T_{uskin} = \frac{U_{supra}}{\dot{U}_{skin}} = -\frac{H_{sub1} \cdot Z_{sub2}}{Z_{sub2} + Z_{skin}} \cdot H_d, \tag{6.11}$$

where H_{sub1} was the flow-flow transfer function of the subglottal section Sub_1 from the glottis to the acceleration location and $H_d = j\omega$ is the ideal derivative filter. Similarly, the coupled transfer function between the ideal airflow source and the acceleration signal, referred to as T_{cskin}, was obtained by:

$$T_{cskin} = \frac{U_o}{\dot{U}_{skin}} = T_{uskin} \cdot T_p, \tag{6.12}$$

where T_p is the same perturbation transfer function that introduced all system coupling for the vocal tract, as defined in equation (6.4). The transfer functions T_{uskin} and T_{cskin} were used to estimate the ideal, supraglottal, and oral airflows directly from the accelerometer signal.

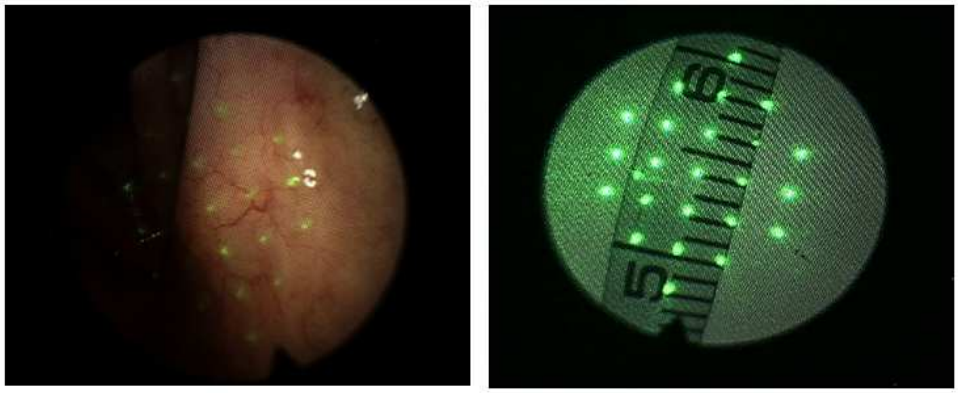
6.1.2 Experimental setup

In order to evaluate the impedance-based inverse filtering scheme, a comprehensive set of measurements was considered. The goal was to obtain estimates of the system behavior through simultaneous recordings of glottal behavior, flow aerodynamics, and acoustic pressures. Thus, the experimental setup corresponded to that of chapter 4 and considered synchronous measurements of skin surface acceleration (ACC), high-speed laryngeal videoendoscopy (HSV), oral volume velocity (OVV), electroglottog-

raphy (EGG), and radiated acoustic pressure (MIC). However, the protocol for these experiments was much simpler (e.g., only sustained vowels were recorded) and a complete calibration of the data was used. Two subjects (male with no vocal training and female with vocal training) completed the required (calibrated) recording sessions for different vowels (/a/ and /i/) and glottal conditions (breathy, chest, falsetto). These subjects had no history of vocal pathologies, were in the 25-35 age range, and are referred to as "calibrated subjects" in this chapter.

Glottal area extraction from laryngeal high-speed videoendoscopy was computed using digital image processing, i.e., segmentation and region merging schemes. The segmentation yielded a glottal area function in squared pixels that was calibrated to obtain the actual glottal area in absolute units (e.g., square centimeters). The HSV calibration scheme was performed using a laser grid that allowed for quantification of features within the endoscopic view [190, 204] and dedicated software developed for this purpose (Endoview). A constant "feature" that repeatedly appeared in selected high-speed videos (e.g., blood vessel on the ventricular folds) was carefully measured in a separate session using videostroboscopy and the laser grid for the participants of the experiments. This "feature" was used as a reference to calibrate each HSV recording, where the optic and glottal conditions could change but the dimensions of the "feature" were assumed to remain constant. Illustrations of the laser grid and calibration feature used for one of the subjects are shown in Fig. 6.6 and Fig. 6.7.

In addition, calibration of the OVV signal was performed by airflow calibration unit (Model MCU-4, Glottal Enterprises) after each recording session. The accelerometer was calibrated using a laser vibrometer as described in [35]. Calibration of the acoustic signal was performed after each recording session by comparing side-by-side recordings of a stable wideband reference tone generator (COOPER-RAND, Luminaud, Inc., Mentor, OH) with the MIC signal and a Class-2 sound level meter (Model NL-20, RION Co., Ltd., Tokyo, Japan) set to linear "C" weighting and "Fast" response time. No calibration of the EGG was undertaken in these experiments.



(a) (b)

Fig. 6.6. Laser grid used for HSV calibration. (a) Calibration feature using a U-shaped blood vessel on the left ventricular fold (b) Same reference laser grid on a metric scale.

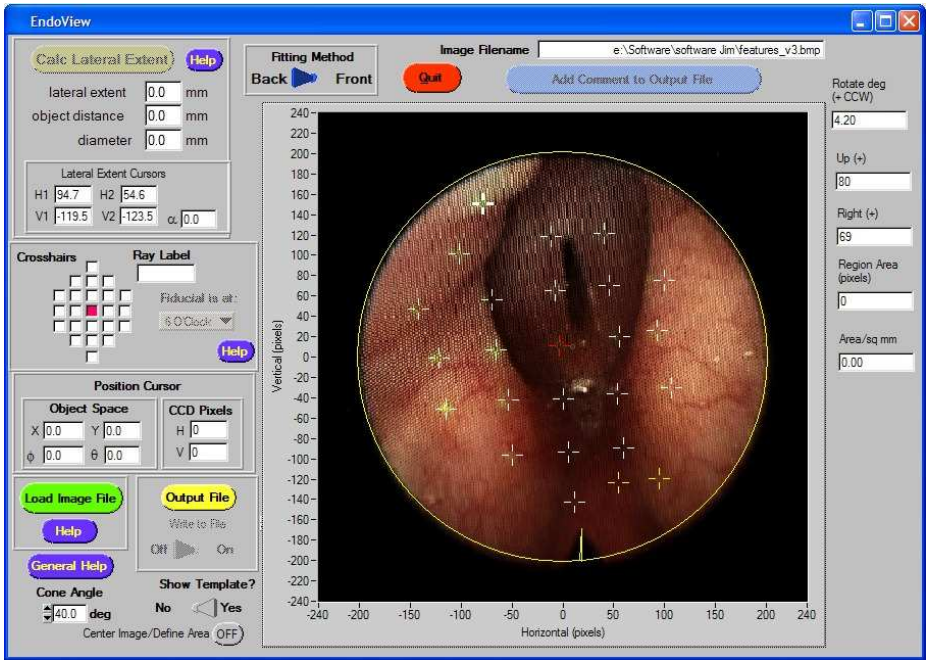


Fig. 6.7. Endoview software used for the laser grid calibration scheme as in [204]. The blood vessel used as calibration feature is observed on the left ventricular fold.

All recordings were obtained in an acoustically treated room at the Laryngeal Surgery & Voice Rehabilitation Center at the Massachusetts General Hospital.

6.1.3 Estimation the tract acoustic impedances

In order to estimate the supraglottal impedance, estimates of the vocal tract area functions were obtained from the oral airflow recordings. The method selected for this purpose was based on a systematic variation of high-resolution, MRI-based vocal tract area functions obtained from a recent imaging study [22]. The recursive algorithm described in [205], where an area function was modified to match target formant frequencies was used. Thus, the area functions were extracted from a selected database [22] and modified to match the formant frequencies of the utterance from the specific subject. Initiating the process with the same vowel was shown to increase the efficiency and minimized the likelihood of obtaining unrealistic tract shapes. The target formant frequencies were selected from poles obtained from estimates of the vocal tract transfer function obtained via CPIF. Poles were selected within the expected range of formant frequencies, thus discarding other features captured by the covariance linear prediction scheme. Estimates of the closed portion boundaries were obtained using the EGG signal as described in [139, 140]. This method provided higher resolution and more reliable area functions than those obtained directly from the linear prediction coefficients [88,206,207]. Even though it has been suggested that the transformation from a set of formant frequencies to a vocal tract area function may not produce a unique solution [61,206,207], this is not expected to be a problem in these experiments due to the simple and common nature of the sustained vowel gestures under study.

Estimates of subglottal impedance were computed in a similar fashion. An acoustical transmission line model of a symmetric branching subglottal representation from previous studies [68, 69] was used. Symmetric anatomical descriptions as those presented by [70] for an average male were found to closely resemble those reported

experimentally [64] and thus were used as default anatomical descriptions. As before, the model was adapted to match subject specific parameters. This was achieved by adjusting the model to match the first subglottal resonance measured from the accelerometer signal. In particular, the length of the trachea, considered to be the main anatomical difference between subjects in the lower airways [208], was modified to produce the observed resonance. A similar scaling method was used to correct anatomical differences when modeling subglottal impedance [64]. The first accelerometer resonance was obtained via the covariance method of linear prediction during the closed phase of the cycle. Even though it is known that this method fails to describe the zeros from the subglottal impedance, preliminary testing with human data and synthetic speech showed that it was sufficiently accurate and stable to estimate the frequency of the first subglottal resonance.

6.1.4 Estimation the glottal acoustic impedance

Estimates of glottal impedance were computed to explore the effects of its time-invariant and time-varying representations in the acoustic coupling phenomena. This approach also allowed for investigation of the effect of incomplete glottal closure using the proposed IBIF scheme. The glottal impedance is a nonlinear quantity that needed to be linearized to meet the Norton equivalent associated with the dipole circuit representation from Fig. 6.1. Thus, the relation between transglottal pressure (ΔP_g), glottal airflow (noted as U_g in the following equations), and glottal area (A_g) was used to linearize the glottal impedance as time-invariant and also time-varying representations. In both cases such a linearization was based on a Taylor series expansion. The time-invariant case decomposed the glottal airflow into DC (bar notation) and AC (tilde notation) components such that

$$\begin{aligned} U_g(t) &= f(\Delta P_g, A_g) \\ &= \bar{U}_g + \tilde{U}_g(t) \\ &= \bar{U}_g + \left.\frac{\partial U_g}{\partial \Delta P_g}\right|_{\Delta\bar{P}_g,\bar{A}_g} \Delta\tilde{P}_g + \left.\frac{\partial U_g}{\partial A_g}\right|_{\Delta\bar{P}_g,\bar{A}_g} \tilde{A}_g + \ \ldots\ , \end{aligned} \tag{6.13}$$

where the AC flow component was approximated as

$$\tilde{U}_g(t) = \left.\frac{\partial U_g}{\partial \Delta P_g}\right|_{\Delta\bar{P}_g,\bar{A}_g} \Delta\tilde{P}_g + \left.\frac{\partial U_g}{\partial A_g}\right|_{\Delta\bar{P}_g,\bar{A}_g} \tilde{A}_g. \tag{6.14}$$

The first partial derivative in this representation was associated with a parallel glottal conductance and the second term was linked with the ideal glottal airflow source proportional to the glottal area. The conductance was expressed as a resistive impedance term using the transglottal pressure from equation (2.2) and it was given by:

$$\tilde{Z}_g = \left.\frac{\partial(\Delta P_g)}{\partial U_g}\right|_{\bar{U}_g,\bar{A}_g} = k_t \frac{\bar{U}_g}{\bar{A}_g^2}. \tag{6.15}$$

This expression constituted the time-invariant expression that was used in the context of the inverse filtering with glottal coupling. In addition, the time-varying linearization described in [47] was also incorporated. This scheme did not remove the mean values and but rather evaluated with respect to the time-varying glottal area $A_g(t)$ and time-varying uncoupled airflow $U_o(t)$. Thus, using the transglottal pressure from equation (2.2) we obtained

$$Z^*_{g-kt}(t) = \left.\frac{\partial(\Delta P_g)}{\partial U_g}\right|_{\substack{U_g=U_o(t) \\ A_g=A_g(t)}} = kt \frac{U_o(t)}{A_g^2(t)}. \tag{6.16}$$

In addition, it was desired to explore the effects of the mean term k_t, for which it was replaced by a time-varying orifice discharge coefficient (c_d) that described the

viscous losses, inertance, and changes due to separation point into the flow. Thus, an modified expression for the transglottal pressure drop became

$$\Delta P_g(t) = c_d^2(t)\frac{1}{2}\frac{U_g^2(t)}{A_g^2(t)}, \tag{6.17}$$

with the linearized time-varying glottal impedance expressed as:

$$Z^*_{g-cd}(t) = \left.\frac{\partial(\Delta P_g)}{\partial U_g}\right|_{\substack{U_g=U_o(t)\\A_g=A_g(t)}} = c_d(t)\frac{U_o(t)}{A_g^2(t)}. \tag{6.18}$$

To obtain the linearized expressions of the glottal impedance in either its time-invariant form ($\tilde{Z}_g$) or time-varying forms ($Z^*_{g-kt}(t)$ and $Z^*_{g-cd}(t)$), recordings of oral volume velocity and high-speed videoendoscopy were employed. It was noted that the time-invariant term determination only required mean values of the glottal area and airflow, for which an inverse filtering scheme for the latter was not needed. However, the time-varying expression required the use of the uncoupled glottal airflow, which was not known *a priori*. Thus, an iterative scheme that started with the inverse filtered oral volume velocity U_{supra} and made use of equation (6.3) to obtain U_o and refine the estimates was utilized.

Estimation of ideal airflow from measurements of oral airflow made use of a quasi-steady regime, i.e., the glottal impedance and coupled transfer functions were estimated and inverse filtered using equation (6.3) for each time step via FFT and its inverse. Only the samples corresponding to the associated time steps were retrieved from the set of inverse filtered waveforms. In other words, if each inverse filtered waveform is put in a matrix form for each time step, then the diagonal of such matrix represented the waveform obtained using a time-varying glottal impedance.

6.1.5 Estimation of skin properties

In the proposed IBIF scheme, default mechanical properties for the skin were taken from early studies [209, 210]. To account for variation of the tissue properties among subjects and cases a calibration method that used a multidimensional waveform matching was implemented. In this method, skin properties were adjusted to match oral airflow waveforms translated to different locations using the multiple transfer functions described in this section, i.e., estimates were contrasted at the mouth, glottis, and neck. Waveforms were aligned and skin properties were selected to minimize the root mean squared error (RMSE).

This approach may have also compensated for other differences in the system that could not be related to the neck surface mechanical properties, such as tracheal diameter and losses in the subglottal system. Nevertheless, it was expected that the primary factor affecting these calibration was indeed the mechanical properties of the skin underlying the accelerometer.

In addition, the mass of accelerometer Knowles BU-7135 with its current mounting configuration was included in Z_{rad} in equation (6.9), having a mass of 1.10 g and a total contact area of 4.15 cm^2 (including its silicone disk), thus yielding a mass per unit area of 0.26 g. These values were maintained constant throughout the simulations.

6.2 Evaluation of the inverse filtering scheme

Two methods were used to evaluate the IBIF inverse filtering scheme for controlled scenarios that represented different quantifiable glottal configurations in normal phonation during sustained vowels. Synthetic speech and recordings of sustained vowels for different glottal configurations were considered.

The numerical simulations presented in chapter 3 were used to initially test the inverse filtering scheme. Two vowels (/a/ and /i/) and two glottal conditions (complete closure and incomplete closure) were used to evaluate the CPIF, IBIF schemes, and system coupling. Differences between the wave reflection analog and transmission line

schemes were noted, particularly for the subglottal system. To minimize the effects associated to this difference, the impedances were taken from measurements obtained in the wave reflection analog using the corresponding frequency domain terms via frequency transformations of an impulse response.

The output of the numerical simulations yielded the "true" glottal airflow, and included ripples in its open phase, skewing, and fluctuations in the closed phase portion when the glottis was set to incomplete closure. This "true" airflow source was compared to the inverse filtering estimates obtained from oral airflow (also provided by the numerical model) for the CPIF and IBIF schemes. Investigations on the system coupling made use of an ideal source also provided through the numerical simulations.

In addition to qualitative observations, glottal measures obtained from the inverse filtering schemes were contrasted to those from the model. This type of comparison has been used recently to explore the ability of the CPIF scheme to correctly estimate the main characteristics of the glottal source [32], where the selected measures of glottal behavior included the difference between the first two harmonics (H1-H2) and the harmonic richness factor (HRF). Two additional measures of interest were included the amplitude of the AC airflow, and maximum flow declination rate (MFDR).

6.3 Results

Three major topics are presented in this section. First, the accuracy of inverse filtering schemes to describe the "true" glottal flow was assessed using synthetic speech as a standard for complete and incomplete glottal closure conditions. Then, estimates of glottal airflow and oral airflow from the acceleration signal were evaluated for various cases. Finally, the coupled IBIF system was used to investigate whether it was possible to retrieve the ideal glottal source for synthetic and human speech.

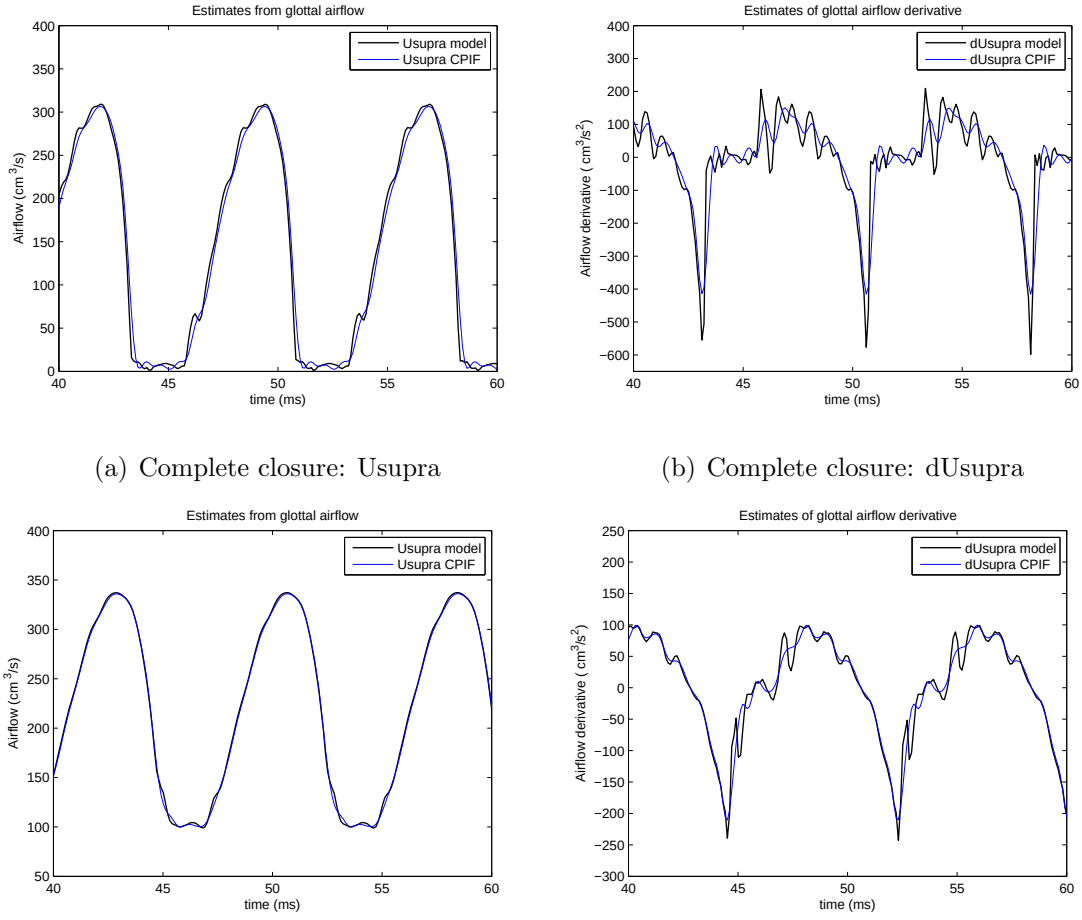
6.3.1 Evaluation of uncoupled inverse filtering schemes

Synthetic speech

Estimates of "true" glottal airflow were obtained through numerical simulations that included incomplete glottal closure. The simulations were obtained using the IF72 parameter set with a lung pressure of 800 Pa and the updated aerodynamic equations [74] studied previously. For the cases with incomplete glottal closure, a posterior glottal opening was included with a fixed area of 0.03 cm^2. The simulations yielded the "true" glottal source as well as the oral airflow. Closed phase inverse filtering was obtained from the latter based on the methods described in [32]. In addition, the results were contrasted with those from the uncoupled IBIF scheme for vocal tract.

The results of the simulations are initially presented for qualitative observations of two vowels (/i/ and /a/) and two glottal conditions (complete and incomplete glottal closure) in Fig. 6.9 and Fig. 6.8. It can be observed from these figures that the CPIF estimates of glottal airflow obtained from oral airflow closely resembled the "true" flow from the numerical model. The performance of the technique appeared to vary between glottal configurations and vowels. The primary differences were observed in the suppression of ripples during the open and closed phase portions of the cycle. It was noted that inverse filter of vowel /i/ was possible, with accuracy comparable to that of the numerical model if ad-hoc changes in the selection of the covariance window were taken into account, i.e., if the covariance window was only composed by samples within the closed phase portion (see [32] for details on the covariance method).

The same cases were subsequently inverse filtered using the uncoupled IBIF model and presented in Fig. 6.10 and Fig. 6.11. These estimates better-captured the temporal variations of the signal with a more consistent performance. However, it was noted that this performance could have been overestimated for the synthetic speech due to similarities in the underlying assumptions with respect to the tract geometry.



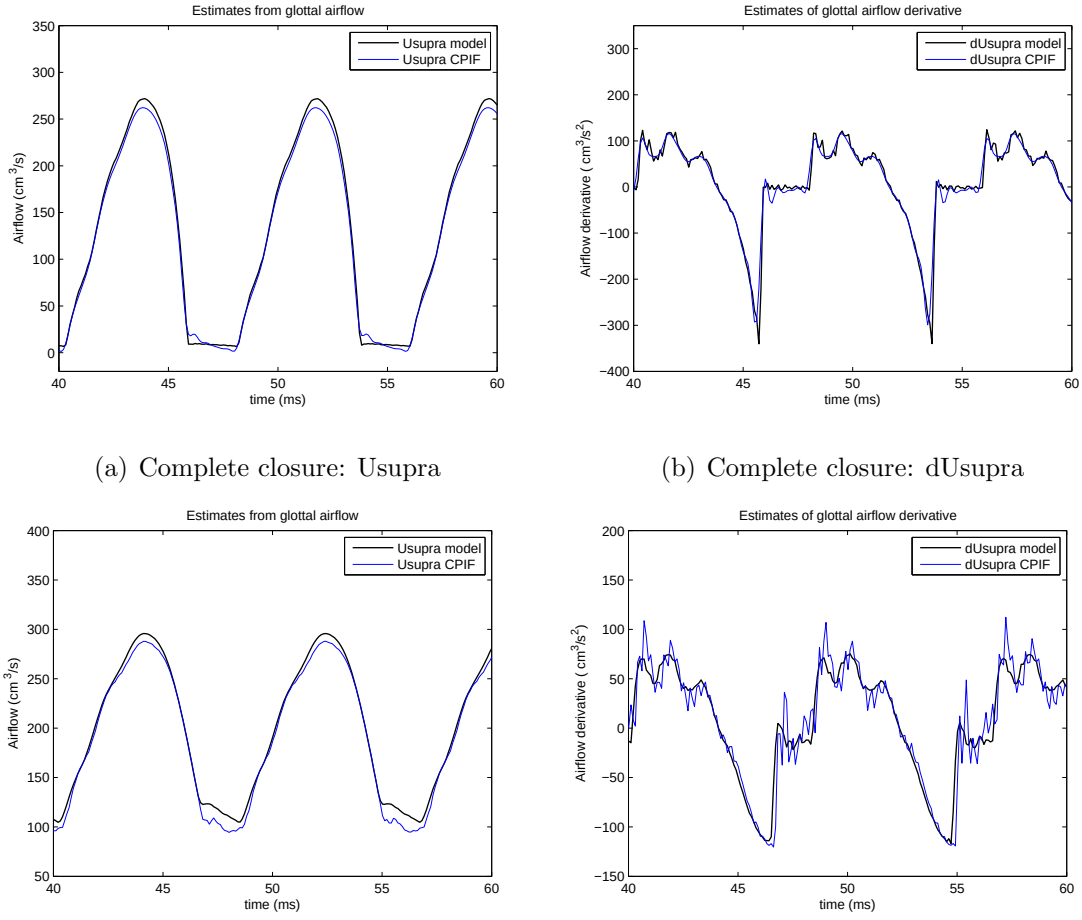
(a) Complete closure: Usupra

(b) Complete closure: dUsupra

(c) Incomplete closure: Usupra

(d) Incomplete closure: dUsupra

Fig. 6.8. Ability of CPIF scheme to capture the "true" glottal airflow for complete and incomplete glottal closure for vowel /a/



(a) Complete closure: Usupra

(b) Complete closure: dUsupra

(c) Incomplete closure: Usupra

(d) Incomplete closure: dUsupra

Fig. 6.9. Ability of CPIF scheme to capture the "true" glottal airflow for complete and incomplete glottal closure for vowel /i/

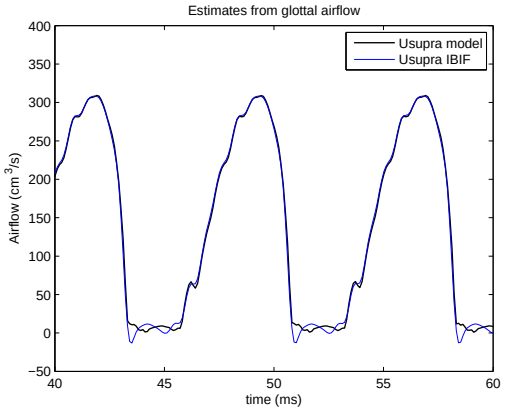
Nevertheless, it was noted that the IBIF scheme was able to retrieve both vowels with reasonable accuracy, illustrating again that inverse filtering of vowel /i/ was possible.

A quantitative view of the differences presented in these cases is shown in Table 6.1. Direct comparison of the waveforms was obtained via the RMSE for both glottal airflow and its derivative. In addition, the differences between selected glottal measures obtained from both inverse filtered signals and the "true" flow from the model was included. Each of these measures was taken from at least 10 subsequent cycles for each case. These measures illustrated that the CPIF represented the flow better than its derivative, as the RMSE for the airflow and the AC flow difference were small, whereas the RMSE for the flow derivative and MFDR differences were relatively larger. It was observed that the error of the harmonic measures was also small. Comparisons of the CPIF and the harmonic measures have been performed before [32] and exhibited similar differences, but no attempts to describe the airflow derivatives and related measures have been made.

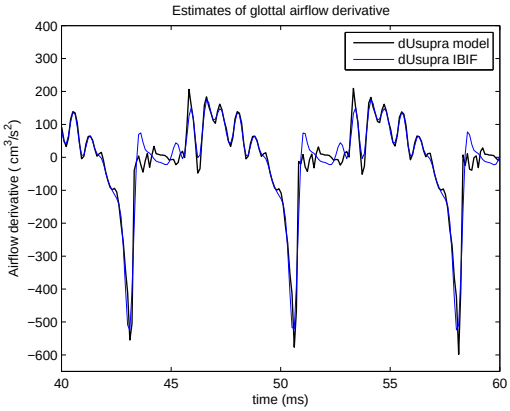
Confirming observations from the waveforms, better quantitative results were obtained using the IBIF scheme for synthetic speech, where most IBIF measures were lower than those from CPIF with a few exceptions. The largest errors were observed during complete glottal closure, a case that is less likely to occur in real speech [25]. This behavior indicated that both schemes were capable to inverse filter and obtain the primary features of the "true" airflow , although the IBIF yielded better results for the synthetic speech case.

Observations from human speech recordings

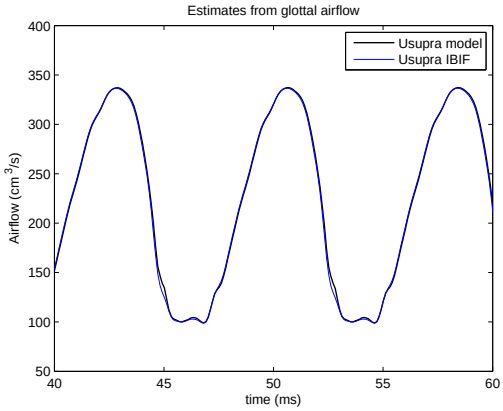
Testing the accuracy of both inverse filtering schemes with human speech data is challenging, as the "true" glottal flow is unknown. However, comparisons between the estimates obtained from both CPIF and IBIF schemes were included to evaluate each under more realistic conditions. It was also desired to investigate whether the schemes were able to inverse-filter human speech recordings of high vowels (e.g., vowel /i/),



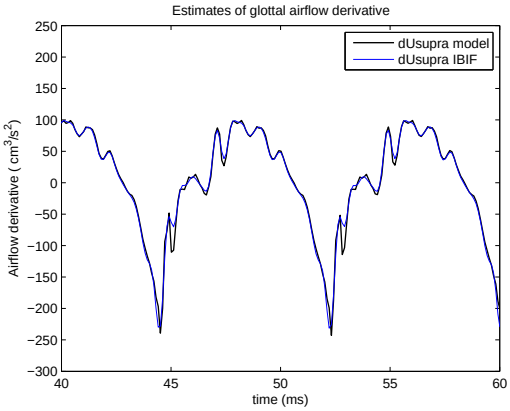
(a) Complete closure: Usupra



(b) Complete closure: dUsupra

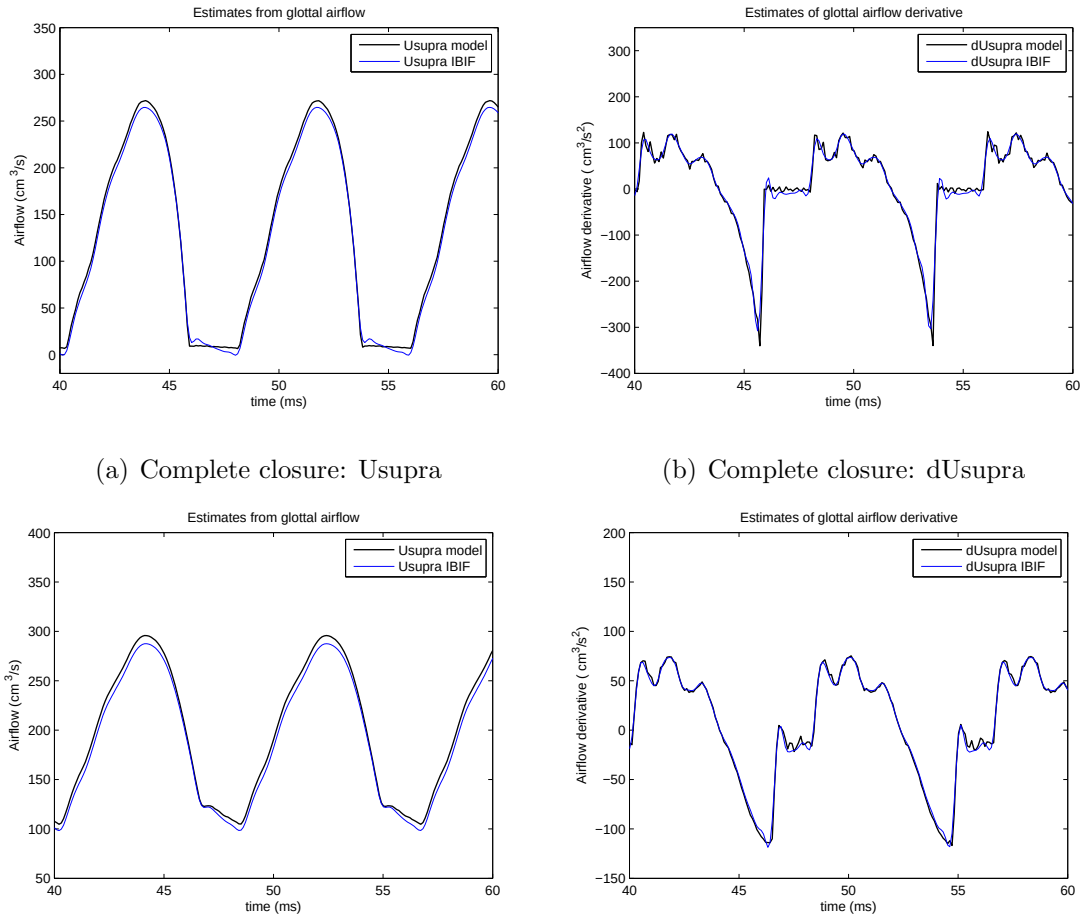


(c) Incomplete closure: Usupra



(d) Incomplete closure: dUsupra

Fig. 6.10. Ability of IBIF scheme to capture the "true" glottal airflow for complete and incomplete glottal closure for vowel /a/



(a) Complete closure: Usupra

(b) Complete closure: dUsupra

(c) Incomplete closure: Usupra

(d) Incomplete closure: dUsupra

Fig. 6.11. Ability of IBIF scheme to capture the "true" glottal airflow for complete and incomplete glottal closure for vowel /i/

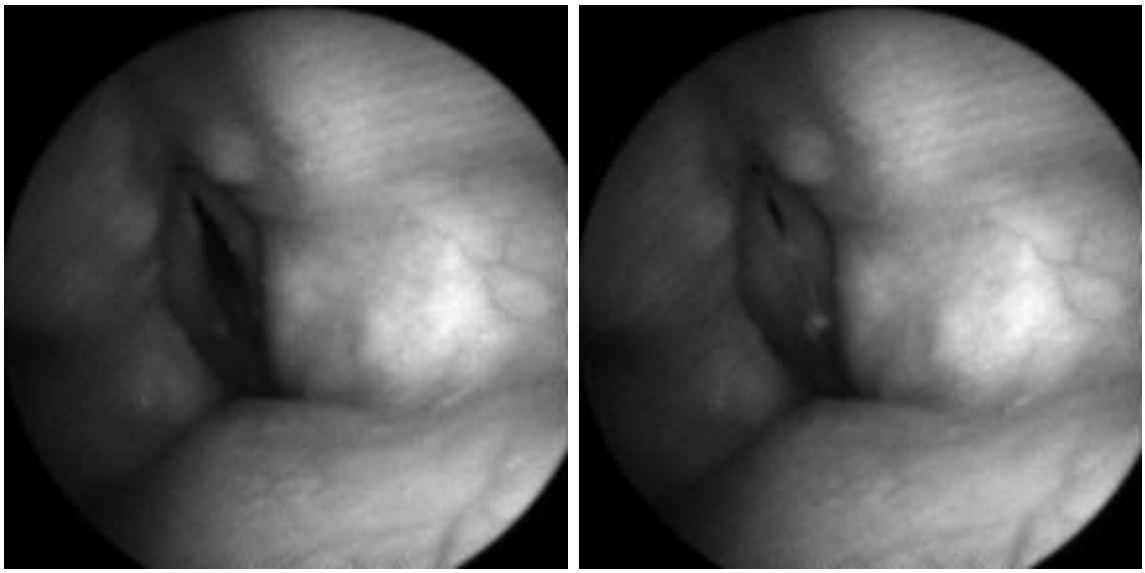
Table 6.1
Error associated with CPIF and IBIF schemes with respect to "true" flow from numerical simulations for complete (com.) and incomplete (inc.) glottal closure

		CPIF				IBIF			
		Vowel /a/		Vowel /i/		Vowel /a/		Vowel /i/	
Measure	Units	com.	inc.	com.	inc.	com.	inc.	com.	inc.
RMSE Usupra	(cm^3/s)	12.3	2.7	5.1	11.5	6.8	3.2	4.2	2.4
RMSE dUsupra	(cm^3/s^2)	62.0	19.0	22.0	16.3	37.6	9.7	15.6	3.9
AC airflow	(cm^3/s)	-4.0	-1.9	-4.4	2.3	13.8	-0.4	0.0	-1.9
MFDR	(cm^3/s^2)	-171	-29	-35	5.0	-55.3	-6.9	-26.7	2.7
H1-H2	(dB)	-0.2	0.1	-1.0	1.7	0.2	0.1	-0.9	-0.9
HRF	(dB)	-0.4	0.2	-0.9	0.7	0.5	0.3	-0.6	-0.6

cases that have not been considered in inverse filtering studies before. However, the results from previous section using synthetic speech suggested that this was possible to achieve in both schemes.

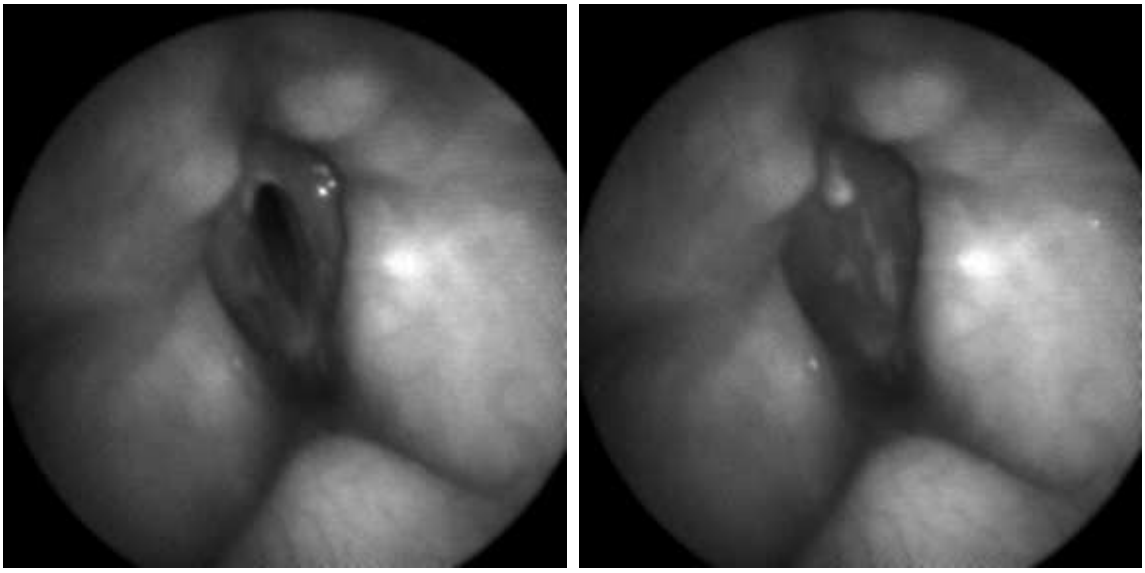
Recordings of sustained vowels on the "calibrated" male subject are presented. Cases included two vowels (/a/ and /i/) for two glottal conditions (chest and falsetto register) representing two different glottal conditions (normal and abducted). The glottal conditions for these recordings are observed from the HSV endoscopic view for the chest register in Fig 6.12 and falsetto register in Fig 6.13. These cases were initially evaluated in this section and extensively analyzed in subsequent sections in this chapter.

The inverse filtered signals for the chest register is presented in Fig 6.14 and for falsetto register in Fig 6.15. Both schemes appeared to be able to retrieve the glottal airflow for both vowels and glottal conditions. It was noted that the IBIF estimates differed more from the CPIF scheme for actual speech recordings than in synthetic case, exhibiting similar trends but with stronger fluctuations, particularly



(a) Vowel /a/ - fully open

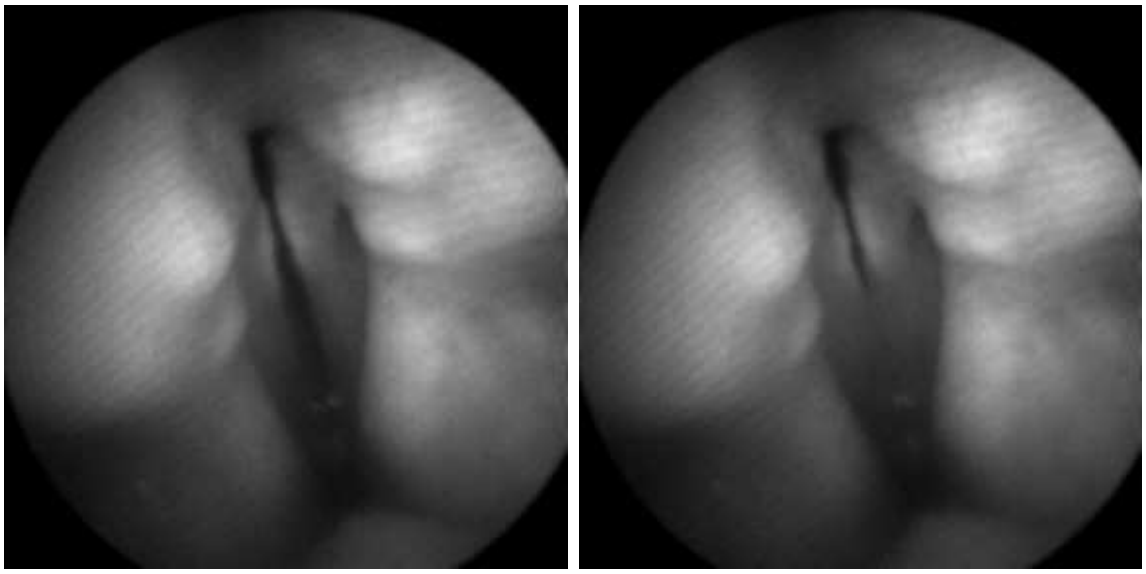
(b) Vowel /a/ - fully closed



(c) Vowel /i/ - fully open

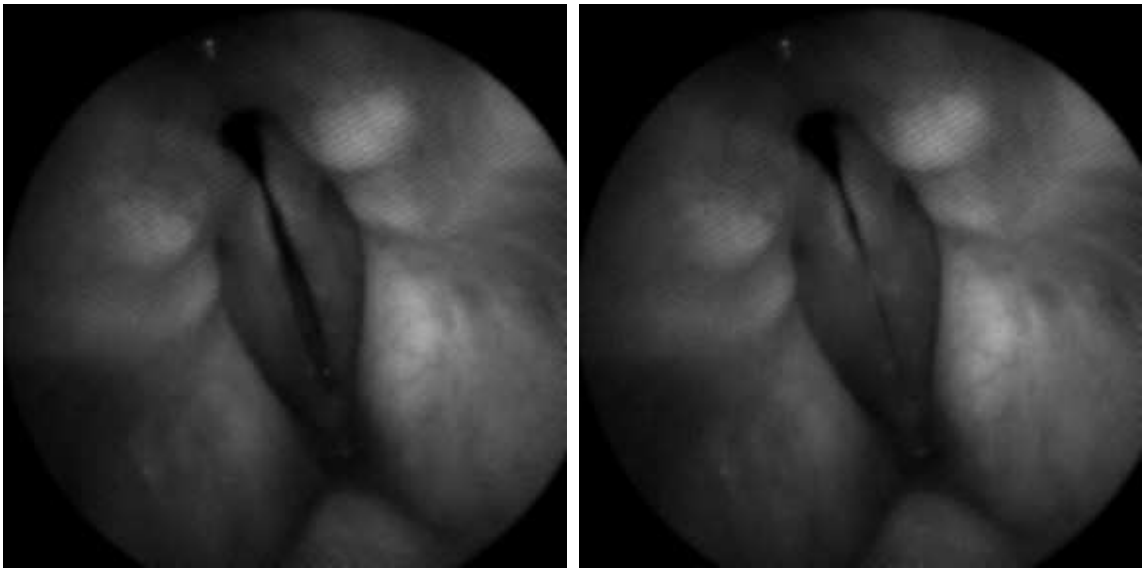
(d) Vowel /i/ - fully closed

Fig. 6.12. Endoscopic view of HSV recordings for both vowels uttered with chest register



(a) Vowel /a/ - fully open

(b) Vowel /a/ - fully closed



(c) Vowel /i/ - fully open

(d) Vowel /i/ - fully closed

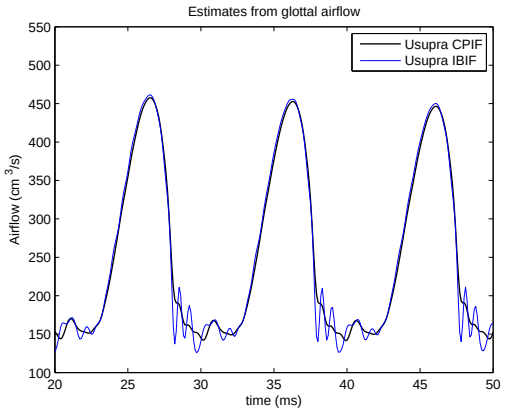
Fig. 6.13. Endoscopic view of HSV recordings for both vowels uttered with falsetto register

for the chest register. These trends were also reflected in the difference between the estimates presented in Table 6.2.

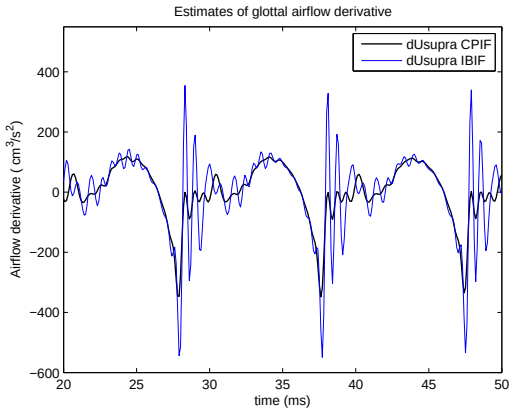
It is interesting to note that the estimates of glottal airflow for vowel /i/ match those from numerical simulations, where glottal pulses were skewed and a single fluctuation (associated to the first formant) in the open phase was noted. The glottal pulse was also less skewed than vowel /a/, as also noted from the less negative peak in the flow derivative (i.e., smaller MFDR). This indicated that vowel /i/ does impose different acoustic loading condition. This differences appeared less pronounced when a more abducted glottal condition was presented, suggesting that this condition was less affected by the acoustic coupling. It will be noted in the subsequent section that similar estimates for this vowel can be obtained with a different approach that used neck surface acceleration, confirming these observations.

Table 6.2
Differences between CPIF and IBIF schemes from oral airflow recordings

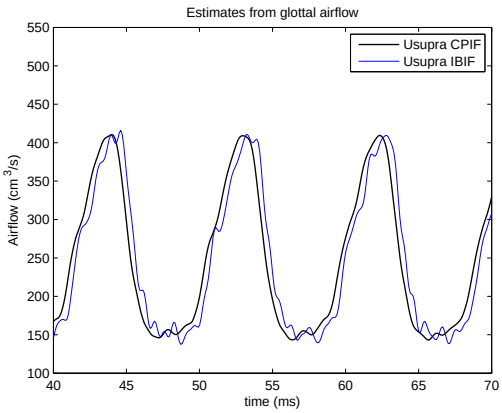
		Chest		Falsetto	
Measure	Units	/a/	/i/	/a/	/i/
F_0	(Hz)	102.5	107.4	227.1	224.6
RMSD Usupra	(cm^3/s)	9.8	21.7	13.3	27.9
RMSD dUspra	(cm^3/s^2)	77.7	44.7	38.6	40.5
AC aiflow	(cm^3/s)	12.6	5.0	8.6	-55.8
MFDR	(cm^3/s^2)	168.3	48.1	-7.3	-32.9
H1-H2	(dB)	0.1	0.1	0.2	1.8
HRF	(dB)	0.4	0.7	-0.5	2.5



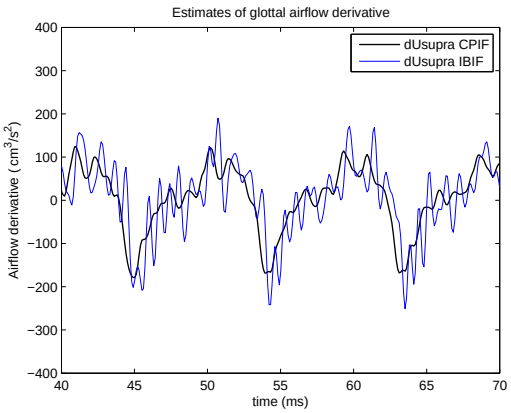
(a) Usupra: Vowel /a/ - chest



(b) dUsupra: Vowel /a/ - chest

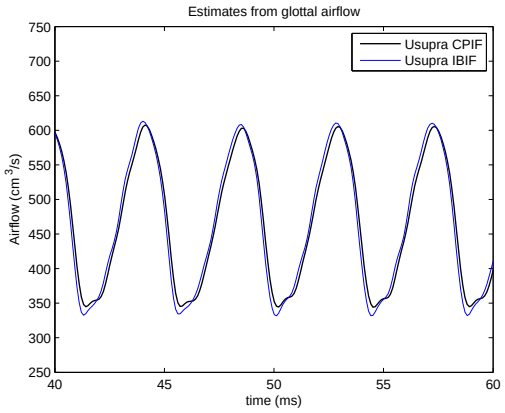


(c) Usupra: Vowel /i/ - chest

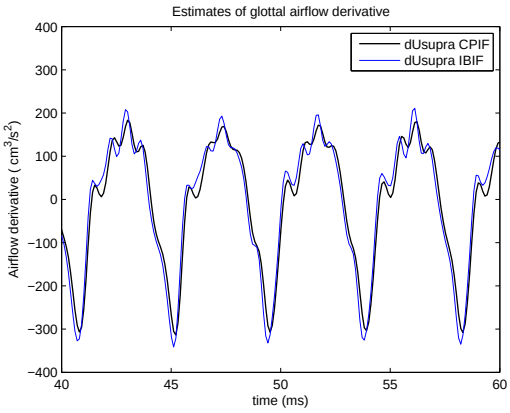


(d) dUsupra: Vowel /i/ - chest

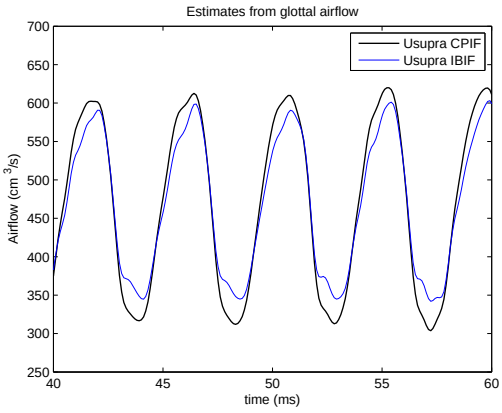
Fig. 6.14. Comparison between CPIF vs IBIF schemes for chest register and both vowels



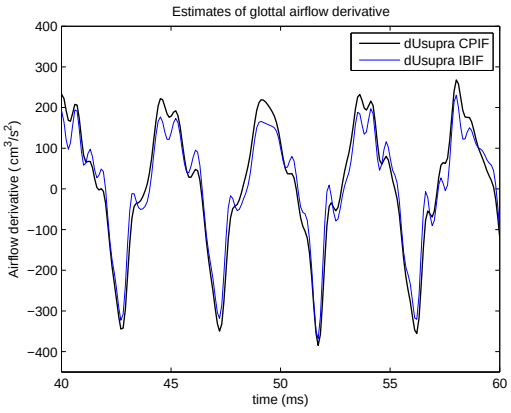
(a) Usupra: Vowel /a/ - falsetto



(b) dUsupra: Vowel /a/ - falsetto



(c) Usupra: Vowel /i/ - falsetto



(d) dUsupra: Vowel /i/ - falsetto

Fig. 6.15. Comparison between CPIF vs IBIF schemes for falsetto register and both vowels

6.3.2 Parameter estimation from neck surface acceleration

As noted in the previous section, the ability to obtain estimates of the "true" airflow that is entering the vocal tract did not depend on the glottal configuration or coupling between the system components. In this section, the results of the IBIF estimates of glottal airflow (U_{supra}) via measurement of neck surface acceleration are presented. The effects of the system coupling on the glottal airflow are explored in the subsequent section.

The results of this section are separated into qualitative observations and quantitative comparisons. For both cases the goal was to assess the ability of the proposed IBIF model to transform neck surface recordings into glottal aerodynamic parameters. The accelerometer signal (ACC) used in this study is known to be an AC signal, for which all estimates were contrasted with reference waveforms with no DC components. It was discussed earlier that strong clinical impact could be obtained from these unsteady representations in patient monitoring scenarios.

The same human speech recordings from the previous section were used, i.e., two vowels imposing different loading conditions (/a/ and /i/) and two different registers (chest and falsetto) representing different glottal conditions.

Qualitative observations

Estimates of glottal airflow (U_{supra}), its derivative (dU_{supra}), and oral airflow (U_m) were obtained from the ACC signal and contrasted with CPIF glottal airflow, its derivative, and the measured oral airflow from the CV mask, respectively. In addition, the transfer functions were also applied so that estimates of the neck surface acceleration were obtained from the oral airflow recordings. The raw waveforms for these cases are presented for vowel /a/ in chest register in Fig. 6.16 , vowel /i/ in chest register in Fig. 6.17, vowel /a/ in falsetto register in Fig. 6.18, and vowel /i/ in falsetto register in Fig. 6.19.

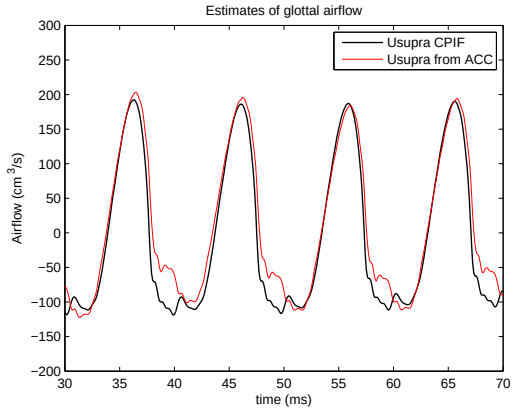
It is noticeable from these comparisons that the IBIF model yielded a close match between ACC-based waveforms and those from CPIF, with an error that appeared to vary between the glottal conditions and vowels. It was also observed that the closest waveform match was obtained during the open phase portion of the cycle.

Quantitative analysis

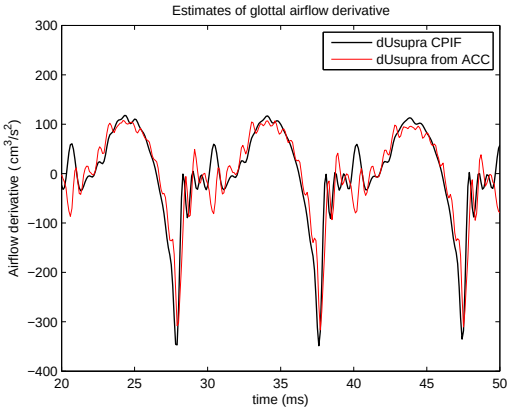
A quantitative analysis of these and additional results is presented in this section. The same measures were used to compare the estimated aerodynamic parameters from the acceleration signal.

Both "calibrated" subjects (male and female) recorded the same two vowels (/a/ and /i/) for three different glottal conditions (normal/chest, high pitch/falsetto, and breathy/abducted). Additionally, two recordings with chest register were considered for the male subject, providing a total of 14 cases under study. The CPIF-based values obtained from simultaneous oral airflow measurements were contrasted with those from ACC, where the raw measures for each case are presented in Table 6.3. It was observed that for the normal chest cases of vowel /a/, the measures were within the expected range for male and female cases from previous studies [25]. It should be noted that vowel /i/ has not been previously studied and thus has no reference for comparisons.

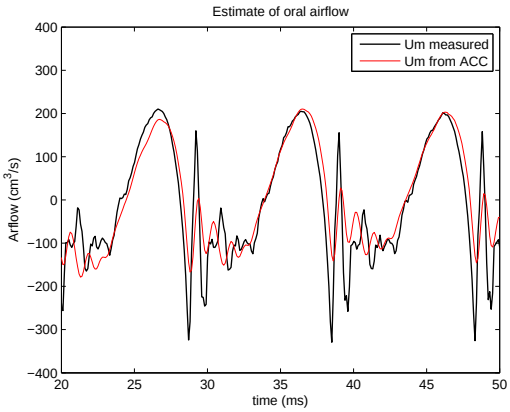
The mean values for the selected measures of glottal behavior and the differences between the CPIF-based and ACC-based measures are presented in Tables 6.4 and 6.5. It was observed that the mean values for the estimation were remarkably close, although there were key variations. The larger of these variations in Table 6.4 were associated with the significant dispersion between the parameters obtained for the different glottal conditions. Thus, it was often more relevant to examine the mean values of the absolute error for each case, as shown in Table 6.5. Here, the error and its variations were considered sufficiently low to potentially make this scheme clinically useful. The RMSE values tend to underestimate the ability of the ACC



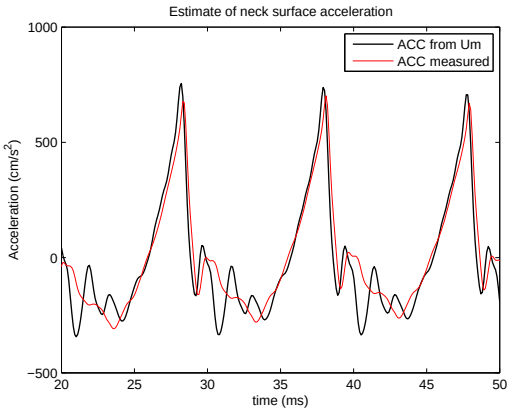
(a) Glottal airflow from ACC



(b) Glottal airflow derivative from ACC

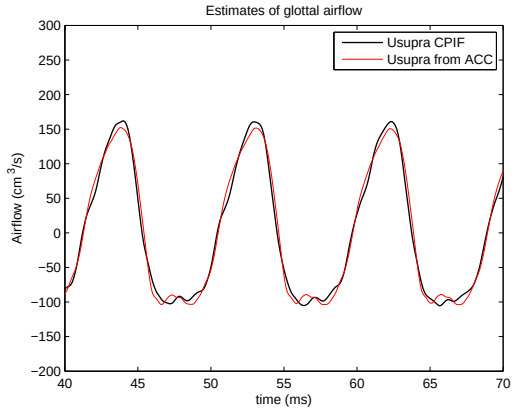


(c) Oral airflow from ACC

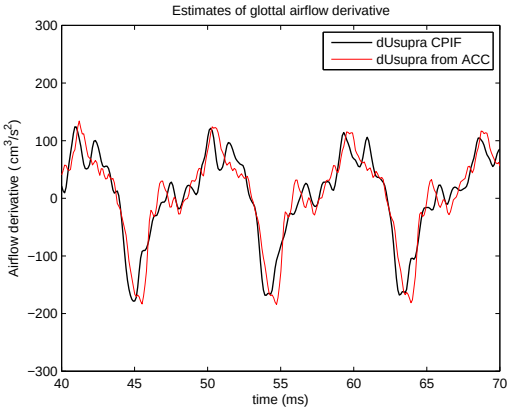


(d) Acceleration from Um

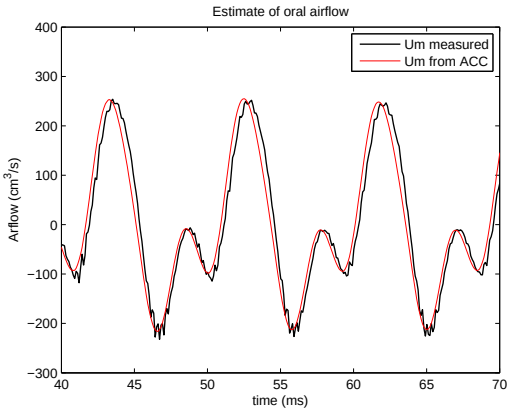
Fig. 6.16. Estimates of multiple signals obtained from measurements of neck surface acceleration (ACC) and oral airflow (Um) for a vowel /a/ in chest register. Estimates from ACC have no DC component.



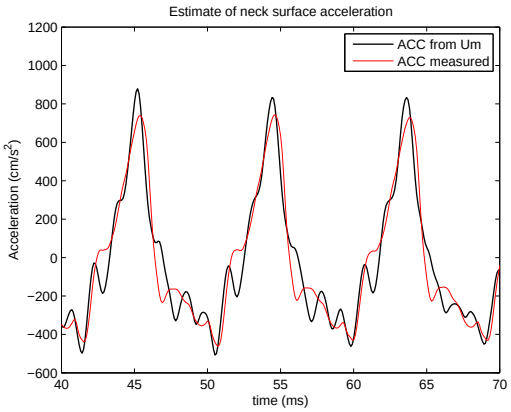
(a) Glottal airflow from ACC



(b) Glottal airflow derivative from ACC

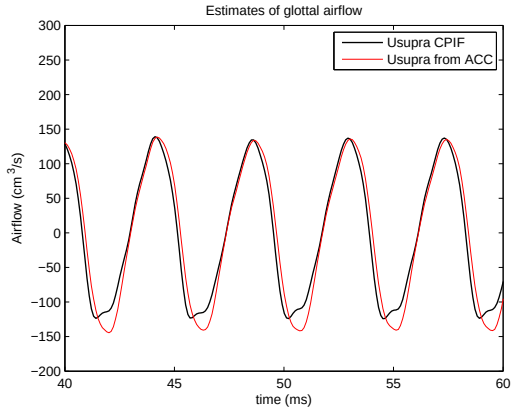


(c) Oral airflow from ACC

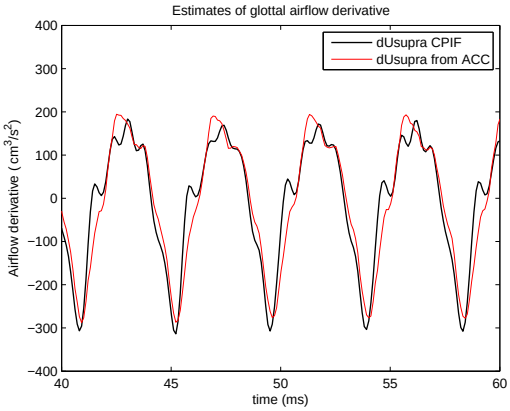


(d) Acceleration from Um

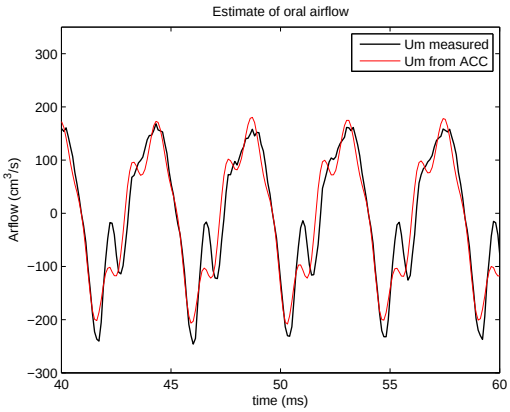
Fig. 6.17. Estimates of multiple signals obtained from measurements of neck surface acceleration (ACC) and oral airflow (Um) for a vowel /i/ in chest register. Estimates from ACC have no DC component.



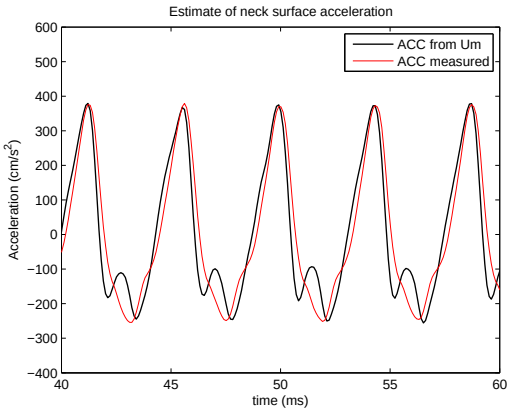
(a) Glottal airflow from ACC



(b) Glottal airflow derivative from ACC

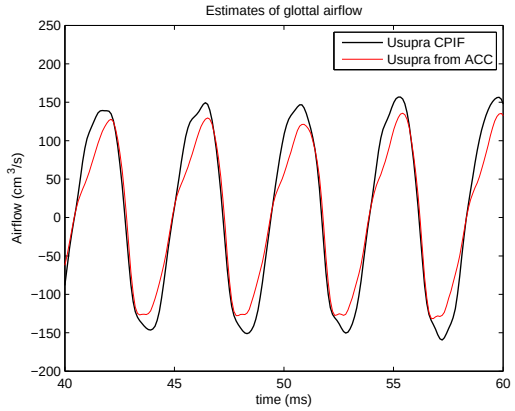


(c) Oral airflow from ACC

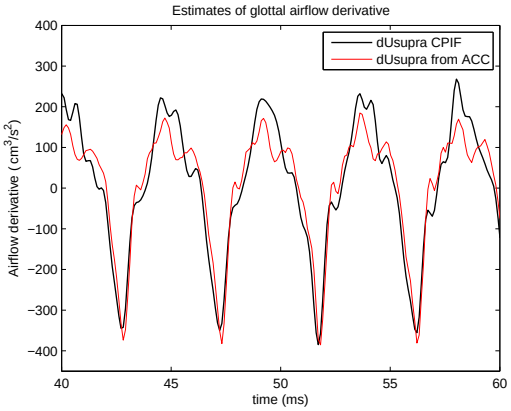


(d) Acceleration from Um

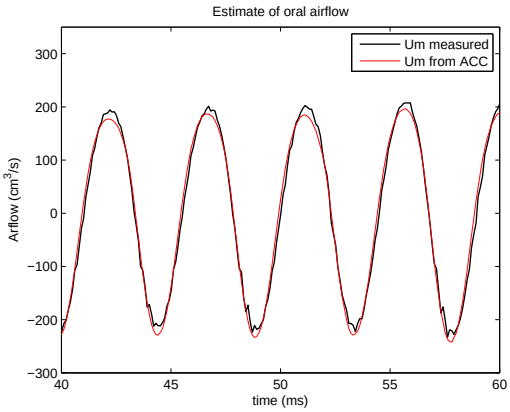
Fig. 6.18. Estimates of multiple signals obtained from measurements of neck surface acceleration (ACC) and oral airflow (Um) for a vowel /i/ in falsetto register. Estimates from ACC have no DC component.



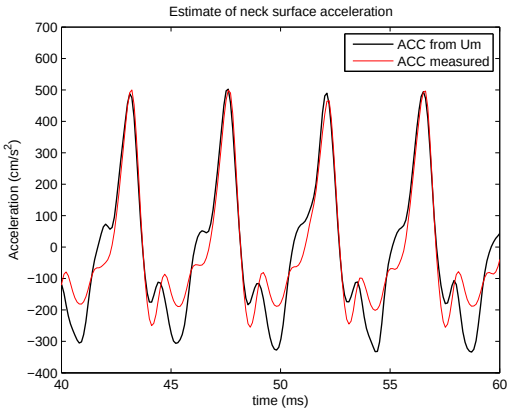
(a) Glottal airflow from ACC



(b) Glottal airflow derivative from ACC



(c) Oral airflow from ACC



(d) Acceleration from Um

Fig. 6.19. Estimates of multiple signals obtained from measurements of neck surface acceleration (ACC) and oral airflow (Um) for a vowel /i/ in falsetto register. Estimates from ACC have no DC component.

Table 6.3
Raw data from CPIF (1) and ACC (2) measures of glottal behavior

Measure	Female subject						Male subject							
	Chest		Breathy		Falsetto		Chest		Breathy		Falsetto		Chest	
	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/	/a/	/i/
fo	225	229	229	237	488	481	117	115	120	117	227	225	103	107
RMSE U_{supra}	24	29	18	9	14	19	24	15	11	10	15	33	26	13
RMSE dU_{supra}	81	162	22	29	72	110	48	23	21	18	50	58	47	27
RMSE U_m	62	68	18	21	107	71	27	15	17	15	52	16	268	14
RMSE ACC	123	110	47	38	154	84	70	50	39	50	82	78	68	116
SPL 1	87	96	76	79	91	88	74	80	67	78	83	86	84	85
SPL 2	89	89	78	77	93	84	74	77	67	72	83	85	80	82
AC flow 1	286	320	202	122	123	140	230	147	150	128	270	302	312	269
AC flow 2	297	371	204	119	127	144	185	150	133	136	282	246	344	263
MFDR 1	467	558	177	142	304	406	214	102	85	80	380	340	351	175
MFDR 2	428	617	184	140	342	439	192	129	72	76	328	337	336	196
H1-H2 1	-15	-9	-26	-10	-9	-5	-10	-12	-23	-17	-16	-21	-9	-12
H1-H2 2	-15	-11	-21	-15	-4	0	-8	-12	-21	-22	-18	-12	-12	-13
HRF 1	-13	-8	-24	-10	-9	-5	-9	-12	-21	-16	-14	-18	-8	-11
HRF 2	-13	-9	-21	-15	-4	0	-7	-10	-20	-21	-17	-11	-11	-11

signal to derive useful quantities. For instance, particular emphasis was given to the AC flow and MFDR estimates which are the most indicative measures of vocal hyperfunction [25] and yielded the most accurate estimates from these experiments.

Table 6.4
Mean values from ACC measures and those from CPIF and measured values

	CPIF	ACC
	Mean ± Std	Mean ± Std
SPL	82.3 ± 7.4	80.6 ± 7.2
AC flow	214.4 ± 77.7	214.4 ± 85.7
MFDR	270.1 ± 152.1	272.6 ± 157.8
H1-H2	-13.7 ± 5.9	-13.1 ± 6.4
HRF	-12.7 ± 5.4	-12.1 ± 6.3

Table 6.5
Estimation error between ACC measures and those from CPIF and measured values

		ACC-based measures
Measures	Units	Mean ± Std
RMSD Usupra	(cm^3/s)	18.5 ± 7.5
RMSD dUspra	(cm^3/s^2)	54.8 ± 40.8
RMSD Um	(cm^3/s)	55.0 ± 67.8
RMSD ACC	(cm/s^2)	79.1 ± 35.3
SPL	(dB)	-1.7 ± 2.9
AC flow	(cm^3/s)	0.0 ± 27.0
MFDR	(cm^3/s^2)	2.5 ± 30.8
H1-H2	(dB)	0.6 ± 4.1
HRF	(dB)	0.6 ± 3.7

The variations of the estimated skin parameters are summarized in Table 6.6. It was observed that the parameters were within measured values [209] with variations observed between values and vowels for each case.

Table 6.6
Mean values of skin properties used in the model. All units are in per-unit-area as in [209].

	Female subject	Male subject
Mm (g)	2.1 ± 0.4	2.4 ± 0.4
Rm (g/s)	1334 ± 550	1244 ± 215
Km (dyne/cm) $\cdot 10^3$	2813 ± 918	614 ± 106

6.3.3 System decoupling: Synthetic speech

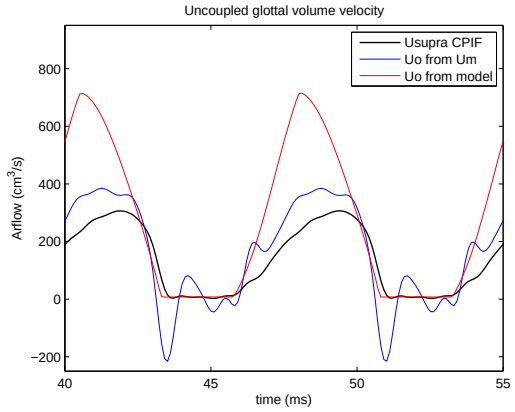
Retrieving the ideal airflow source U_o in the dipole model (see Fig. 6.1) from recordings of oral airflow was referred to as decoupling the system, as this source was expected to be free from effects introduced by acoustic coupling. The coupled transfer function from equation (6.3) was used to decouple the CPIF synthetic speech cases. Two vowels (/a/ and /i/) and two glottal configurations (complete closure and incomplete closure) were analyzed. Simulations were run for different types of glottal impedance terms: i) linearized time-invariant ($\tilde{Z}_g$), ii) linearized time-varying with mean k_t loss term (Z^*_{g-kt}), iii) linearized time-varying with time-varying $cd(t)$ loss term (Z^*_{g-cd}), and iv) the original nonlinear time-varying impedance (Z_g) from equation (2.3). In addition, the ideal uncoupled flow U_o from the numerical model was defined as in [92]

$$U_o(t) = A_g(t) \cdot \sqrt{\frac{2\ P_s}{\rho}}. \tag{6.19}$$

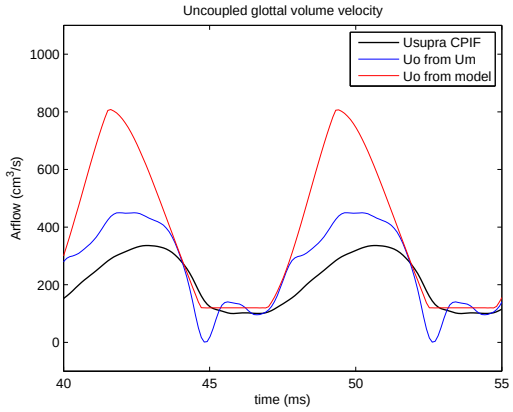
Time-invariant impedance

The CPIF glottal airflow (U_{supra}), estimated uncoupled airflow (U_o from U_m), and ideal uncoupled airflow (U_o from model) are shown in Fig. 6.20 for the time-

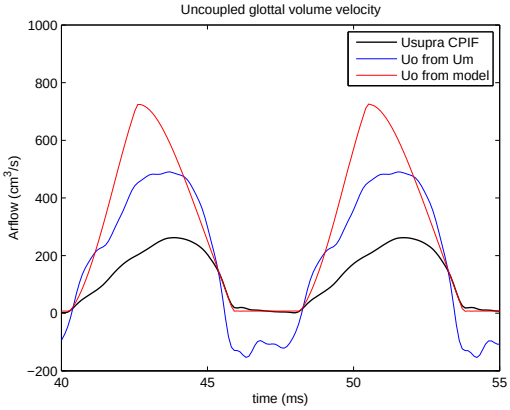
invariant impedance ($\tilde{Z}_g$), different vowels, and glottal configurations. As, none of the estimated uncoupled airflow yielded waveforms similar to that of the ideal case, and some exhibited large negative fluctuations, a time-invariant glottal impedance appeared limited to describe the effects of source filter interactions and it was not expected to illustrate the effect of incomplete glottal closure. In addition, the mean impedances were generally low, yet slightly higher values for cases with complete glottal closure were noted. This was expected given the relatively large glottal area ($\sim 0.2\ cm^2$) that the numerical simulations yielded.



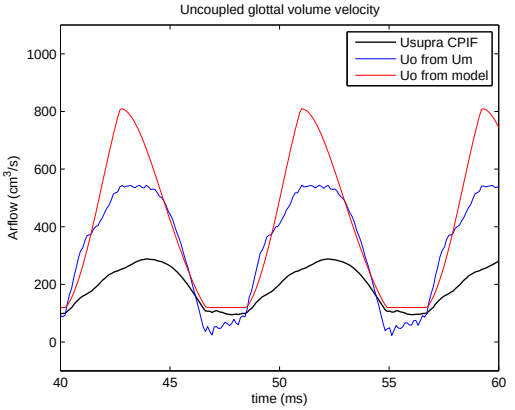
(a) vowel /a/ complete closure, $\tilde{Z}g$=34 Ω



(b) vowel /a/ incomplete closure, $\tilde{Z}g$=26 Ω



(c) vowel /i/ complete closure, $\tilde{Z}g$=27 Ω



(d) vowel /i/ incomplete closure, $\tilde{Z}g$=22 Ω

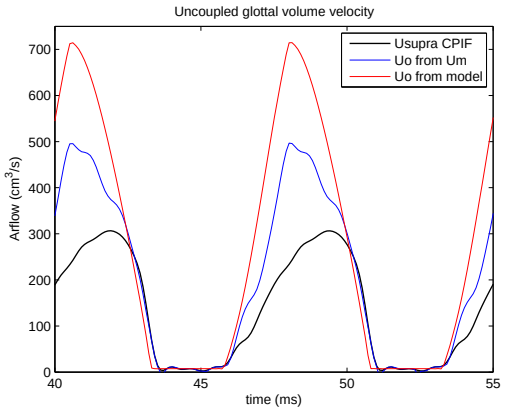
Fig. 6.20. Estimated uncoupled glottal airflow from synthetic speech with time-invariant impedance. All impedance units are in cgs.

Time-varying impedance

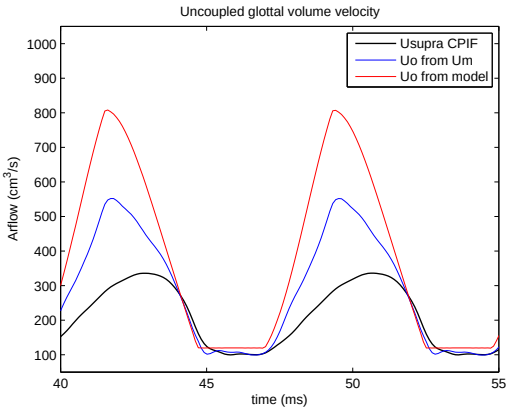
The CPIF glottal airflow (U_{supra}), estimated uncoupled airflow (U_o from U_m), and ideal uncoupled airflow (U_o from model) are shown in Fig. 6.21 for the multiple time-varying impedances ($\tilde{Z}_g$), vowel /a/, and both glottal configurations. It was observed for all cases shown in Fig. 6.21 that the addition of a time-varying expression significantly improved the estimation of the uncoupled glottal airflow. All cases exhibited a degree of de-skewing of the glottal airflow pulse, illustrating that the origin of this phenomena in the synthetic signals was produced by the acoustic coupling.

Among the time-varying glottal impedances used in this experiment, the nonlinear one yielded the best match with respect to the ideal uncoupled case. The linearized impedance with a mean k_t factor yielded the second best results, as it better preserved the shape of the ideal pulse but was not able to retrieve its complete amplitude. The linearized impedance with a time-varying c_d factor exhibited a good match at the beginning and end, but not in its peak. This raised the question whether the orifice discharge coefficient was simply proportional to the tissue velocity or if such an estimate was sufficiently accurate. In addition, it was observed that the nonlinear cases were more unstable, meaning that only one iteration was possible (whereas other cases were obtained with up to 8 iterations) and artifacts were observed during the closed portion of the cycle, particularly during complete closure. The incomplete glottal closure cases for all impedances showed the same trends but yielded smoother waveforms, probably due to the better estimates from CPIF for this case.

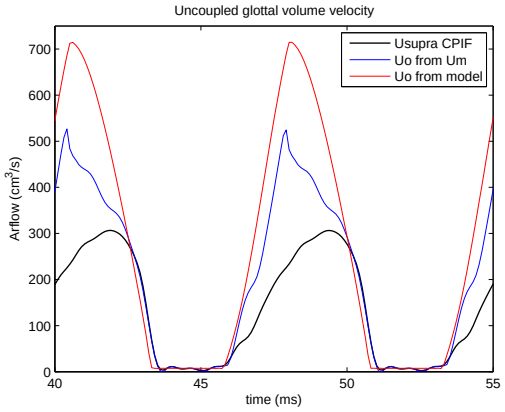
A representation of the temporal patterns of the glottal impedance for the synthetic vowel /a/ for complete and incomplete glottal closure is presented in Fig. 6.22. The temporal variations of each of the impedances were presented along with the mean impedance $\tilde{Z}_g$. An identical expression but associated with minimum glottal airflow and glottal area values represented a linearized time-invariant impedance of the glottal gap. In addition, a subtle trace of the glottal area (with no units) was included to elucidate the open and closed portions of the cycle.



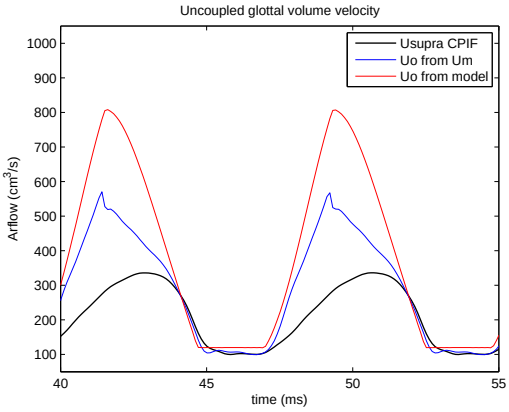
(a) Uo from Zg via kt for complete closure



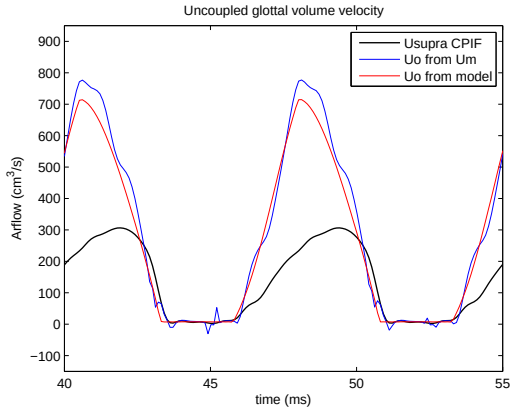
(b) Uo from Zg via kt for incomplete closure



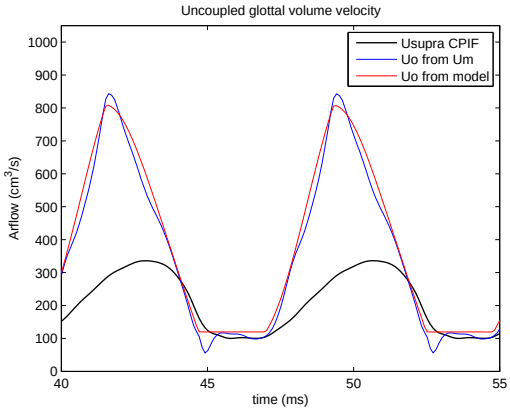
(c) Uo from Zg via cd for complete closure



(d) Uo from Zg via cd for incomplete closure



(e) Uo from Zg nonlinear for complete closure



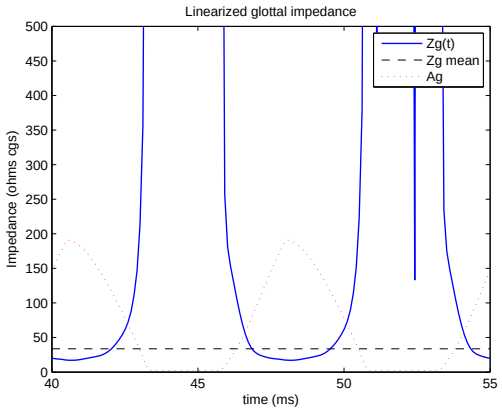
(f) Uo from Zg nonlinear for incomplete closure

Fig. 6.21. Glottal impedance for synthetic vowel /a/ with complete and incomplete glottal closure

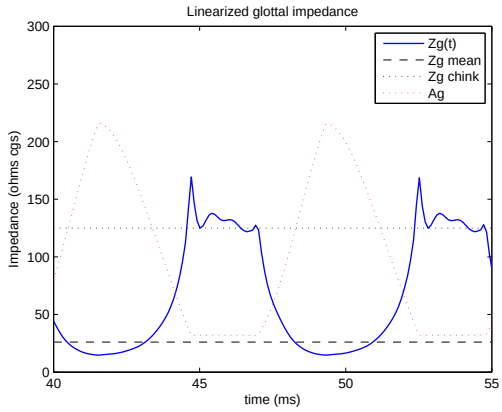
Observations of these patterns illustrated the effect of the gap impedance in the system response, varying from finite values for incomplete closure to orders of magnitude higher for complete closure. The gap cases oscillated around the point of the equivalent gap impedance, which indicated a behavior similar to that of a parallel impedance. This finding was interesting, as the expressions used for these experiments did not explicitly assumed this parallel configuration.

The unstable behavior from the nonlinear case during complete closure appeared evident in the temporal representation of its impedance. This is a product of the nonlinear impedance being smaller by a factor of two with respect to the linearized k_t case, as noted from the linearization performed previously. Both linearized impedances differed in the degree of asymmetry during the open phase of the cycle, which was related to the effects that they produced on the uncoupled airflows.

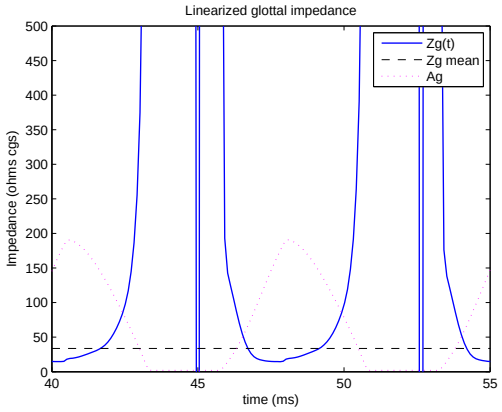
The same uncoupled airflow and temporal patterns presented previously are shown for vowel /i/ in Fig. 6.23 and Fig. 6.24. The same trends as in vowel /a/ were observed, where all cases exhibited de-skewing of the glottal pulses and the estimates that better matched the ideal uncoupled flow were from the nonlinear impedance for both complete and incomplete glottal closure. A more unstable behavior was noted for the complete glottal closure in the linearized impedances, most likely due to the limitation of the CPIF estimates for this vowel. The equivalence to the parallel impedances was again observed for the incomplete glottal closure.



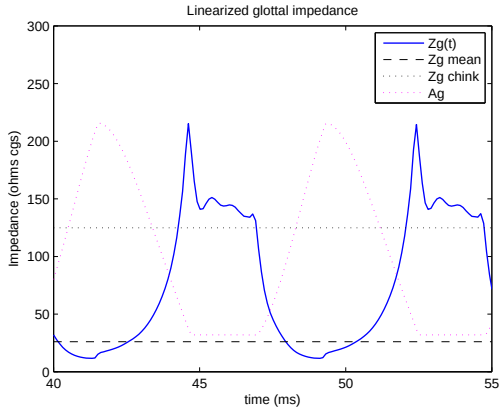
(a) Zg via kt for complete closure



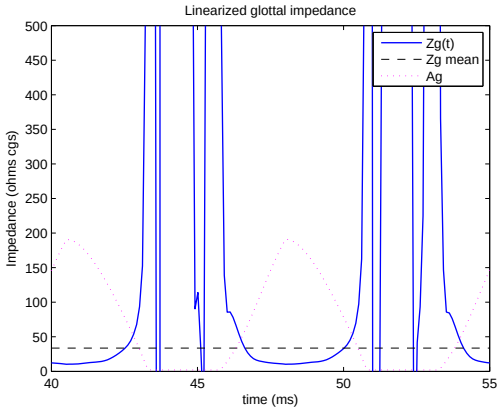
(b) Zg via kt for incomplete closure



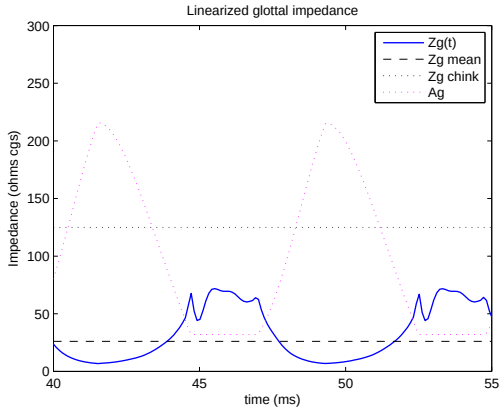
(c) Zg via cd for complete closure



(d) Zg via cd for incomplete closure

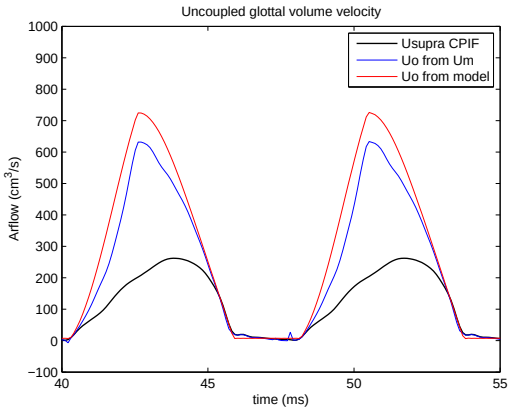


(e) Zg nonlinear for complete closure

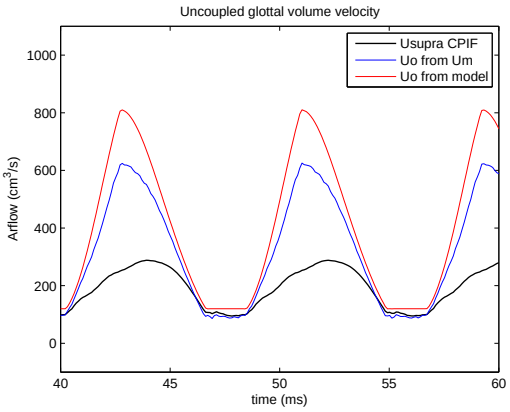


(f) Zg nonlinear for incomplete closure

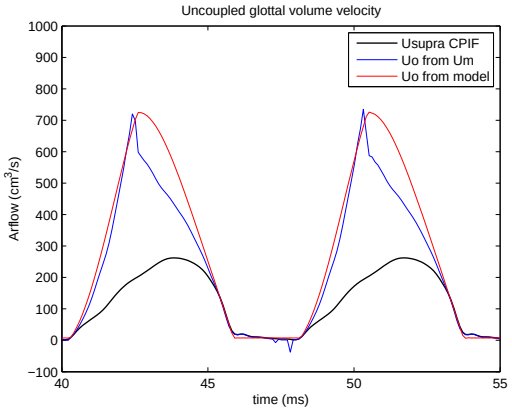
Fig. 6.22. Glottal impedance for synthetic vowel /a/ with complete and incomplete glottal closure



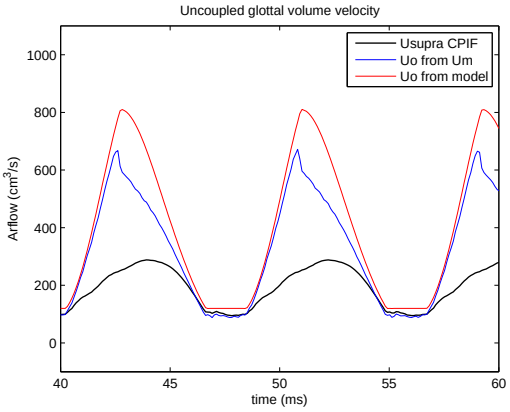
(a) Uo from Zg via kt for complete closure



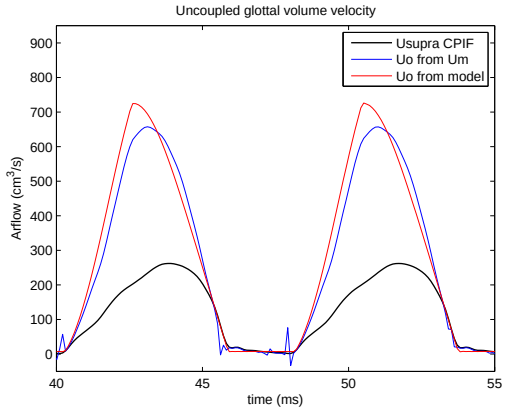
(b) Uo from Zg via kt for incomplete closure



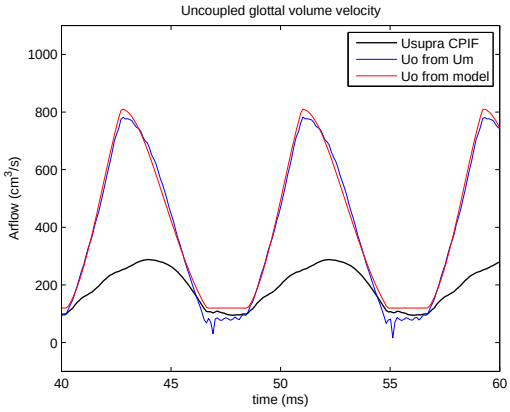
(c) Uo from Zg via cd for complete closure



(d) Uo from Zg via cd for incomplete closure

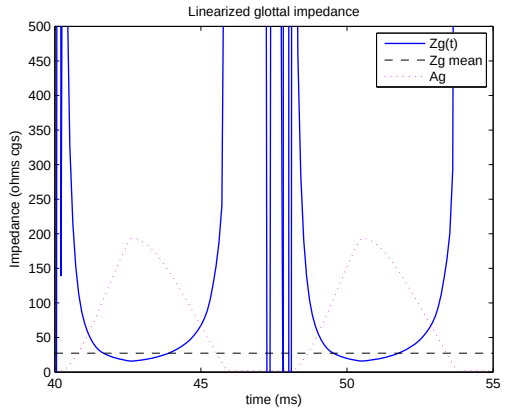


(e) Uo from Zg nonlinear for complete closure

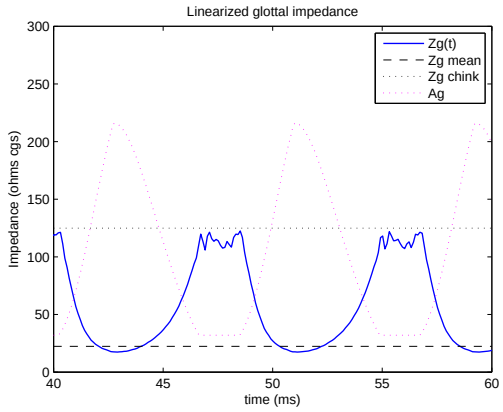


(f) Uo from Zg nonlinear for incomplete closure

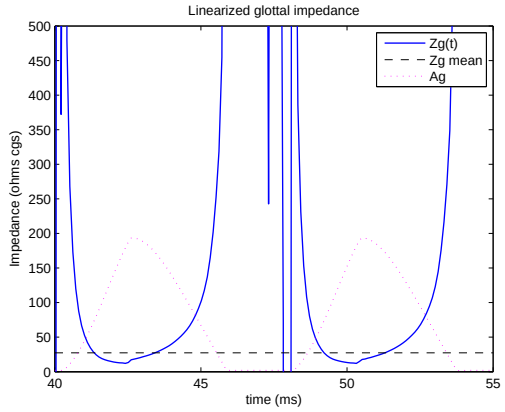
Fig. 6.23. Uncoupled airflow for synthetic vowel /i/ with complete and incomplete glottal closure



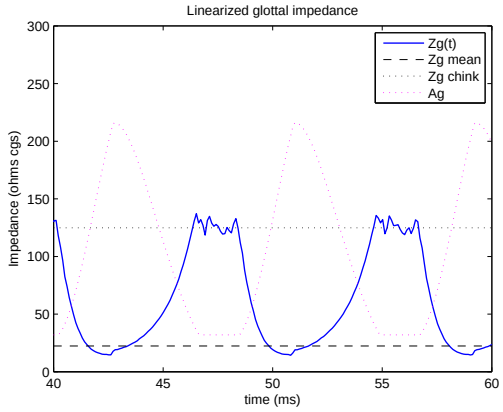
(a) Zg via kt for complete closure



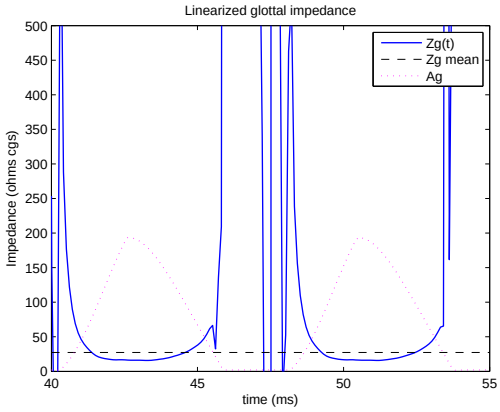
(b) Zg via kt for incomplete closure



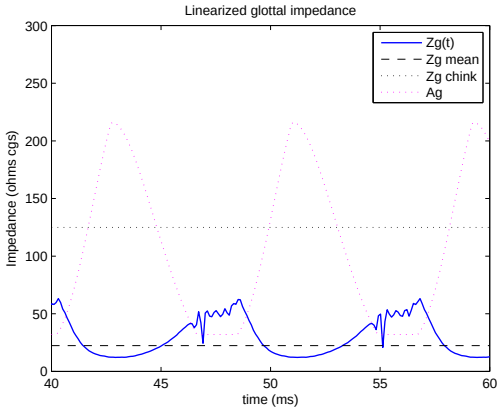
(c) Zg via cd for complete closure



(d) Zg via cd for incomplete closure



(e) Zg nonlinear for complete closure



(f) Zg nonlinear for incomplete closure

Fig. 6.24. Time-varying glottal impedance for synthetic vowel /i/ with complete and incomplete glottal closure

6.3.4 System decoupling: Human subject recordings

The goal of this study component was again to retrieve the ideal source U_o from the dipole model, now from actual recordings of oral airflow. The methods used in this section were the same as before, but U_{supra} was now obtained from CPIF of the CV mask signal and the glottal area was estimated from HSV. The same impedance cases were evaluated. No ideal uncoupled airflow was computed in this case, as no direct measurement of the subglottal pressure was obtained. Thus, the waveforms are presented twice, first normalized and superimposed with the glottal area to evaluate the degree of de-skewing, and then in their absolute units. In addition, estimates of glottal airflow (U_{supra}) obtained from the accelerometer were also used to compute the uncoupled glottal airflow U_o.

To further emphasize the effect of different glottal configurations, the same two vowels (/a/ and /i/) and registers (chest and falsetto) were analyzed for the male "calibrated" subject. It was observed that the chest register recordings had no significant glottal gap from the HSV, but aerodynamic and stroboscopic calibrations illustrated that a non-visible gap was in fact present, most likely due to arytenoid hooding. Thus, all glottal area waveforms were corrected to include an additional gap area using the numerical simulations with incomplete closure as a reference. Nevertheless, the chest voice had a much lower incomplete glottal closure than that of falsetto, for which both cases were still considered to represent normal and abducted closure conditions.

Time-invariant impedance

The normalized CPIF glottal airflow (U_{supra}), estimated uncoupled airflow (U_o from U_m), and glottal area for the time-invariant impedance ($\tilde{Z}_g$) for both vowels and glottal configurations are shown in Fig. 6.25. Given that the normalization visually distorted the signals, the estimates with absolute units for U_o from CPIF and ACC along with U_{supra} for these cases are then shown Fig. 6.26 to correct this phenomenon.

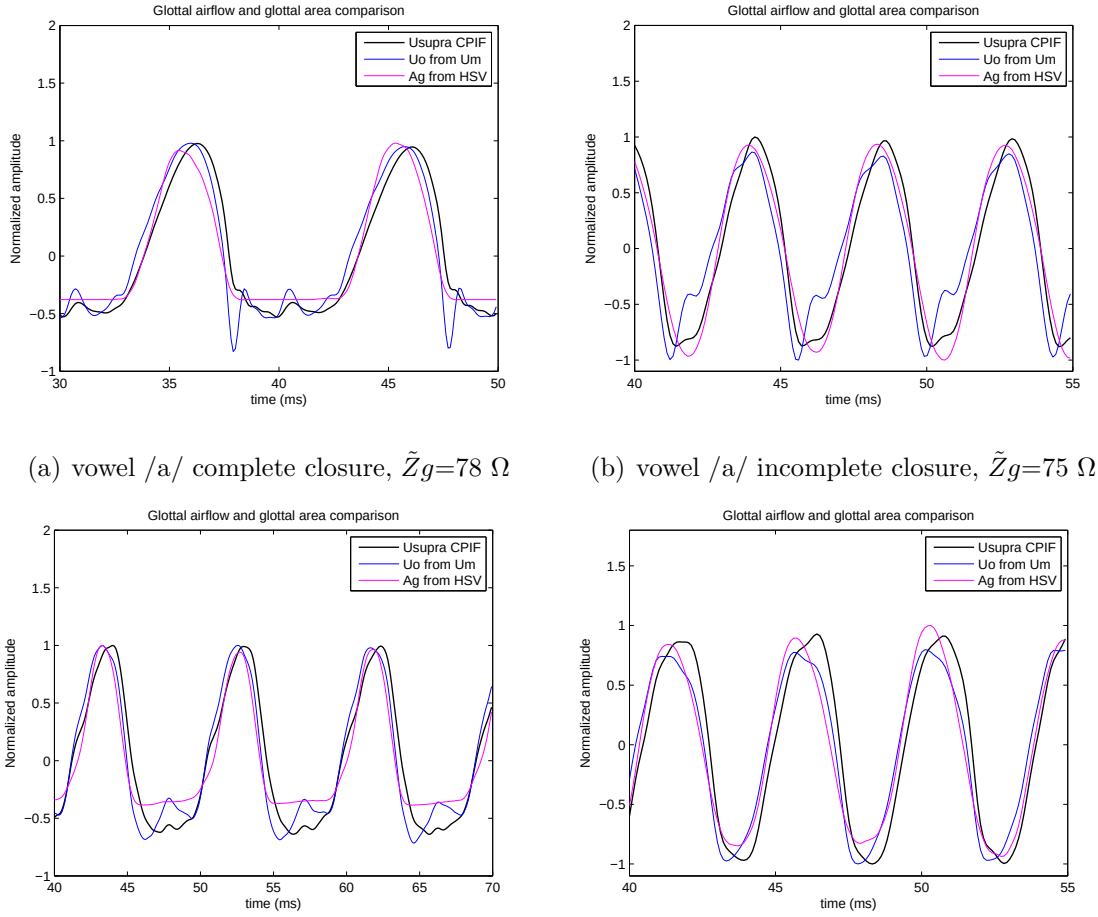
The estimated uncoupled airflows yielded waveforms with fewer distortions compared with the synthetic case, primarily due to the higher mean impedances of all cases in these recordings. The fact that the impedances were higher also implied that less changes were noted with respect to the coupled U_{supra}. No clear de-skewing was observed and artifacts were introduced (negative fluctuations at the point of closure) for vowel /a/ in both registers. Slightly better results were observed for the chest register during vowel /i/, where some degree of skewing was noted, but with undesired fluctuations during the closed phase. In all cases, the estimates from oral flow and acceleration yielded comparable results. Although better than in the synthetics speech case, the time-invariant glottal impedance was again considered limited to describe the effects of source filter interactions.

Time-varying impedance

As before, the normalized CPIF glottal airflow (U_{supra}), estimated uncoupled airflow (U_o from U_m), and glottal area are presented for the two glottal configurations of vowel /a/ in Fig. 6.27, but now for the three time-varying expressions of the glottal impedance described earlier. The same cases with absolute units are then presented in Fig. 6.28 for the same vowel.

It was clear for all chest register cases shown in Fig. 6.21 that the addition of a time-varying expression significantly enhanced the estimation of the uncoupled glottal airflow. These cases exhibited a clear degree of de-skewing of the glottal airflow pulse, being well aligned with the glottal area waveform. The estimates of decoupled airflow for the falsetto register case did exhibit a minor degree of de-skewing, yet less pronounced and with some added artifacts, particularly during the peak and closed phase of the pulse. This was believed to be related to a less significant coupling for cases exhibiting large gap and a larger uncertainty in the impedance estimates.

Among the time-varying glottal impedances, the linearized impedance with a mean k_t factor yielded the best match with respect to the normalized glottal area wave-



(a) vowel /a/ complete closure, $\tilde{Z}g$=78 Ω

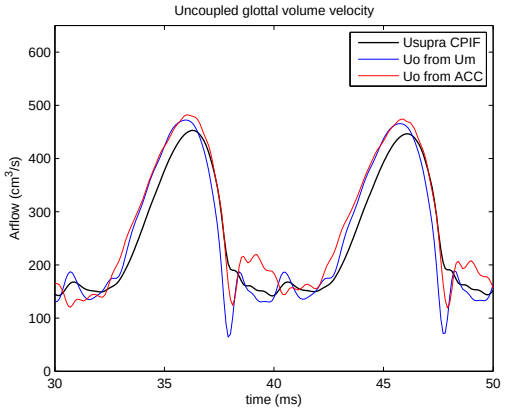
(b) vowel /a/ incomplete closure, $\tilde{Z}g$=75 Ω

(c) vowel /i/ complete closure, $\tilde{Z}g$=95 Ω

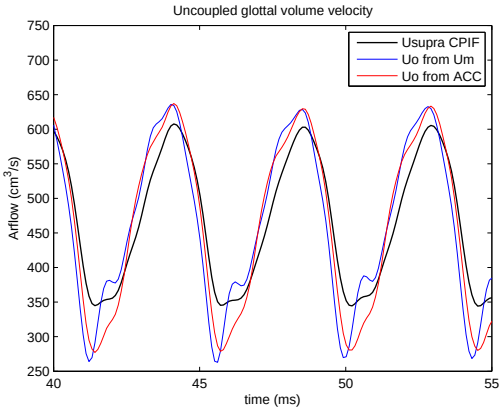
(d) vowel /i/ incomplete closure, $\tilde{Z}g$=81 Ω

Fig. 6.25. Normalized uncoupled glottal airflow with time-invariant impedance from recorded vowels for different glottal configurations. All impedance units are cgs.

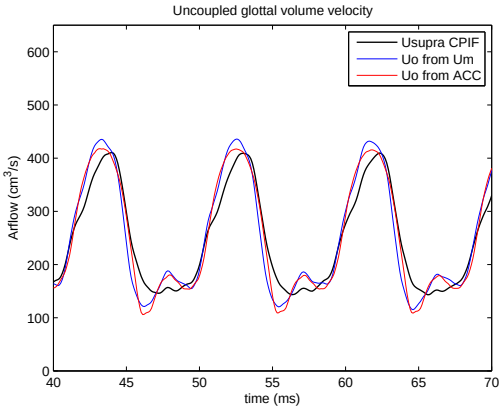
form for both registers. The other two cases exhibited larger deviations during the closing and closed portions of the cycle in both registers. These fluctuations were not expected in the decoupled airflow and were considered estimation errors. The estimates with absolute units from Fig. 6.28 showed that the nonlinear impedance yielded the largest amplitudes, degree of de-skewing, and estimation errors for both registers. The accelerometer-based estimates yielded similar trends but with reduced error, illustrating the potential of the these estimates to represent the glottal aerodynamics.



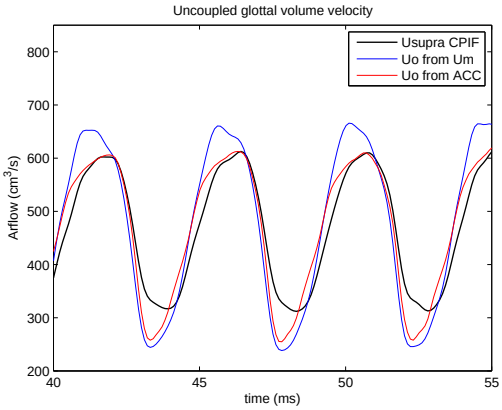
(a) vowel /a/ complete closure, $\tilde{Z}g$=78 Ω



(b) vowel /a/ incomplete closure, $\tilde{Z}g$=75 Ω



(c) vowel /i/ complete closure, $\tilde{Z}g$=95 Ω

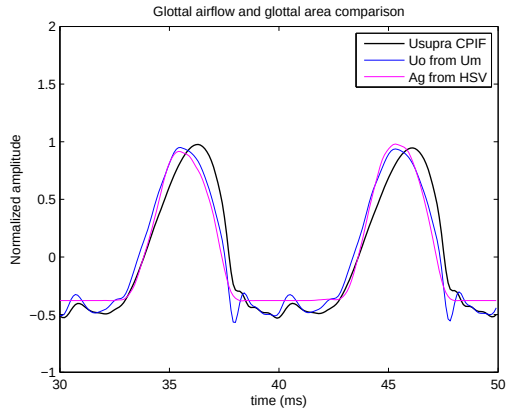


(d) vowel /i/ incomplete closure, $\tilde{Z}g$=81 Ω

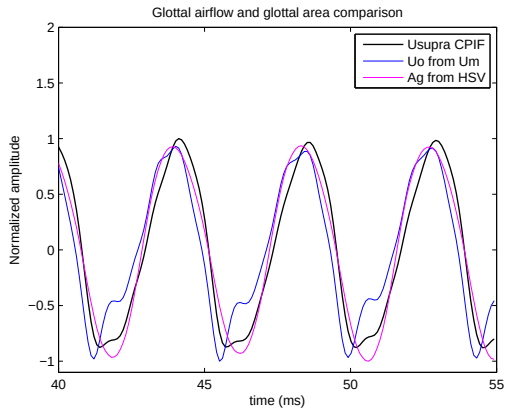
Fig. 6.26. Uncoupled glottal airflow with time-invariant impedance from recorded vowels for different glottal configurations. All impedance units are CGS.

A representation of the temporal patterns of glottal impedance for the speech recordings of vowel /a/ for both registers are presented in Fig. 6.29. As in the synthetic case, the mean impedance $\tilde{Z}_g$, gap impedance, and a trace of the glottal area waveform were shown in dotted lines to facilitate the analysis.

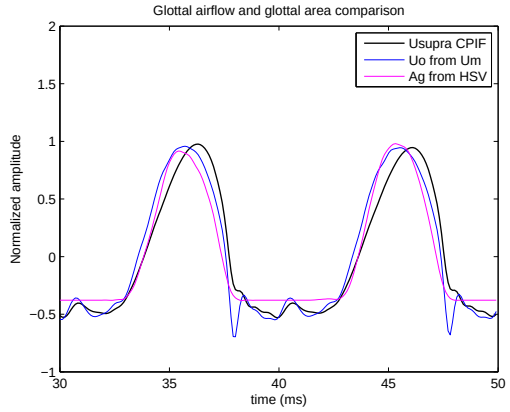
The presence of a DC flow component in normal chest voice significantly reduced the impedance during the closed portion of the cycle when compared with the synthetic case. Interestingly, the difference between the impedances of both registers was considered minor, particularly when evaluating their mean impedances. As in



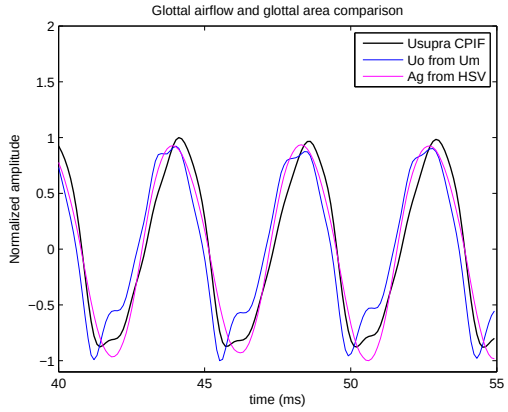
(a) Uo from Zg via kt for chest voice



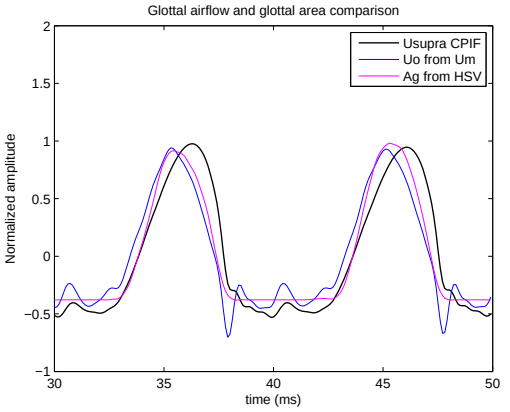
(b) Uo from Zg via kt for falsetto



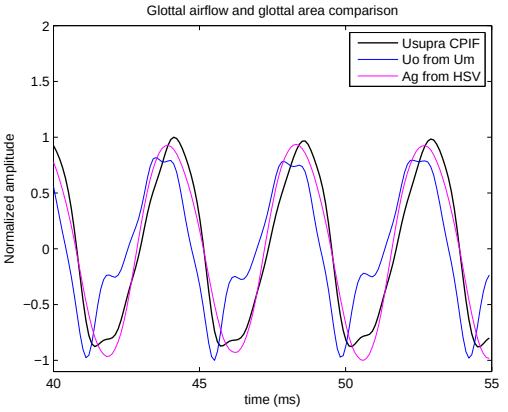
(c) Uo from Zg via cd(t) for chest voice



(d) Uo from Zg via cd(t) for falsetto

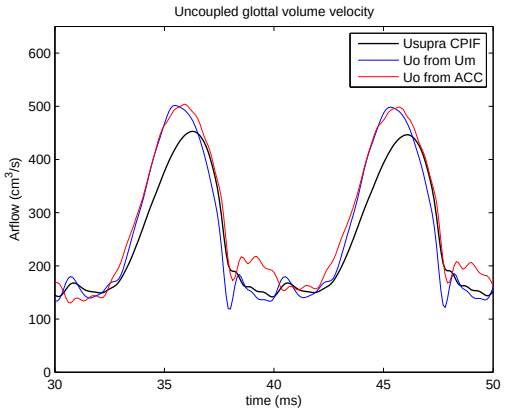


(e) Uo from Zg nonlinear for chest voice

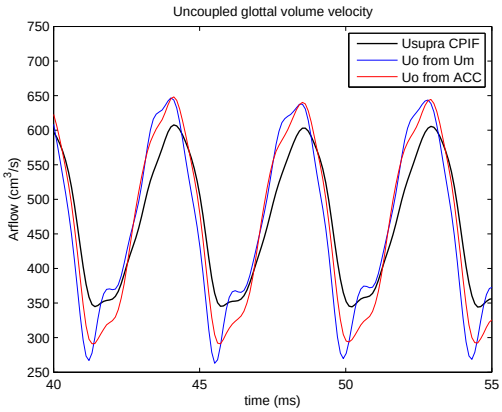


(f) Uo from Zg nonlinear for falsetto

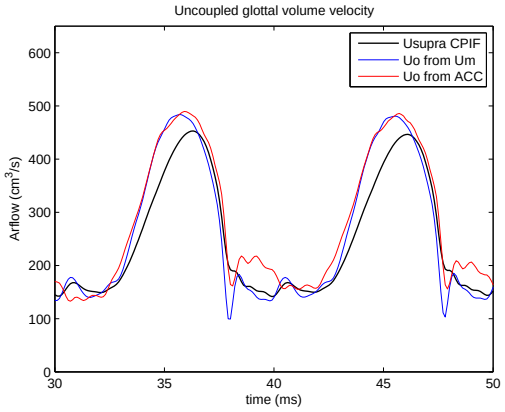
Fig. 6.27. Estimates of uncoupled glottal airflow: Uo vs. Usupra vs. Ag for vowel /a/. All units normalized.



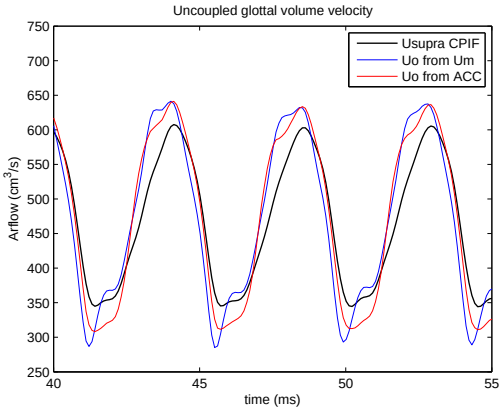
(a) Uo from Zg via kt for chest voice



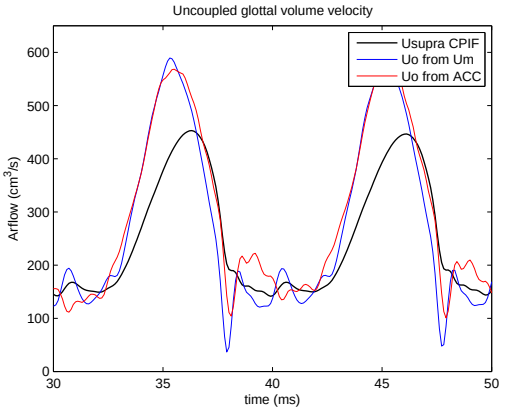
(b) Uo from Zg via kt for falsetto



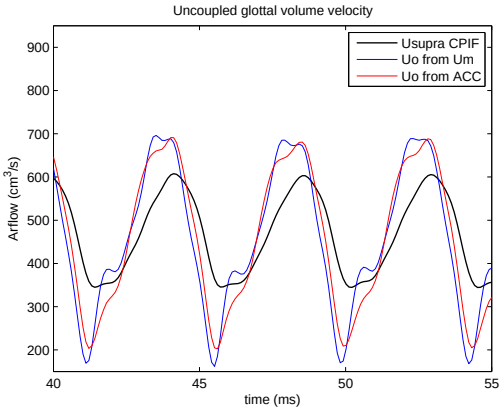
(c) Uo from Zg via cd(t) for chest voice



(d) Uo from Zg via cd(t) for falsetto



(e) Uo from Zg nonlinear for chest voice



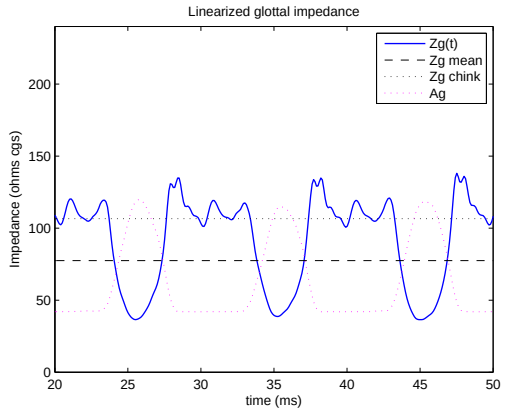
(f) Uo from Zg nonlinear for falsetto

Fig. 6.28. Estimates of uncoupled glottal airflow: Uo vs. Usupra for vowel /a/ from Um and ACC

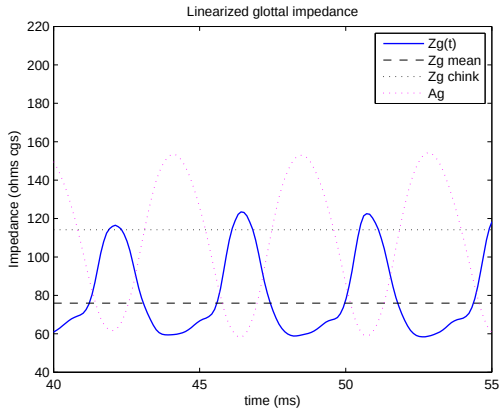
the synthetic speech case, the idea of a parallel gap fits well with both normal and abducted conditions.

The nonlinear impedance was much smaller than the linearized ones and exhibited a more stable behavior with respect to the synthetics speech cases. Both linearized impedance differed in the degree of asymmetry during the open phase of the cycle, which was noted more clearly in the falsetto register case.

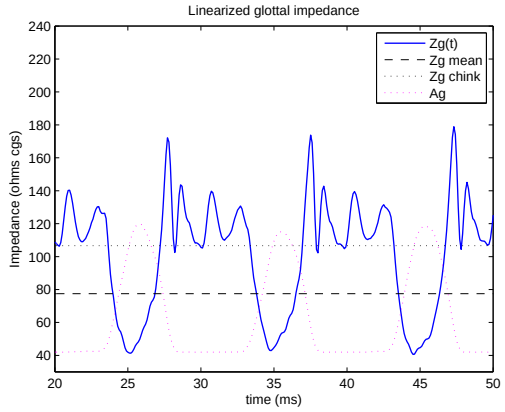
The same uncoupled airflow and temporal patterns presented previously are shown for vowel /i/ in Fig. 6.30, Fig. 6.31, and Fig. 6.32. This vowel exhibited less skewing than vowel /a/, for which less decoupling was needed. De-skewing of glottal pulses was achieved by all types of impedances, but the best fit was obtained by the linearized impedance from the mean factor k_t for both registers. This behavior matched the observations for vowel /a/. The other impedances appeared to over de-skew the glottal pulses and yielded rather asymmetric pulses in the opposite direction. Once again, the amplitude of the uncoupled case with nonlinear impedance was higher than the others, but with large fluctuations during the closing and closed portions of the cycle. The acceleration estimates were comparable and sometimes with less artifacts than those from CPIF for these cases. The equivalence to the parallel impedances and the idea that chest register has slightly higher mean impedance were again observed for all cases in vowel /i/.



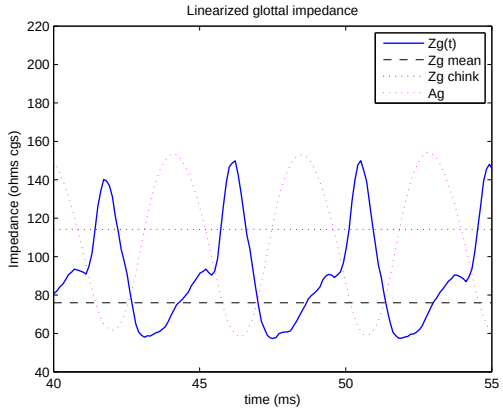
(a) Uo from Zg via kt for chest voice



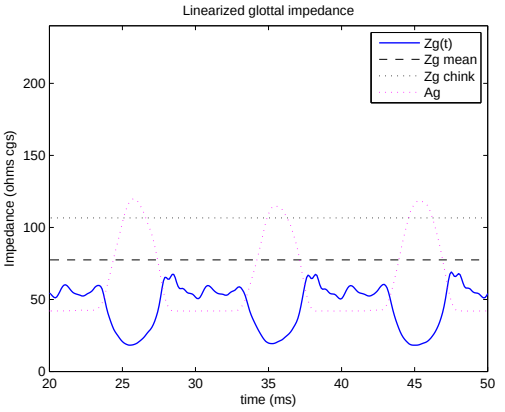
(b) Uo from Zg via kt for falsetto



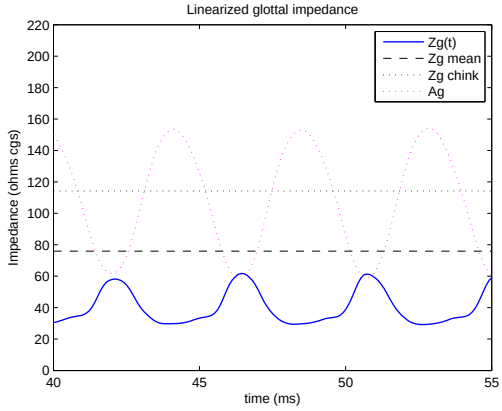
(c) Uo from Zg via cd(t) for chest voice



(d) Uo from Zg via cd(t) for falsetto

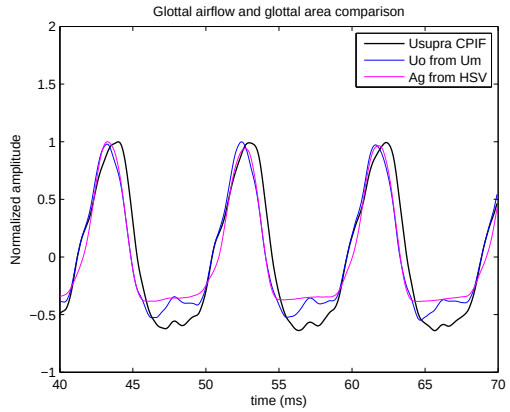


(e) Uo from Zg nonlinear for chest voice

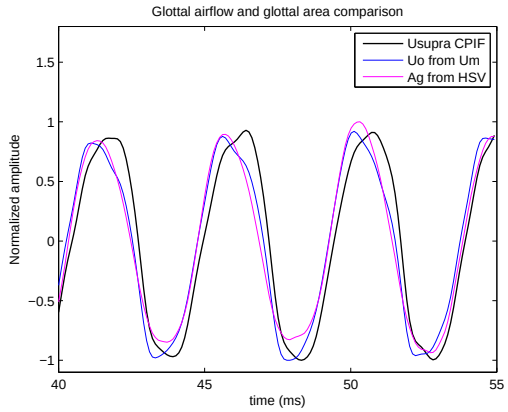


(f) Uo from Zg nonlinear for falsetto

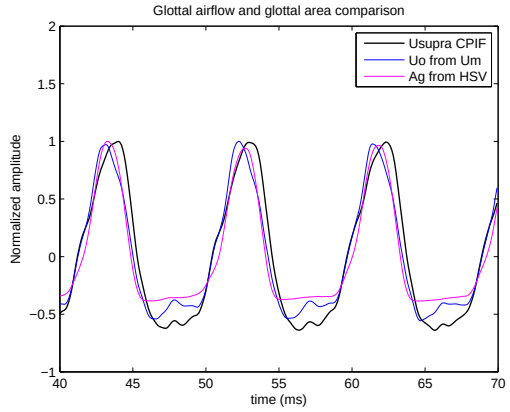
Fig. 6.29. Time-varying glottal impedance for synthetic vowel /a/ for normal and abducted glottal conditions



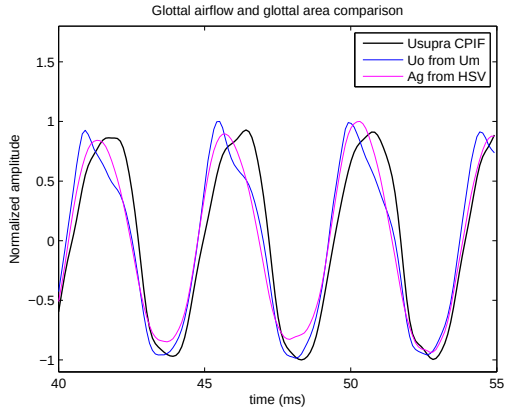
(a) Uo from Zg via kt for chest voice



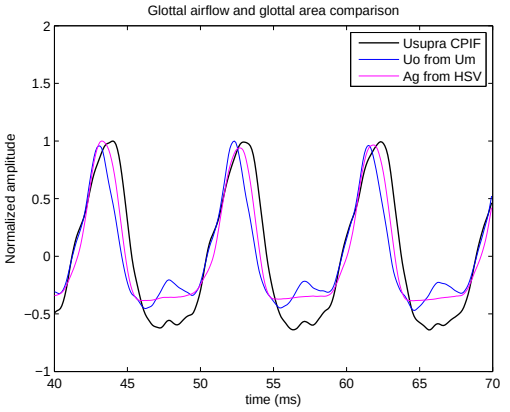
(b) Uo from Zg via kt for falsetto



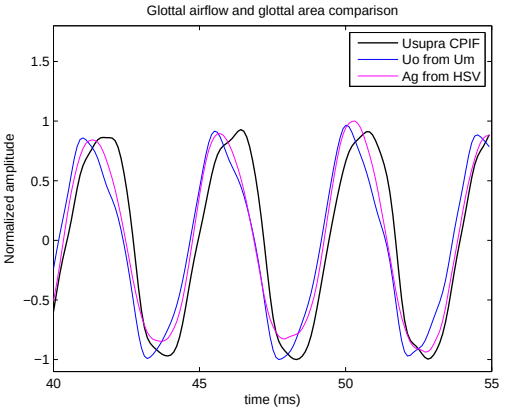
(c) Uo from Zg via cd(t) for chest voice



(d) Uo from Zg via cd(t) for falsetto

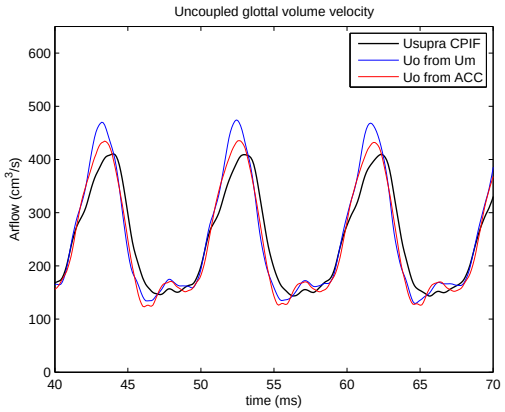


(e) Uo from Zg nonlinear for chest voice

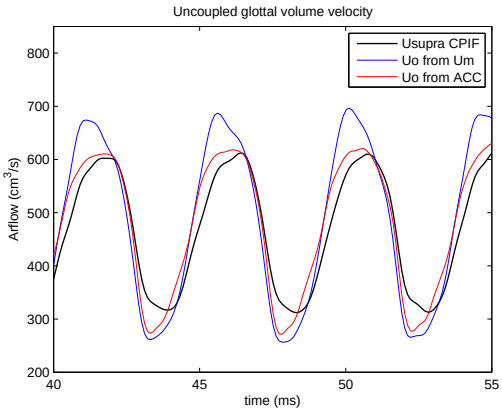


(f) Uo from Zg nonlinear for falsetto

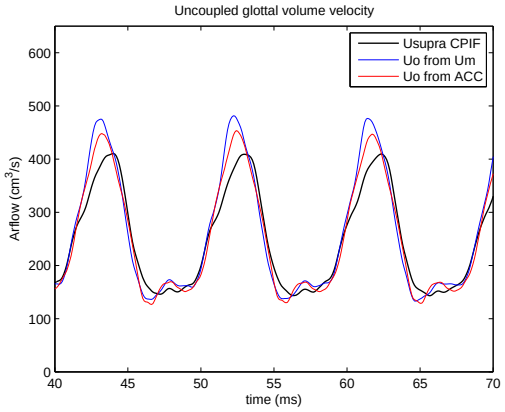
Fig. 6.30. Estimates of uncoupled glottal airflow: Uo vs. Usupra vs. Ag for vowel /i/. All units normalized.



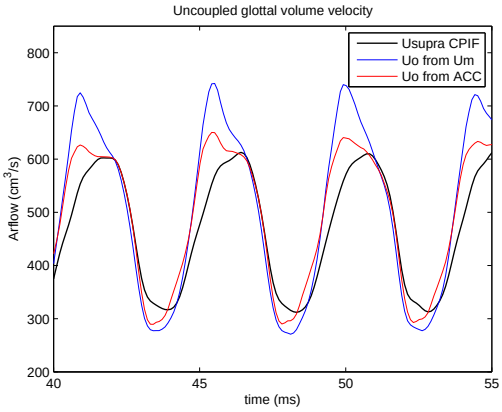
(a) Uo from Zg via kt for chest voice



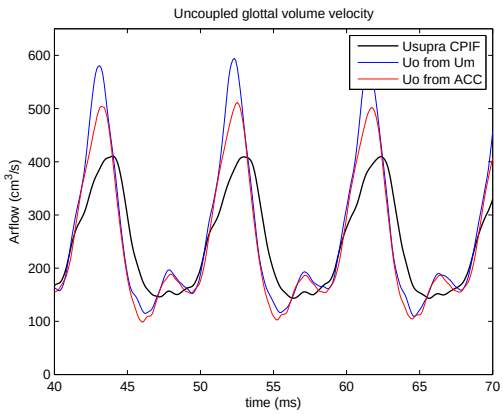
(b) Uo from Zg via kt for falsetto



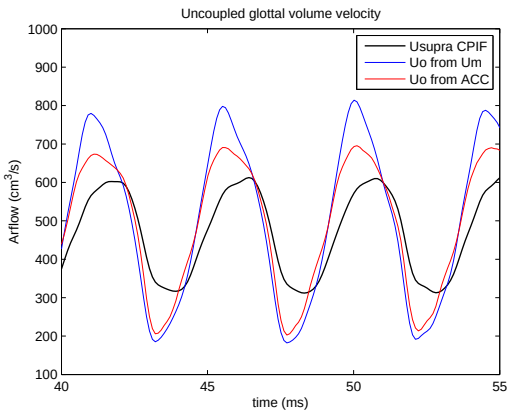
(c) Uo from Zg via cd(t) for chest voice



(d) Uo from Zg via cd(t) for falsetto

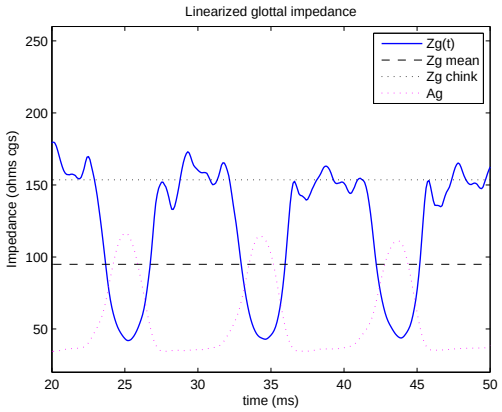


(e) Uo from Zg nonlinear for chest voice

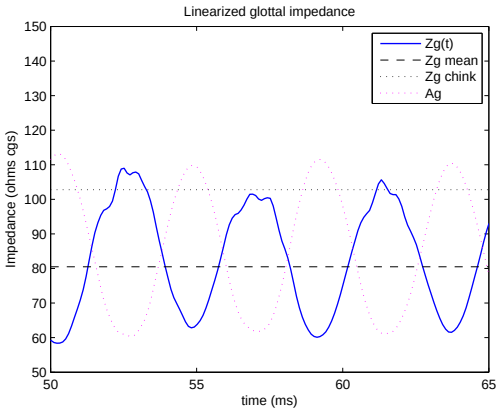


(f) Uo from Zg nonlinear for falsetto

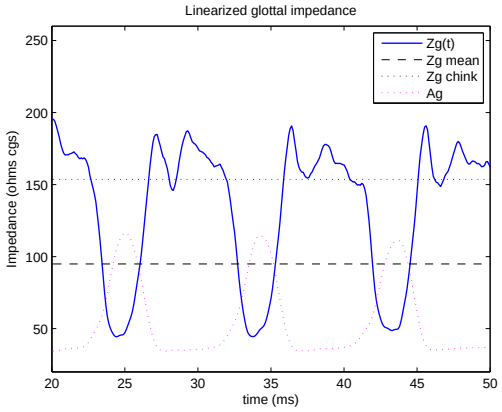
Fig. 6.31. Estimates of uncoupled glottal airflow: Uo vs. Usupra for vowel /i/ from Um and ACC



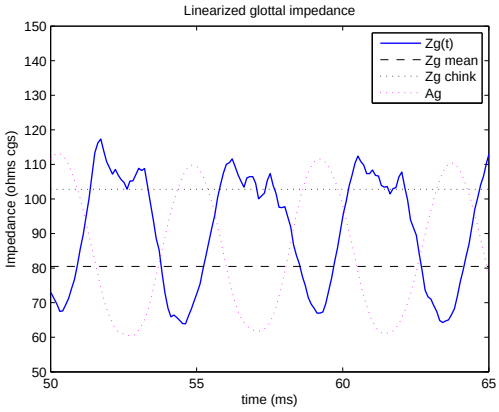
(a) Uo from Zg via kt for chest voice



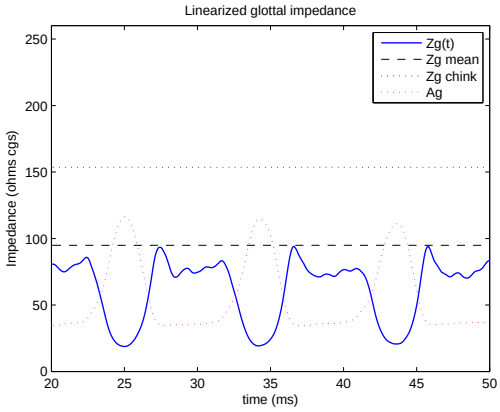
(b) Uo from Zg via kt for falsetto



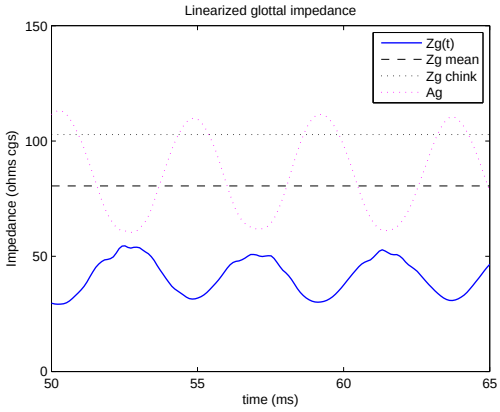
(c) Uo from Zg via cd(t) for chest voice



(d) Uo from Zg via cd(t) for falsetto



(e) Uo from Zg nonlinear for chest voice



(f) Uo from Zg nonlinear for falsetto

Fig. 6.32. Time-varying glottal impedance for synthetic vowel /i/ for normal and abducted glottal conditions

6.4 Discussion

Numerical simulations were applied to evaluate two inverse filtering schemes (CPIF and IBIF) for different glottal conditions and different vowels that had significantly different acoustic loading conditions. Using interactive and physiologically relevant synthetic speech allowed for comparing the inverse filtered versions with that of the "true" glottal airflow provided by the numerical model.

The results of this section indicated that glottal coupling was not required for the estimation of "true" glottal airflow. Both inverse filtering schemes yielded results that had the same desired structure, yet with deviations and estimation errors that have not been noticed before. These errors included differences in the fluctuations during the open and closed phase, but did not affect significantly the estimates of glottal measures. It was noted that the CPIF yielded better estimates than the IBIF applied to the vocal tract in real speech cases, and that both were able to provide good estimates of clinically useful measures. Further exploration of the accuracy of CPIF via numerical modeling is proposed in future studies for additional glottal conditions and pathological cases.

Both inverse filtering techniques were able to remove the effects of the vocal tract for high vowels such as /i/ in real and synthetic speech, a problem that had been generally neglected due to limitations of previous methods. An inspection of inverse filtered waveforms from human speech recordings indicated that the glottal pulse for vowel /i/ did fit well estimates from numerical simulations exhibiting skewing and ripples in the open phase [46, 60, 75]. However, vowel /i/ exhibited less skewing of the glottal pulses, as noted in a more symmetric waveform and reduced MFDR, and also a reduced AC flow. These observations fit with observations from chapter 3 and with the less proportional relation between MFDR and SPL for this vowel, as noted previously through numerical simulations [150]. It was also observed that the degree of glottal airflow pulse skewing appeared to vary with the glottal conditions, where

less significant DC components (i.e., chest register) had more pronounced coupling effects.

Obtaining the "true" glottal airflow via uncoupled inverse filtering allowed a simple yet accurate estimation of glottal aerodynamic parameters from neck surface acceleration through the subglottal branch of the IBIF model. This subglottal inverse filtering yielded very close estimates with respect to the current criterion standard (CPIF of CV mask recordings) of glottal airflow measures commonly used in clinical practice. A particularly low error was found for AC flow and MFDR, both of great clinical interest to detect vocal hyperfunction. Comparing these results with previous efforts [35, 36] elucidated the importance of the proposed transmission line model of the subglottal system, the added compliance of the skin, and the calibration with the oral airflow mask via waveform matching and RMSE minimization. The subglottal IBIF model could be used as an independent entity of the dipole model and applied to the ambulatory monitoring of vocal function. However, changes of the skin properties due to neck movements, certain vowel dependency, and other related factors need further attention, particularly when applying the method for running speech. Future efforts will be devoted to explore the factors that control the changes in the skin properties and potentially optimize single values for the ambulatory assessment of vocal function.

Exploration of inverse filtering using the coupled version of IBIF provided strong evidence that the skewing of the glottal airflow pulses was produced by acoustic coupling. This finding was suggested decades ago through numerical simulations but had not been properly validated using actual speech recordings. Estimates of the uncoupled airflow were obtained from oral airflow and from neck surface acceleration measurements, indicating that both components are interconnected and both tract impedances are needed in the inverse filtering process. For each case, clear deskewing of glottal airflow pulses was observed, particularly in those cases where the skewing was more pronounced. The impedance analysis suggested that cases where a stronger DC flow component was present had slightly lower mean impedances and

less pronounced coupling (i.e., skewing of the glottal airflow pulses). The results of the de-coupling experiments relate the ideal airflow source to the glottal area, as has been suggested in previous studies.

Different types of glottal impedances were studied, and the time-invariant type was found insufficient to represent the coupling phenomenon. Even though the time-varying impedances did not explicitly assumed a parallel gap configuration, the results are well-aligned with such behavior. This finding is consistent with previous representation of incomplete closure as a parallel impedance [14, 15]. Among the time-varying expressions, the nonlinear version yielded the best estimates for synthetic speech while the linearized version that used a mean coefficient k_t yielded the best estimates for actual human speech recordings. However, the nonlinear impedance exhibited more unstable behavior and violated the assumptions of the dipole model. The remaining case, based on linearized glottal impedance with a time-varying orifice discharge coefficient, was not sufficiently accurate during the change from convergent to divergent and yielded distortions. This behavior may be related to the fact that the orifice discharge coefficient term arises from experimental data from a physical model with physiological differences and that it was simply adapted to follow the derivative of the glottal area. Further studies are needed to more accurately adapt time-varying versions of the glottal impedance. The uncertainty of the impedance estimates was primarily affected by the glottal area estimates, as phenomena such as arytenoid hooding and the assumption that a constant calibration feature could severely alter the estimated values. Further work is needed to improve the estimates of this important quantity from laryngeal high-speed video.

6.5 Conclusions

This chapter integrated simulations, experiments, and the knowledge acquired during the complete thesis research. A transmission line based inverse filtering was used to estimate parameters of the glottal airflow from neck surface acceleration

measurements with high accuracy. The same model was used to study the effects of acoustic coupling for different vowels and glottal configurations. The results indicated that accurate estimates of glottal airflow can be obtained from uncoupled inverse filtering models, which facilitates the application of the proposed subglottal inverse-filtering scheme in the ambulatory assessment of vocal function. The effect of coupling on the glottal airflow in human speech recordings indicated that the skewing of the glottal pulse is related to the effects of acoustic coupling. That said, it was also noted that the the degree of coupling was not necessary related to the degree of skewing, as cases with incomplete glottal closure exhibited large coupling but less skewing of the glottal pulses, matching prediction from the numerical simulations in chapter 3. This indication of the importance of coupling was most evident from human speech recordings and their analysis.

7. CONCLUSIONS

The principal aims of this thesis research were to investigate the effects of complete and incomplete glottal closure on source-filter interactions and inverse filtering of speech sounds. Four research components were undertaken to address this goal: 1) numerical simulations of acoustic coupling, 2) experiments of tissue instabilities due to acoustic coupling, 3) biosensing evaluation of neck surface acceleration, and 4) experiments on inverse filtering with and without acoustic coupling. Observations from numerical simulations provided evidence that supported previous studies as well as contributed many new insights on the nature of the coupling during incomplete glottal closure. The effect of acoustic coupling was also observed in human subjects through *in vivo* recordings of vocal fold tissue motion under strong coupling conditions. These findings supported the representation of source-filter interaction based on impedance representations of the system components, for which an impedance-based description was applied to a novel inverse filtering method. Such a scheme allowed for assessment of the effects of acoustic coupling through a variable glottal impedance during the glottal airflow retrieval process. It was found that glottal coupling was not needed to estimate "true" airflow parameters from either oral airflow or neck surface acceleration for complete and incomplete glottal scenarios. This finding, along with a revision of related biosensing issues, highlighted the potential clinical applications of neck surface acceleration for the ambulatory assessment of vocal function. The addition of glottal coupling in the proposed inverse filtering scheme ascertained the skewing of the glottal airflow pulses as an effect of coupling, as observed in real and synthetic speech. In addition, observations on human subjects supported findings from numerical simulations on the effects of incomplete glottal closure.

Numerical simulations of acoustic coupling for complete and incomplete glottal closure illustrated that incomplete glottal closure does alter the interaction and energy transfer during phonation. It was observed that during incomplete glottal closure, mean glottal impedance was reduced and acoustic coupling increased. However, such stronger coupling did not generate significant skewing or ripples in the glottal pulses (effects referred to as harmonic distortion) for this condition. The offset in the glottal area increased the coupling and made it proportional to a transglottal pressure term that was smooth and relatively proportional to the glottal area. This finding could be explained based on Titze's airflow solver [59], which also matched the coupling transfer function principles that described changes in glottal airflow pulses in terms of the system impedances.

Variations of acoustic coupling were noted between vowels, where more than one formant was needed to account for the differences in the interaction, thus supporting the need for a complete tract impedance representation [17]. Although the overall degree of coupling increased during incomplete closure, the differences between vowels were maintained. It was noted that higher MFDR and H1-H2 glottal measures were indicative of the degree of harmonic distortion and a lower AC airflow measure was related to the degree of coupling. These measures exhibited the same trends when evaluated in human subject recordings.

The proposed scheme to incorporate a parallel glottal gap in the self-sustained model was shown to replicate human subject data [39, 189], mimicked previous descriptions of closed phase ripples based on parametric representations [14, 15], and reduced the energy transfer from the fluid to the tissue motion that, when compensated with an increased subglottal pressure, also matched observations of vocal hyperfunction [25]. The parallel gap representation was shown to better mimic the effects introduced by DC offset in normal human speech than other methods associated to abduction and pre-collision. Future studies on numerical modeling are proposed to validate assumptions from the model through acoustical experiments on near field conditions.

The biosensing experiments indicated that bioacoustic sensors designed to capture tissue-borne components (e.g., a light-weight accelerometer) could be corrupted by air-borne noise. Furthermore, the technical specifications of such sensor are altered when they were placed on soft-tissue such as skin. Experiments to account for these changes indicated that the accelerometer used in subsequent experiments was the least sensitive (of the sensors tested) to air-borne corruption. The signal from the sensor was not expected to be corrupted by the subject's own voice, but could be affected under loud background noise conditions. These findings support the general use of an accelerometer for the ambulatory assessment of vocal function. Future research considers optimizing the sensor design with additional coatings and/or air chambers for special cases where vocal assessment may be needed under loud background environments (e.g., singers in concerts).

Experiments on the effects of acoustic coupling on tissue instability suggested that acoustic coupling did affect the tissue motion under strong coupling conditions, as those around sign changes in the reactance of the vocal tract. Instabilities with and without the presence of strong acoustic coupling did exhibit differences in the unstable regime, although both were classified as subcritical Hopf bifurcations. The acoustic coupling forces appeared to affect the anterior-posterior configuration during the strong inertive regime, an effect that disappeared after the reactance change due to a pitch glide crossing a vocal tract formant. These observations supported the use of impedance representations of the complete system, which were latter used for inverse filtering purposes. Future research is proposed to explore similar tasks with a larger pool of subjects and investigate unstable behavior due to acoustic coupling via numerical simulations.

Subsequent investigations integrated ideas and observations from many components of this thesis research. Groundwork on inverse filtering illustrated that closed phase inverse filtering yielded results comparable to those from "true" glottal airflow, particularly when extracting measures of glottal function. A simple variation from the scheme proposed by [32] allowed for inverse filtering of high vowels (e.g., vowel

/i/), which was critical to explore the different effects of acoustic coupling. Results from human subjects observation were in agreement with those where vowel /i/ exhibited lower AC airflow, MFDR, and H1-H2 measures with respect to a vowel /a/. This finding raises the question of which vowel better represents normal speech, and how should an assessment of vocal function best be evaluated during running speech. Future research is needed to address these important issues.

The subglottal, uncoupled impedance-based inverse filtering scheme proposed for the acceleration signal yielded results on a par with those from closed phase inverse filtering. The acceleration inverse filtering had a simple time-invariant implementation that primarily depended on the skin properties to represent subject specific cases. The proposed method enhanced previous vocal system models [35, 36] by improving the subglottal and skin properties descriptions and establishing a calibration method via multiple cross-correlations with oral airflow. Although the scheme was unable to retrieve steady DC components from the glottal airflow due to the AC nature of the sensor, the most significant clinical parameters were related to unsteady measures of glottal airflow. Thus, this approach has a strong potential for application of ambulatory monitoring of vocal function, which will be a topic of future research.

The impedance-based inverse filtering also allowed for exploring the decoupling of the glottal airflow based on lumped expressions of the system impedances. A time-varying version of this decoupling yielded clear glottal pulse de-skewing for all cases. This decoupling provided a comprehensive validation using human subject data on that the pulse skewing is a product of acoustic coupling. The clinical/practical applications of the resulting decoupled airflow remain to be explored. Nevertheless, the decoupling illustrated the inverse filtering did not benefit from the addition of a glottal impedance term, i.e., the "true" glottal flow was that entering to the vocal tract and only the tract geometry was needed to retrieve it. This observation supported representations such as the uncoupled inverse-filtering scheme for the acceleration signal and the closed phase inverse filtering scheme.

LIST OF REFERENCES

LIST OF REFERENCES

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

[2] G. Fant, K. Ishizaka, J. Lindqvist-Gauffin, and J. Sundberg, "Subglottal formants," *Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm*, vol. 13, no. 1, pp. 001–012, 1972.

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

[4] M. Rothenberg, "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology* (K. N. Stevens and M. Hirano, eds.), pp. 305–328, University of Tokyo Press, 1980.

[5] M. Rothenberg, "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.

[6] T. V. Ananthapadmanabha and G. Fant, "Calculation of true glottal flow and its components," *Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm*, vol. 23, no. 1, pp. 001–030, 1982.

[7] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of the glottal flow," *Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm*, vol. 26, no. 4, pp. 1–14, 1985.

[8] G. Fant and Q. Lin, "Glottal source - vocal tract acoustic interaction," *Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm*, vol. 28, no. 1, pp. 013–027, 1987.

[9] M. Rothenberg, "Source-tract acoustic interaction in breathy voice," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control* (I. R. Titze and R. C. Scherer, eds.), pp. 465–481, The Denver Center for the Performing Arts, 1984.

[10] D. H. Klatt and L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among male and female talkers," *J. Acoust. Soc. Am.*, vol. 87, no. 2, pp. 820–856, 1990.

[11] H. M. Hanson and K. N. Stevens, "Sub-glottal resonances in female speakers and their effect on vowel spectra," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, vol. 3, (Stockholm), pp. 182–185, 1995.

[12] H. M. Hanson, "Glottal characteristics of female speakers: Acoustic correlates," *J. Acoust. Soc. Am.*, vol. 101, no. 1, pp. 466–481, 1997.

[13] K. N. Stevens, *Acoustic phonetics*. Cambridge, Mass.: MIT Press, 1st ed., 1998.

[14] B. Cranen and L. Boves, "On subglottal formant analysis," *J. Acoust. Soc. Am.*, vol. 81, no. 3, pp. 734–746, 1987.

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

[16] I. R. Titze, "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice," *J. Voice*, vol. 18, no. 3, pp. 292–298, 2004.

[17] I. R. Titze, "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 2733–2749, 2008.

[18] B. H. Story, I. R. Titze, and E. A. Hoffman, "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 100, no. 1, pp. 537–554, 1996.

[19] K. Honda, H. Hirai, J. Estill, and Y. Tohkura, "Contribution of vocal tract shape to voice quality: MRI data and articulatory modeling," in *Vocal Fold Physiology* (O. Fujimura and M. Hirano, eds.), pp. 23–38, Singular Publishing Group, 1995.

[20] 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.

[21] 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," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 1037–1049, 2006.

[22] B. H. Story, "Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002," *J. Acoust. Soc. Am.*, vol. 123, no. 1, pp. 327–335, 2008.

[23] I. R. Titze and B. H. Story, "Acoustic interactions of the voice source with the lower vocal tract," *J. Acoust. Soc. Am.*, vol. 101, no. 4, pp. 2234–2243, 1997.

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

[25] 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," *J. Speech Hear. Res.*, vol. 32, no. 2, pp. 373–392, 1989.

[26] J. S. Perkell, E. B. Holmberg, and R. E. Hillman, "A system for signal processing and data extraction from aerodynamic, acoustic, and electroglottographic signals in the study of voice production," *J. Acoust. Soc. Am.*, vol. 89, no. 4, pp. 1777–1781, 1991.

[27] I. R. Titze and J. Sundberg, "Vocal intensity in speakers and singers," *J. Acoust. Soc. Am.*, vol. 91, no. 5, pp. 2936–2946, 1992.

[28] T. F. Quatieri and R. J. McAulay, "Shape invariant of time-scale and pitch modification of speech," *IEEE Trans. Audio Speech Lang. Process.*, vol. 40, no. 3, pp. 497–510, 1992.

[29] Y. Stylianou, "Applying the harmonic plus noise model in concatenative speech-synthesis," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 1, pp. 21–29, 2001.

[30] D. G. Childers and H. T. Hu, "Speech synthesis by glottal excited linear prediction," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2026–2036, 1994.

[31] M. D. Plumpe, T. F. Quatieri, and D. A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification," *IEEE Trans. Speech Audio Proces.*, vol. 7, no. 5, pp. 569–586, 1999.

[32] P. Alku, C. Magi, S. Yrttiaho, T. Bäckström, and B. Story, "Closed phase covariance analysis based on constrained linear prediction for glottal inverse filtering," *J. Acoust. Soc. Am.*, vol. 125, no. 5, pp. 3289–3305, 2009.

[33] J. Walker and P. Murphy, "Advanced methods for glottal wave extraction," in *Nonlinear Analyses and Algorithms for Speech Processing* (M. Faundez-Zanuy, ed.), pp. 139–149, Springer Berlin/Heidelberg, 2005.

[34] J. Walker and P. Murphy, "A review of glottal waveform analysis," in *Progress in Nonlinear Speech Processing* (Y. Stylianou, ed.), pp. 1–21, Springer Berlin/Heidelberg, 2007.

[35] H. A. Cheyne, *Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck*. PhD thesis, Harvard-MIT Division of Health Sciences and Technology, 2002.

[36] H. A. Cheyne, H. M. Hanson, R. P. Genereux, K. N. Stevens, and R. E. Hillman, "Development and testing of a portable vocal accumulator," *J. Speech Lang. Hear. Res.*, vol. 46, no. 6, pp. 1457–1467, 2003.

[37] J. G. Švec, I. R. Titze, and P. S. Popolo, "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," *J. Acoust. Soc. Am.*, vol. 117, no. 3 Pt 1, pp. 1386–1394, 2005.

[38] R. E. Hillman, J. T. Heaton, A. Masaki, S. M. Zeitels, and H. A. Cheyne, "Ambulatory monitoring of disordered voices," *Ann. Otol. Rhinol. Laryngol.*, vol. 115, no. 11, pp. 795–801, 2006.

[39] E. B. Holmberg, R. E. Hillman, and J. S. Perkell, "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," *J. Acoust. Soc. Am.*, vol. 84, no. 2, pp. 511–529, 1988.

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

[41] J. B. Park and L. Mongeau, "Experimental investigation of the influence of a posterior gap on glottal flow and sound," *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 1171–1179, 2008.

[42] B. H. Story and I. R. Titze, "Voice simulation with a body-cover model of the vocal folds," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1249–60, 1995.

[43] J. van den Berg, "Myoelastic-aerodynamic theory of voice production," *J. Speech Hear. Res.*, vol. 1, no. 3, pp. 227–244, 1958.

[44] J. van den Berg, "Register problems," *Ann. N.Y. Acad. Sci.*, vol. 155, no. 1, pp. 129–134, 1968.

[45] 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.

[46] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two mass model of the vocal cords," *Bell Syst. Tech.*, vol. J51, pp. 1233–1268, 1972.

[47] M. Rothenberg and S. Zahorian, "Nonlinear inverse filtering technique for estimating the glottal-area waveform," *J. Acoust. Soc. Am.*, vol. 61, no. 4, pp. 1063–1070, 1977.

[48] D. G. Miller and H. K. Schutte, "'Mixing' the registers: Glottal source or vocal tract?," *Folia Phoniatr. Logop.*, vol. 57, pp. 278–291, 2005.

[49] I. R. Titze, T. Riede, and P. Popolo, "Nonlinear source-filter coupling in phonation: Vocal exercises," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. 1902–1915, 2008.

[50] I. R. Titze and A. S. Worley, "Modeling source-filter interaction in belting and high-pitched operatic male singing," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1530–1540, 2009.

[51] K. N. Stevens and A. S. House, "Development of a quantitative description of vowel articulation," *J. Acoust. Soc. Am.*, vol. 27, no. 3, pp. 484–493, 1955.

[52] M. M. Sondhi, "Model for wave propagation in a lossy vocal tract," *J. Acoust. Soc. Am.*, vol. 55, no. 5, pp. 1070–1075, 1974.

[53] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Commun.*, vol. 1, no. 3-4, pp. 199–229, 1982.

[54] 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.

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

[56] J. Liljencrants, *Speech Synthesis with a Reflection-type Line Analog*. PhD thesis, Dept. of Speech Commun. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden, 1985.

[57] M. G. Rahim, *Artificial Neural Networks for Speech Analysis/Synthesis*. New York: Kluwer Academic Publishers, 1994.

[58] B. H. Story, *Physiologically-Based Speech Simulation using an Enhanced Wave-Reflection Model of the Vocal Tract*. PhD thesis, University of Iowa, 1995.

[59] I. R. Titze, "Parameterization of the glottal area, glottal flow, and vocal fold contact area," *J. Acoust. Soc. Am.*, vol. 75, no. 2, pp. 570–580, 1984.

[60] M. Zañartu, L. Mongeau, and G. R. Wodicka, "Influence of acoustic loading on an effective single mass model of the vocal folds," *J. Acoust. Soc. Am.*, vol. 121, no. 2, pp. 1119–1129, 2007.

[61] B. H. Story, A. M. Laukkanen, and I. R. Titze, "Acoustic impedance of an artificially lengthened and constricted vocal tract," *J. Voice*, vol. 14, no. 4, pp. 455–469, 2000.

[62] S. F. Austin and I. R. Titze, "The effect of subglottal resonance upon vocal fold vibration," *J. Voice*, vol. 11, no. 4, pp. 391–402, 1997.

[63] Z. Zhang, J. Neubauer, and D. A. Berry, "The influence of subglottal acoustics on laboratory models of phonation," *J. Acoust. Soc. Am.*, vol. 120, no. 3, pp. 1558–1569, 2006.

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

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

[66] G. R. Wodicka, K. N. Stevens, H. L. Golub, and D. C. Shannon, "Spectral characteristics of sound transmission in the human respiratory system," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1130–1135, 1990.

[67] G. R. Wodicka and D. C. Shannon, "Transfer function of sound transmission in subglottal human respiratory system at low frequencies," *J. Appl. Physiol.*, vol. 69, no. 6, pp. 2126–2130, 1990.

[68] P. Harper, *Respiratory tract acoustical modeling and measurements*. PhD thesis, Purdue University, 2000.

[69] P. Harper, S. S. Kraman, H. Pasterkamp, and G. R. Wodicka, "An acoustic model of the respiratory tract," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 543–550, May 2001.

[70] E. R. Weibel, *Morphometry of the Human Lung*. New York: Springer, first ed., 1963.

[71] K. Horsfield, G. Dart, D. E. Olson, G. F. Filley, and G. Cumming, "Models of the human bronchial tree," *J. Appl. Physiol.*, vol. 31.

[72] H. C. Yeh and G. M. Schum, "Models of human-lung airways and their application to inhaled particle deposition," *Bull. Math. Biol.*, vol. 42, no. 3, pp. 461–480, 1980.

[73] R. H. Habib, R. B. Chalker, B. Suki, and A. C. Jackson, "Airway geometry and wall mechanical properties estimated from subglottal input impedance in humans," *J. Appl. Physiol.*, vol. 77, no. 1, pp. 441–451, 1994.

[74] I. R. Titze, "Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model," *J. Acoust. Soc. Am.*, vol. 111, no. 1, pp. 367–376, 2002.

[75] B. H. Story, "An overview of the physiology, physics and modeling of the sound source for vowels," *Am. Sci. & Tech.*, vol. 23, no. 4, pp. 195–206, 2002.

[76] R. L. Wegel, "Theory of vibration of the larynx," *Bell Syst. Tech. J.*, vol. 9, pp. 207–227, 1930.

[77] R. C. Scherer, I. R. Titze, and J. F. Curtis, "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," *J. Acoust. Soc. Am.*, vol. 73, no. 2, pp. 668–676, 1983.

[78] J. van den Berg, J. T. Zantema, and P. Doornenbal Jr., "On the air resistance and the Bernoulli effect of the human larynx," *J. Acoust. Soc. Am.*, vol. 29, no. 5, pp. 626–631, 1957.

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

[80] N. Binh and J. Gauffin, "Aerodynamic measurements in an enlarged static laryngeal model," *Speech Transmission Laboratory Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm*, vol. 2, no. 3, pp. 315–345, 1983.

[81] I. R. Titze, "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.*, vol. 83, no. 4, pp. 1536–1552, 1988.

[82] F. Alipour and R. C. Scherer, "On pressure-frequency relations in the excised larynx," *J. Acoust. Soc. Am.*, vol. 122, no. 4, pp. 2296–2305, 2007.

[83] 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," *J. Acoust. Soc. Am.*, vol. 102, no. 2, pp. 1121–1133, 1997.

[84] J. B. Park and L. Mongeau, "Instantaneous orifice discharge coefficient of a physical, driven model of the human larynx," *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 442–455, 2007.

[85] 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," *J. Acoust. Soc. Am.*, vol. 96, pp. 3416–3431, 1994.

[86] B. D. Erath and M. W. Plesniak, "The occurrence of the coanda effect in pulsatile flow through static models of the human vocal folds," *J. Acoust. Soc. Am.*, vol. 120, pp. 1000–1011, 2006.

[87] B. D. Erath and M. W. Plesniak, "An investigation of jet trajectory in flow through scaled vocal fold models with asymmetrical glottal passages," *Experiments in Fluids*, vol. 41, pp. 735–748, 2006.

[88] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Prentice-Hall signal processing series, Englewood Cliffs, NJ: Prentice Hall, 1978.

[89] S. Adachi and M. Yamada, "An acoustical study of sound production in biphonic singing, xoomij," *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2920–2932, 1999.

[90] S. M. Lulich, *The role of lower airway resonances in defining vowel feature contrasts.* PhD thesis, Harvard University–MIT Division of Health Sciences and Technology, 2006.

[91] X. Chi and M. Sonderegger, "Subglottal coupling and its influence on vowel formants," *J. Acoust. Soc. Am.*, vol. 122, no. 3, pp. 1735–1745, 2007.

[92] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *J. Acoust. Soc. Am.*, vol. 97, no. 3, pp. 1874–1884, 1995.

[93] J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," *J. Acoust. Soc. Am.*, vol. 110, no. 4, pp. 2120–2128, 2001.

[94] 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*, vol. 16, pp. 023118:1–8, 2006.

[95] C. Tao, Y. Zhang, and J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 794–801, 2007.

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

[97] I. R. Titze and B. H. Story, "Rules for controlling low-dimensional vocal fold models with mucle activation," *J. Acoust. Soc. Am.*, vol. 112, no. 3, pp. 1064–1076, 2002.

[98] J. Kuo, *Voice source modeling and analysis of speakers with vocal-fold nodules.* PhD thesis, Harvard-MIT Division of Health Sciences and Technology, 1998.

[99] M. S. Howe and R. S. McGowan, "On the single-mass model of the vocal folds," *Fluid Dyn. Res.*, vol. 42, pp. 1–16, 2010.

[100] R. S. McGowan and M. S. Howe, "Comments on single-mass models of vocal fold vibration," *J. Acoust. Soc. Am.*, vol. 127, no. 5, pp. EL215–EL221, 2010.

[101] R. S. McGowan, L. L. Koenig, and A. L

[102] 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," *J. Acoust. Soc. Am.*, vol. 117, no. 3, pp. 1362–1372, 2005.

[103] 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," *J. Acoust. Soc. Am.*, vol. 95, no. 6, pp. 3595–3604, 1994.

[104] S. L. Thomson, L. Mongeau, and S. H. Frankel, "Aerodynamic transfer of energy to the vocal folds," *J. Acoust. Soc. Am.*, vol. 118, no. 3, pp. 1689–1700, 2005.

[105] C. Tao and J. J. Jiang, "Anterior-posterior biphonation in a finite element model of vocal fold vibration," *J. Acoust. Soc. Am.*, vol. 120, no. 3, pp. 1570–1577, 2006.

[106] T. Wurzbacher, M.Dllinger, R. Schwarz, U. Hoppe, U. Eysholdt, and J. Lohscheller, "Spatiotemporal classification of vocal fold dynamics by a multimass model comprising time-dependent parameters," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. 2324–2334, 2010.

[107] F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 3003–3012, 2000.

[108] I. R. Titze and F. Alipour, *The myoelastic aerodynamic theory of phonation.* Denver, CO ; Iowa City, IA: National Center for Voice and Speech, 2006.

[109] W. Zhao, C. Zhang, S. H. Frankel, , and L. Mongeau, "Computational aeroacoustics of phonation, part i: Computational methods and sound generation mechanisms," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 2134–2146, 2002.

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

[111] R. W. Chan and I. R. Titze, "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2351–2362, 2006.

[112] N. H. Fletcher, "Autonomous vibration of simple pressure controlled valves in gas flows," *J. Acoust. Soc. Am.*, vol. 93, no. 4, p. 21722180, 1993.

[113] F. Alipour, D. Montequin, and N. Tayama, "Aerodynamic profiles of a hemilarynx with a vocal tract," *Ann. Otol. Rhinol. Laryngol.*, vol. 110, no. 6, pp. 550–555, 2001.

[114] Z. Zhang, J. Neubauer, and D. A. Berry, "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2841–2849, 2006.

[115] L. J. Chen, M. Zanartu, D. C. Cook, and L. Mongeau, "Effects of acoustic loading on the self-oscillations of a synthetic model of the vocal folds," in *Proceedings of the Ninth International Conference on Flow-Induced Vibrations* (Zolotarev and Horacek, eds.), (Prague), pp. 1–6, 2008.

[116] J. S. Drechsel and S. L. Thomson, "Influence of supraglottal structures on the glottal jet exiting a two-layer synthetic, self-oscillating vocal fold model," *J. Acoust. Soc. Am.*, vol. 123, no. 6, pp. 4434–4445, 2008.

[117] M. L. L. Harries, J. M. Walker, D. M. Williams, S. Hawkins, and I. A. Hughesd, "Changes in the male voice at puberty," *Arch. Dis. Child*, vol. 77, pp. 445–447, 1997.

[118] E. Curry, "Voice breaks and pathological larynx conditions," *J. Speech Disord.*, vol. 14, no. 4, pp. 356–358, 1949.

[119] D. A. Berry, H. Herzel, I. R. Titze, and B. H. Story, "Bifurcations in excised larynx experiments," *J. Voice*, vol. 10, no. 2, pp. 129–138, 1996.

[120] J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel, "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes," *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3179–3192, 2001.

[121] J. G. Švec, H. K. Schutte, and D. G. Miller, "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1523–1531, 1999.

[122] D. G. Miller, J. G. Švec, and H. K. Schutte, "Measurement of characteristic leap interval between chest and falsetto registers," *J. Voice*, vol. 16, no. 1, pp. 8–19, 2002.

[123] I. T. Tokuda, J. Horáček, J. G. Švec, and H. Herzel, "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," *J. Acoust. Soc. Am.*, vol. 122, no. 1, pp. 519–531, 2007.

[124] I. T. Tokuda, J. Horáček, J. G. Švec, and H. Herzel, "Bifurcations and chaos in register transitions of excised larynx experiments," *Chaos*, vol. 18, pp. 013102:1–12, 2008.

[125] P. Mergell and H. Herzel, "Modelling biphonation - The role of the vocal tract," *Speech Commun.*, vol. 22, pp. 141–154, 1997.

[126] H. Hatzikirou, W. T. Fitch, and H. Herzel, "Voice instabilities due to source-tract interactions," *Acta Acust. united Ac.*, vol. 92, no. 3, pp. 468–475, 2006.

[127] I. R. Titze, *Principles of voice production*. Iowa City, IA: National Center for Voice and Speech, 2000.

[128] I. T. Tokuda, M. Zemke, M. Kob, and H. Herzel, "Biomechanical modeling of register transitions and the role of vocal tract resonators," *J. Acoust. Soc. Am.*, vol. 127, no. 3, pp. 1528–1536, 2010.

[129] H. Herzel, D. Berry, I. R. Titze, and M. Saleh, "Analysis of vocal disorders with methods from nonlinear dynamics," *J. Speech Hear. Res.*, vol. 37, pp. 1008–1019, 1994.

[130] D. A. Berry, Z. Zhang, and J. Neubauer, "Mechanisms of irregular vibration in a physical model of the vocal folds," *J. Acoust. Soc. Am.*, vol. 120, no. 3, pp. EL36–EL42, 2006.

[131] P. Mergell, W. T. Fitch, and H. Herzel, "Modeling the role of nonhuman vocal membranes in phonation," *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 2020–2028, 1999.

[132] M. Echternach, S. Dippold, J. Sundberg, S. Arndt, M. F. Zander, and B. Richter, "High-speed imaging and electroglottography measurements of the open quotient in untrained male voices' register transitions," *J. Voice*, p. doi:10.1016/jvoice.2009.05.003, 2010. In Press.

[133] J. van den Berg, "Subglottal pressures and vibration of vocal folds," *Folia Phoniatr.*, vol. 9, p. 6571, 1957.

[134] I. R. Titze, "A framework for the study of vocal registers," *J. Voice*, vol. 2, no. 3, pp. 183–194, 1988.

[135] J. Makhoul, "Linear prediction: A tutorial review," in *Proc. of the IEEE*, vol. 63, pp. 561–580, 1975.

[136] D. Y. Wong, J. D. Markel, and A. H. Gray, "Least-squares glottal inverse filtering from the acoustic speech waveform," *IEEE Trans. Acoust Speech Sig. Process.*, vol. 27, no. 4, pp. 350–355, 1979.

[137] J. Deller Jr., "Some notes on closed phase glottal inverse filtering," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 4, pp. 917–919, 1981.

[138] T. F. Quatieri, *Discrete-time speech signal processing: Principles and practice.* Prentice-Hall signal processing series, Upper Saddle River, NJ: Prentice Hall, 2002.

[139] A. Krishnamurthy and D. Childers, "Two-channel speech analysis," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. 34, no. 4, pp. 730–743, 1986.

[140] D. Childers and A. Chieteuk, "Modeling the glottal volume-velocity waveform for three voice types," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 505–519, 1995.

[141] S. Parthasarathy and D. Tufts, "Excitation-synchronous modeling of voiced speech," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. 35, no. 9, pp. 1241–1249, 1987.

[142] M. Frohlich, D. Michaelis, and H. W. Strube, "Simsimultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals," *J. Acoust. Soc. Am.*, vol. 110, no. 1, pp. 479–488, 2001.

[143] Q. Fu and P. Murphy, "Robust glottal source estimation based on joint source-filter model optimization," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 2, pp. 492–501, 2006.

[144] P. Alku, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Communication*, vol. 11, no. 2-3, pp. 109–118, 1992.

[145] R. E. Hillman, W. W. Montgomery, and S. M. Zeitels, "Appropriate use of objective measures of vocal function in the multidisciplinary management of voice disorders," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 5, pp. 172–175, 1997.

[146] D. D. Mehta and R. E. Hillman, "Use of aerodynamic measures in clinical voice assessment," *Perspectives on Voice and Voice Disorders, ASHA Div.3*, vol. 17, no. 3, pp. 14–18, 2007.

[147] D. D. Mehta and R. E. Hillman, "Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods," *Curr. Opin. Otolaryngol. Head Neck. Surg.*, vol. 16, no. 3, pp. 211–215, 2008.

[148] J. G. Švec, J. Sundberg, and S. Hertegøard, "Three registers in an untrained female singer analyzed by videokymography, strobolaryngoscopy and sound spectrography," *J. Acoust. Soc. Am.*, vol. 123, no. 1, pp. 347–353, 2008.

[149] D. D. Mehta, D. D. Deliyski, S. M. Zeitels, M. Zañartu, and R. E. Hillman, "Integration of ultra high-speed color videoendoscopy with time-synchronized measures of vocal function," in *Eastern Section Meeting*, Triological Society, January 2009. Boston, MA.

[150] I. R. Titze, "Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation," *J. Speech Lang. Hear. Res.*, vol. 49, no. 2, pp. 439–447, 2006.

[151] D. G. Childers and C. K. Lee, "Vocal quality factors: Analysis, synthesis, and perception," *J. Acoust. Soc. Am.*, vol. 90, no. 5, pp. 2394–2410, 1991.

[152] E. Buder, *Voice Quality Measurement*, ch. Acoustic analysis of voice quality: A tabulation of algorithms 1902-1990, pp. 119–244. San Diego: Singular, 1st ed., 2000.

[153] R. D. Kent, ed., *The MIT Encyclopedia of Communication Disorders*. Cambridge: MIT Press, 1st ed., 2003.

[154] J. Kreiman and B. Gerratt, "Perception of aperiodicity in pathological voice," *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2201–2211, 2005.

[155] M. Rothenberg, "Correcting low-frequency phase distortion in electroglottograph waveforms," *J. Voice*, vol. 16, no. 1, pp. 32–36, 2002.

[156] J. Sundberg, "Chest wall vibrations in singers," *J. Speech Hear. Res.*, vol. 26, no. 3, pp. 329–340, 1983.

[157] A. Lamarche and S. Ternstrom, "An exploration of skin acceleration level as a measure of phonatory function in singing," *J. Voice*, vol. 22, no. 1, pp. 10–22, 2008.

[158] T. Carroll, J. Nix, E. Hunter, K. Emerich, I. R. Titze, and M. Abaza, "Objective measurement of vocal fatigue in classical singers: A vocal dosimetry pilot study," *Otolary. Head Neck Surg.*, vol. 135, no. 4, pp. 595–602, 2006.

[159] K. N. Stevens, D. N. Kalikow, and T. R. Willemain, "A miniature accelerometer for detecting glottal waveforms and nasalization," *J. Speech Hear. Res.*, vol. 18, no. 3, pp. 594–599, 1975.

[160] R. P. Lippmann, "Detecting nasalization using a low-cost miniature accelerometer," *J. Speech Hear. Res.*, vol. 24, no. 3, pp. 314–317, 1981.

[161] T. F. Quatieri, K. Brady, D. Messing, J. P. Campbell, W. M. Campbell, M. S. Brandstein, C. J. Weinstein, J. D. Tardelli, and P. D. Gatewood, "Exploiting nonacoustic sensors for speech encoding," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 2, pp. 533–544, 2006.

[162] R. T. H. Laennec, *De l'auscultation mdiate, ou, Traité du diagnostic des maladies des poumons et du coeur : fondé principalement sur ce nouveau moyen d'exploration*. Paris: J.-A. Brosson, et J.-S. Chaudé, 1819.

[163] J. Jones, F. L., "Poor breath sounds with good voice sounds. a sign of bronchial stenosis," *Chest*, vol. 93, no. 2, pp. 312–313, 1988.

[164] A. Cohen and A. D. Berstein, "Acoustic transmission of the respiratory system using speech stimulation," *IEEE Trans. Biomed. Eng.*, vol. 38, no. 2, pp. 126–132, 1991.

[165] J. D. Sapira, "About egophony," *Chest*, vol. 108, no. 3, pp. 865–867, 1995.

[166] J. Xu, J. Cheng, and Y. Wu, "A cepstral method for analysis of acoustic transmission characteristics of respiratory system," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 660–664, 1998.

[167] J. E. Huber, "Use of external cueing to treat hypophonia in parkinsons disease." Grant abstract NIH 1R01DC009409-01, May 2008.

[168] H. Vermarien and E. van Vollenhoven, "The recording of heart vibrations: A problem of vibration measurement on soft tissue," *Med. Biol. Eng. Comp.*, vol. 22, pp. 168–178, 1984.

[169] S. S. Kraman, G. A. Pressler, H. Pasterkamp, and G. R. Wodicka, "Design, construction, and evaluation of a bioacoustic transducer testing (BATT) system for respiratory sounds," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1711–1715, 2006.

[170] S. S. Kraman, G. R. Wodicka, G. A. Pressler, and H. Pasterkamp, "Comparison of lung sound transducers using a bioacoustic transducer testing system," *J. Appl. Physiol.*, vol. 101, no. 2, pp. 469–476, 2006.

[171] M. Zañartu, J. C. Ho, S. S. Kraman, H. Pasterkamp, J. E. Huber, and G. R. Wodicka, "Air-borne and tissue-borne sensitivities of acoustic sensors used on the skin surface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 2, 2009. 443-451.

[172] International Organization for Standardization, *Methods for the calibration of vibration and shock pick-ups. Part 11, Testing of transverse vibration sensitivity.* ISO 5347-11-1993, Geneve, Switzerland: International Organization for Standardization, 1st ed., 1993.

[173] International Organization for Standardization, *Methods for the calibration of vibration and shock pick-ups. Part 15, Testing of acoustic sensitivity.* ISO 5347-15-1993, Geneve, Switzerland: International Organization for Standardization, 1st ed., 1993.

[174] H. Pasterkamp, S. S. Kraman, P. D. DeFrain, and G. R. Wodicka, "Measurement of respiratory acoustical signals. comparison of sensors," *Chest*, vol. 104, no. 5, pp. 1518–1525, 1993.

[175] G. R. Wodicka, S. S. Kraman, G. M. Zenk, and H. Pasterkamp, "Measurement of respiratory acoustic signals. effect of microphone air cavity depth," *Chest*, vol. 106, no. 4, pp. 1140–1144, 1994.

[176] S. S. Kraman, G. R. Wodicka, Y. Oh, and H. Pasterkamp, "Measurement of respiratory acoustic signals. effect of microphone air cavity width, shape, and venting," *Chest*, vol. 108, no. 4, pp. 1004–1008, 1995.

[177] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds. advances beyond the stethoscope," *Am. J. Respir. Crit. Care Med.*, vol. 156, no. 3 Pt 1, pp. 974–987, 1997.

[178] S. B. Patel, T. F. Callahan, M. G. Callahan, J. T. Jones, G. P. Graber, K. S. Foster, K. Glifort, and G. R. Wodicka, "An adaptive noise reduction stethoscope for auscultation in high noise environments," *J. Acoust. Soc. Am.*, vol. 103, no. 5 Pt 1, pp. 2483–91, 1998.

[179] H. Pasterkamp, G. R. Wodicka, and S. S. Kraman, "Effect of ambient respiratory noise on the measurement of lung sounds," *Med. Biol. Eng. Comput.*, vol. 37, no. 4, pp. 461–465, 1999.

[180] M. A. Spiteri, D. G. Cook, and S. W. Clarke, "Reliability of eliciting physical signs in examination of the chest," *Lancet*, vol. 1, no. 8590, pp. 873–875, 1988.

[181] J. Cayley, W. E., "Diagnosing the cause of chest pain," *Am. Fam. Physician.*, vol. 72, no. 10, pp. 2012–21, 2005.

[182] M. Zañartu, "Influence of acoustic loading on the flow-induced oscillations of single mass models of the human larynx," Master's thesis, Purdue University, 2006.

[183] B. Cranen and L. Boves, "On the measurement of glottal flow," *J. Acoust. Soc. Am.*, vol. 84, no. 3, pp. 888–900, 1988.

[184] B. Cranen and J. Schroeter, "Physiologically motivated modelling of the voice source in articulatory analysis/synthesis," *Speech Communication*, vol. 19, no. 1, pp. 1–19, 1996.

[185] B. Cranen and L. Boves, "Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production," *J. Acoust. Soc. Am.*, vol. 77, no. 4, pp. 1543–1551, 1985.

[186] 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," *J. Acoust. Soc. Am.*, vol. 122, no. 4, pp. 2270–2278, 2007.

[187] B. D. Erath, *An Experimental and Theoretical Assessment of Flow Asymmetries in Normal and Pathological Speech*. PhD thesis, Purdue University, 2009.

[188] D. Cook and M. Zañartu, "Toward patient-specific vocal fold models: Objective determination of lumped vocal fold model parameters from continuum vocal fold models," in *International Conference on Voice Physiology and Biomechanics*, (Wisconsin, Madison), 2010.

[189] 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," *J. Acoust. Soc. Am.*, vol. 96, no. 2, pp. 695–698, 1994.

[190] J. B. Kobler, R. E. Hillman, S. M. Zeitels, and J. Kuo, "Assessment of vocal function using simultaneous aerodynamic and calibrated videostroboscopic measures," *Ann. Otol. Rhinol. Laryngol.*, vol. 107, no. 6, pp. 477–485, 1998.

[191] W. Y. Ma and B. S. Manjunath, "Edgeflow: A technique for boundary detection and image segmentation," *IEEE Trans. Imag. Process.*, vol. 9, no. 8, pp. 1375–1388, 2000.

[192] V. Osma-Ruiz, J. I. Godino-Llorente, N. Saenz-Lechon, and R. Fraile, "Segmentation of the glottal space from laryngeal images using the watershed transform," *Comput. Med. Imaging Graph.*, vol. 32, no. 3, pp. 193–201, 2008.

[193] J. F. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.

[194] J. G. Švec, F. Šram, and H. K. Schutte, "Videokymography in voice disorders: What to look for?," *Ann. Otol. Rhinol. Laryngol.*, vol. 116, no. 3, pp. 172–180, 2007.

[195] H. S. Bonilha, D. D. Deliyski, and T. T. Gerlach, "Phase asymmetries in normophonic speakers: Visual judgments and objective findings," *Am. J. Speech Lang. Pathol.*, vol. 17, pp. 367–376, 2008.

[196] D. D. Mehta, D. D. Deliyski, S. M. Zeitels, and R. E. Hillman, "Voice production mechanisms following phonosurgical treatment of early glottic cancer," *Ann. Otol. Rhinol. Laryngol.*, vol. 119, no. 1, pp. 1–9, 2010.

[197] J. Lohscheller, U. Eysholdt, H. Toy, and M. Döllinger, "Phonovibrography: Mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics," *IEEE Trans. Med. Imag.*, vol. 27, no. 3, pp. 300–309, 2008.

[198] J. A. Fiz, J. Gnitecki, S. S. Kraman, G. R. Wodicka, and H. Pasterkamp, "Effect of body position on lung sounds in healthy young men," *Chest*, vol. 133, no. 3, pp. 729–736, 2008.

[199] Acoustical Society of America and American National Standards Institute, *American national standard specification for octave-band and fractional-octave-band analog and digital filters*. ANSI:S1.11-1995 (R2004)), New York, N.Y.: Acoustical Society of America and American National Standards Institute, 1986.

[200] Acoustical Society of America and American National Standards Institute, *Criteria for evaluating room noise*. ANSI:S12.2-1995 (R1999), New York, N.Y.: Acoustical Society of America and American National Standards Institute, 1995.

[201] Acoustical Society of America and American National Standards Institute, *Microphone-in-real-ear and acoustic test fixture methods for the measurement of insertion loss of circumaural hearing protection devices*. ANSI:S12.42-1995 (R2004), New York, N.Y.: Acoustical Society of America and American National Standards Institute, 1995.

[202] J. G. Švec, P. S. Popolo, and I. R. Titze, "Measurement of vocal doses in speech: Experimental procedure and signal processing," *Logoped. Phoniatr. Vocol.*, vol. 28, no. 4, pp. 181–192, 2003.

[203] P. J. Lynch, "Lungs-simple diagram of lungs and trachea." Creative Commons Attribution 2.5 License 2006, December 2006.

[204] J. B. Kobler, D. I. Rosen, J. A. Burns, L. M. Akst, M. S. Broadhurst, S. M. Zeitels, and R. E. Hillman, "Comparison of a flexible laryngoscope with calibrated sizing function to intraoperative measurements," *Ann. Otol. Rhinol. Laryngol.*, vol. 115, no. 10, pp. 733–740, 2006.

[205] B. H. Story, "Technique for tuning vocal tract area functions based on acoustic sensitivity functions," *J. Acoust. Soc. Am.*, vol. 119, no. 2, pp. 715–718, 2006.

[206] H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms," *IEEE Trans. Audio Electroacoust.*, vol. 21, no. 5, pp. 417– 427, 1973.

[207] J. Schroeter and M. M. Sondhi, "Techniques for estimating vocal-tract shapes from the speech signal," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 1, pp. 113–150, 1994.

[208] I. Sanchez and H. Pasterkamp, "Tracheal sound spectra depend on body height," *Am Rev Respir Dis*, vol. 148, no. 4-1, pp. 1083–1087, 1993.

[209] K. Ishizaka and J. L. Flanagan, "Direct determination of vocal-tract wall impedance," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. ASSP-23, pp. 370–373, 1975.

[210] E. K. Franke, "Mechanical impedance of the surface of the human body," *J. Appl. Physiol.*, vol. 3, no. 10, pp. 582–590, 1951.

VITA

VITA

Before initiating his doctoral studies in electrical and computer engineering at Purdue University, Mr. Matías Zañartu Salas received the B.S. degree and professional title in acoustical engineering from Universidad Tecnológica Vicente Pérez Rosales, Santiago, Chile, in 1996 and 1999, respectively, and the M.S. degree in electrical and computer engineering in 2006 from Purdue University, West Lafayette, IN. Since 2006, he was a Research Assistant at Purdue University. During 2008, he was a Visiting Researcher at the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston. In 2009 he was a Graduate Intern Technical at the Physical Technologies Laboratory from Intel Corporation. His current research interests include modeling of nonlinear systems, biomedical acoustics, speech production, and biomedical signal and image processing.

Among other distinctions, Mr. Zañartu was the recipient of a Fulbright Scholarship, an Institute of International Education IIE-Barsa Scholarship, a Qualcomm Q Award of Excellence, and a Best Student Paper in Speech Communication in the 157th meeting of the Acoustical Society of America in Portland, 2009. He is a member of the Acoustical Society of America, the Institute of Electrical and Electronics Engineers, the Biomedical Engineering Society, the International Speech Communication Association, and the Society of Hispanic Professional Engineers.

Following his doctoral studies, Mr. Zañartu will begin a post-doctoral appointment in a joint project between Purdue University and Intel Corporation on speech enhancement for automatic speech recognition. His future plans include pursuing a career in academia in his home country, Chile.