A Three-Dimensional Model of the Larynx and the Laryngeal Constrictor Mechanism: Visually Synthesizing Pharyngeal and Epiglottal Articulations Observed in Laryngoscopy

by

Scott Moisik
B.A., University of Calgary, 2006

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

A Three-Dimensional Model of the Larynx and the Laryngeal Constrictor Mechanism:
Visually Synthesizing Laryngeal and Pharyngeal Articulations Observed in Laryngoscopy

by

Scott Moisik
B.A., University of Calgary, 2006

Supervisory Committee:

Dr. John H. Esling, Supervisor
(Department of Linguistics)

Dr. Sonya Bird, Departmental Member
(Department of Linguistics)

Dr. George Tzanetakis, Outside Member
(Department of Computer Science)

Dr. Peter Driessen, External Examiner
(Department of Electrical and Computer Engineering)

Supervisory Committee:

Dr. John H. Esling, Supervisor
(Department of Linguistics)

Dr. Sonya Bird, Departmental Member
(Department of Linguistics)

Dr. George Tzanetakis, Outside Member
(Department of Computer Science)

Dr. Peter Driessen, External Examiner
(Department of Electrical and Computer Engineering)

ABSTRACT

This thesis documents the creation of a three-dimensional model of the larynx. The focus is on synthesizing the movement and appearance of laryngeal and pharyngeal sounds, with the intention of elucidating the physiological performance required of the larynx to produce these articulations. The model serves three primary purposes: the analysis of laryngeal articulation, an interactive tool for learning about linguistically relevant anatomy, and a foundation for future modeling developments such as acoustic synthesis.

There are two methodological topics of discussion concerning the techniques used to generate the three-dimensional model of the larynx. The first concerns the morphological aspect of the laryngeal architecture. Laryngeal structures were segmented from a series of histological images using a process known as vertex tracing to generate wire-frame computer representations, or meshes, of the laryngeal structures featured in the model. The meshes were then carefully placed within the three-dimensional space used to generate a scene of the larynx that could be rendered and presented to the user of the program. Frame hierarchies, an organization scheme for vertices, were imposed on flexible tissue meshes to attach and manipulate various moving structures found in the larynx. Finally, basic mechanical features of laryngeal movement derived from research into the biomechanics of laryngeal physiology were implemented.

The second methodological topic pertains to the analysis of laryngoscopic videos to obtain data that describes the movement patterns used to generate the laryngeal and pharyngeal articulations of interest. There are three image analysis techniques applied to the laryngoscopy. The first uses normal speed laryngoscopy to assess end-state articulations, by comparing various geometrical aspects of laryngeal landmarks as they differ between the maximally open setting (used for deep inspiration), and the articulatory target setting. With this technique, various phonation types and segmental articulations are assessed using videos of a phonetician carefully performing the articulations. Some comparison of these articulations to their analogues in the speech of native speakers from various languages is made for the sake of illustration and verification. The second image analysis technique used is applied to high-speed laryngoscopic video of aryepiglottic trilling, which is an important function of the laryngeal constrictor mechanism. The left and right aryepiglottic apertures during trilling are analyzed using binary-conversion and area measurement. The third technique takes the same high-speed laryngoscopic video of aryepiglottic trilling and extracts motion vectors between frame pairs to characterize the directionality and magnitude of motion occurring for each of the folds.

Using the image analysis data, model movements are constrained and synchronized to recreate the articulations observed in the laryngoscopic videos. One of the major innovations of this model is a biomechanical simulation of aryepiglottic fold trilling, based primarily upon the data collected from the high-speed laryngoscopic videos. Overall the model represents one of the first attempts to visually recreate laryngeal articulatory function in a way that is dynamic and interactive. Future work will involve dynamic acoustic synthesis for laryngeal states represented by the model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

Acknowledgements

I am fortunate to have had the opportunity to learn, be challenged, and ultimately create a 3D articulatory model, and this thesis, which documents it. The circumstances that gave rise to this project and the actual task of model creation have been buttressed by the unyielding support of numerous individuals, both in the past and the present, to whom I am deeply grateful.

Most immediately, I owe thanks to my loving fiancée, Carly Jaques, who has been a part of the thesis tumult first hand, and stood by my side in support, despite the stormy seas. Her belief in me helped the project materialize in to something substantial. Then there is my brother, Ryan Moisik, who has had a profound and positive impact on my personality, and helped me through with the power of laughter and absurd humour. I also owe much thanks to my parents, Susan & David Moisik, for their support throughout this thesis, and my entire life. All of the pragmatic problems that arose were brooked more easily because of my mother's expertise in life and University matters. I owe thanks to my father for instilling in me a great appreciation for sculpture, architecture, philosophy, and the power of representation. Finally, I wish to thank my Ukrainian Baba, Grams Moisiuk, for her thrift in saving money to help me pay for my Undergraduate Degree, and her constant unconditional love (and reminding me not to spray deodorant in my eyes).

The thanks I have that go to an academic tune are as follows. First, I want to thank Dr. Michael Dobrovolsky, who saw a phonetician in me and taught me the beauty of sound, both linguistic and musical. I am eternally in awe of Michael as a professor, mentor, and friend; I thank him for encouraging me to go to study with Dr. John Esling. The Department of Linguistics at the University of Calgary was a wonderfully supportive environment to grow as an underling in linguistics, and there are a number of people that helped me thrive. Dr. Darin Howe/Flynn, Dr. John Archibald and Dr. Susan Carroll all gave me the opportunity to be more than just a student, but a researcher as well. I also am grateful to Dr. Elizabeth Ritter for helping me write linguistically. A big thank-you goes out the Linda Toth, the super-powered administrative wonder woman of the U of C Linguistics department and Corey Telfer, for all of our stimulating discussions.

At the University of Victoria, I have been in the constant company of stimulating, passionate, and supportive people, with a fair share of comedians as well. I have had the unique privilege to work with Dr. John Esling, my supervisor, who has been pivotal in the development

of the 3D model and the articulatory theory behind it. With his excellent guidance and insights he has helped me to further myself, both professionally and intellectually. I also wish to thank Dr. Hua Lin, who has allowed me to explore challenging academic territories and been fully supportive on a personal and academic level. I also wish to thank Dr. Suzanne Urbancyzk, Dr. Sonya Bird, Dr. Ewa Czaykowska-Higgins, Dr. Tae-Jin Yoon and my committee member and external examiner, Dr. George Tzanetakis and Dr. Peter Driessen respectively, all of whom have enriched my experience at the University of Victoria. My peers at the University of Victoria all deserve a big round of thanks, particularly Allison Benner, fellow phonetician who has lent her consummate proof reading skills more than once. I also wish to thank Janet Leonard, Thomas Magnuson, Izabelle Grenon, Carolyn Pytlyk, Huang Shu-min, Nicholas Welch, Ya Li, and Pauliina Saarinen, among the rest of the wonderful graduate students at the University of Victoria.

There are a number of individuals abroad that also played a role in the realization of this work. Foremost, I extend my gratitude to Dr. Lise Crevier-Buchman and Coralie Vincent of the Laboratoire de Phonétique et Phonologie at UMR 7018, CNRS/Sorbonne-Nouvelle in Paris. Both of these individuals played a pivotal role in the acquisition of the high-speed laryngoscopic videos, discussed in this thesis, by generously offering their expertise, time, and imaging equipment. Special thanks also goes out to Dr. Kiyoshi Honda of the École Nationale Supérieure des Télécommunications in Paris for his suggestions concerning the image analysis of the high-speed videos and Ken-Ichi Sakakibara of NTT Communication Science Laboratories in Kyoto, whose work has been helpful in elucidating the physiology of the laryngeal vocal tract.

On a final note, there are a handful of individuals who have made the rough road of calculus, physics, and mechanical engineering a little safer to travel for an ill-equipped art/linguistics student. Robert Prinz helped me understand Titze's challenging paper describing his mathematical model of the vocal folds. Anthony Bowers has been a great friend and helped me see just how cool the mathematics of engineering, physics, and biomechanics can be.

DEDICATION

This is for Jordan MacLennan. I miss you.

"Walking in falling leaves: procrastination, procrastination..."
14 Directions, Children

Chapter One

MODELING THE LARYNGEAL VOCAL TRACT

"Articulatory synthesis needs considerable understanding of the speech act itself"
R. Carlson (1995: 9932)

1.1 Introduction

This thesis documents the rational, background assumptions, methodology, and, ultimately, the creation of a three-dimensional (3D) computer model of the laryngeal vocal tract[1], with focus on the articulatory function of the laryngeal constrictor mechanism (LCM). Vocal tract modeling serves three primary purposes. First, vocal tract models allow for the visualization of articulators and their relationship to one another during the production of linguistic sounds. The laryngeal vocal tract is less accessible than other parts of the vocal tract because of its anatomical position within the body. Unlike the oral vocal tract, it is not easily palpated internally with the fingers or the tongue. It is also difficult to acquire clear and visually intuitive images of the larynx with current imaging technology, particularly during movement. Consequently, the larynx is arguably one of the most poorly visualized regions of the vocal tract. A 3D model can facilitate our understanding about the morphological character and physiological behaviour of its complex structure.

Beyond this visual and potentially pedagogical application, 3D models also serve as the basis for the next generation of speech synthesizers. Vocal tract area measurements obtained from 3D models can lead to more accurate and realistic synthetic speech made in real time as the model is deformed during a simulated articulation. Finally, the *analysis* of complex systems can be facilitated by using *synthesis* to probe, test, and challenge the depths of our understanding, often helping us to discover new insight into old problems. The quality of the synthesis is determined, in part, by the quality of the assumptions that make synthesis possible. In moving from the continuous, infinitely variable, and typically non-linear domain of the real, to the discrete, finite, and typically linear domain of the simulated, assumptions are used to form

[1] A term derived from Esling (2005), which is based upon the observation that the vocal tract can be conceptually divided into two sections: the oral and the laryngeal vocal tracts. The laryngeal vocal tract constitutes all of the linguistically relevant structure below the level of the uvula and velo-pharyngeal port.

approximations of what is being represented. It is through reflection on the translation from the real to the simulated domains that understanding can be refined and new questions posed.

1.2 Statement of purpose

The objective of creating a 3D model of the laryngeal vocal tract is to produce a user-oriented, dynamic and highly interactive simulation of the laryngeal architecture to explore its role in the production of linguistic and anthropophonic[2] sounds originating in the laryngeal vocal tract[3]. The primary modality of the model is visual representation[4]; pre-recorded auditory accompaniment is used to provide context for the simulated articulations. Anatomically, the model is focused on the representation of the aryepiglottic folds, with a biomechanical model of aryepiglottic fold movement to illustrate the physiology of aryepiglottic trilling. Linguistically, the model is designed to illustrate the articulatory contribution of the aryepiglottic folds to constricted laryngeal and pharyngeal sounds, both segmental and phonatory[5]. For contrast, the distinction between constricted sounds and non-constricted ones will play a major role in how the sounds are presented to the user. Interactivity is a key feature of the model and is provided by means of two parameter sets, one physiological and the other linguistic, that control model movements. The model is also intended to demonstrate how laryngoscopic videos can be used to contribute to the development of dynamic models of laryngeal function.

1.3 Models of the Vocal Tract

The model developed here is by no means the first of its kind; many linguistic scientists and researchers from diverse fields have endeavoured to create representations of the vocal tract to collect our knowledge of it into a coherent and visually intelligible whole. Typically these models have been static, but thanks to computers, they can now be made dynamic, and even interactive. The purpose of this section is to provide an overview of the field of vocal tract modeling, with the intention of elucidating the goals set for the present model and establishing the greater context in which this research takes place.

---

[2] Referring to the entire sound producing capacity of the human vocal organs.
[3] Such as whispered speech, which is not employed phonemically, but is used pervasively as a paralinguistic effect.
[4] Acoustic synthesis is planned for later phases of model development.
[5] Initiation (i.e. ejectives and implosives; see Catford 1977a) is not considered at this phase in model development.

1.3.1 *Two-dimensional vocal tract models*

The traditional approach to modeling the vocal tract is to provide a simple, two-dimensional (2D), midsagittal view of the tongue, lips, nasal cavity, and pharyngeal-laryngeal tube (e.g. Heinz & Stevens 1965, Coker & Fujimura 1966, Maeda 1972, Mermelstein 1973). These models typically take their data from x-rays or fleshpoint tracking (i.e. using x-ray microbeam or electromagnetometry[6]); however, some of the two-dimensional models were reverse-engineered using acoustic information (see Stone & Lundberg 1996). The typical application of these models is to allow for a derivation of the vocal tract area function for the purpose of speech synthesis (e.g. Mermelstein 1973). While many of the acoustic characteristics of the vocal tract can be derived from midsagittal geometry (Yehia & Tiede 1997: 1619), 2D models present limitations to synthetic speech derivation. Acoustically, 2D models are insufficient to model non-linear sound generation and transverse modes of propagation (both critical in deriving non-vowel sounds and frequencies above 4 kHz). Additionally, midsagittal distance does not sufficiently characterize the area of transverse vocal tract sections, as the vocal tract is not a uniform tube, but rather morphologically complex along its entire length.

1.3.2 *Three-dimensional vocal tract models*

Improving speech synthesis is one of the prime motivations behind the creation of 3D vocal tract models. With the advent of more powerful imaging technologies, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), data from the third dimension (i.e. depth) is accessible to modelers (see Whalen 2004). With 3D data, more complex forms of the vocal tract area function can be employed to produce more accurate synthetic speech (e.g. Fant 1992, 1993; Fant & Båvegård 1997; Kitamura et al. 2004). Furthermore, 3D models also make it possible to calculate a comprehensive model of sound wave propagation in the vocal tract (Birkholz & Jackel 2003), which would yield even greater accuracy.

There are a handful of three-dimensional vocal tract models that have been published in both the linguistic and medical literature (e.g. Stone 1990; Kahrilas et al. 1995; Stone & Lundberg 1996; Yehia & Tiede 1997; Engwall 1999, 2000; Badin et al. 2000; Birkholz & Jackel

[6] For a review of vocal tract imaging and data collection modalities used for linguistic research see Whalen (2004).

2003; Fels et al. 2006; Birkholz & Kröger 2007; Granat et al. 2007). These models primarily feature the oral vocal tract and face, with particular focus on the tongue. One of the common goals for many of these models is the construction of 3D *talking heads*. The research conducted at l'Institut de la Communication Parlée (ICP) in Grenoble (e.g. Badin et al. 2000) and Kungliga Tekniska högskolan (KTH) in Stockholm (e.g. Engwall 1999, 2000) exemplifies this approach. The applications for these talking heads (Beskow 2003: 43-48) include multimodal speech synthesis to facilitate articulation training for speech-impaired and profoundly deaf individuals and to improve the perception of synthetic speech by providing the listener with visual cues.

3D vocal tract models are used to investigate both embryological and evolutionary aspects of vocal tract development. A simulation of vocal tract growth in three dimensions was conducted by Birkholz & Kröger (2007). Their model can be used to generate hypotheses about how children and adults use different articulatory strategies to obtain the same acoustic output for a given vowel articulation, given the differences in vocal tract morphology between the two. On the basis of their model, Birkholz and Kröger conclude that acoustic characteristics do not scale proportionally with vocal tract growth (Ibid.: 380). This conclusion is reached based on differences in the synthetic acoustic vowel space of the infant vocal tract simulation compared with that of adults. On an evolutionary level, 3D modelling is being used to test hypotheses about the articulatory capacity of our distant ancestors (e.g. Neanderthals, *Homo ergaster*, *Australopithecus africanus, afarensis*: Granat et al. 2007: 382). From the 3D reconstructions of ancestral vocal tracts, acoustic simulations provide evidence that these vocal tracts were capable of producing acoustic output similar to modern humans.

In none of these models does the laryngeal vocal tract figure prominently or receive sufficient attention to meet the linguistic requirements of the model proposed herein. Kahrilas et al. (1995) present a medically oriented model that has an anatomically elaborated laryngeal vocal tract designed for the purpose of modeling deglutition. The model represents the hyoid bone, cervical spine, and the thyroid, cricoid, epiglottic, and tracheal cartilages. Additionally, the hypopharyngeal lumen is also included, based on a rendering of a liquid bolus (10 mL of liquid barium) as it passes through the vocal tract during swallowing. While the model is dynamic, it is not interactive. The model's movements are predetermined by the swallowing dataset. In models that do provide movement parameters to recreate vowel articulations (such as Badin et al. 2000; Engwall 2000; Birkholz & Jackel 2003), the laryngeal vocal tract is generally represented as a

tube with limited anatomical details. For example, in Birkholz & Jackel's (2003) model, the position of the hyoid bone, which determines the antero-posterior position of the tongue root, the volume of the pharynx, and the height of the larynx constitute the complete set of parameterized model components for controlling laryngeal vocal tract shape.

1.3.3 *Laryngeal vocal tract models*

Generally macroscopic models of the vocal tract, like those just discussed, leave out fine-grained detail concerning the structure and parameters of the laryngeal vocal tract. Within the past eight years, however, there has been a considerable interest in creating 3D representations of the vocal folds and associated structures. Much of the work has been spearheaded by Titze and his collaborators at the National Center for Voice and Speech, in Iowa City. Their modeling work is strongly grounded in anatomical empiricism; simulation of physical processes and tissue deformation are based on data obtained from research into the biomechanical properties of the vocal folds and associated structures. This research has been collected into a single volume (Titze 2006), which outlines the mathematical basis of the Myoelastic Aerodynamic theory of phonation. The three-dimensional component of the overall simulation is based on the earlier research of Hunter, Titze & Alipour (2004). The model they present is based on the finite-element method[7], which attempts to discretize a deformable body, such as the vocal folds, into a computationally tractable set of elements, which are given properties to reflect the physical system. Both adduction and abduction processes are modeled by applying abstract vector forces to represent the influence of the various intrinsic laryngeal muscles responsible for vocal fold posturing. Conveniently, finite-element models can be rendered as 3D meshes for viewing purposes.

Apart from Titze's research, there is a moderate amount of work being done to create 3D models of laryngeal structures. Montagnoli, Rubert, Guido, and Pereira (2006) present a three-dimensional deformable model of the vocal folds. The model uses two meshes for the vocal folds, whose vertices[8] are chained together through parallel spring-damper systems[9]. An

[7] Fundamentally, this mesh based method of structural analysis is designed to provide approximate solutions to partial differential equations.
[8] Each vertex is assigned a mass of 0.012g, a spring constant of 0.1Nmm, and a damping coefficient of 0.01 Nsmm.
[9] These are discussed in greater detail in Section 2.3.

unspecified driving force is used to generate glottal pulses at a frequency of 175 Hz. There are also two models which feature 3D reconstructions of the major laryngeal cartilages segmented from MRI images. Selbie, Gewalt, & Ludlow's (2002) model is an attempt to establish a reference by which anatomical coordinates of the laryngeal cartilages from different larynges could be registered within the same coordinate space, for the sake of morphological comparisons. The model of Han et al. (2002) is developed as a static model of the laryngeal cartilages for the purpose of displaying on the internet[10].

A final consideration of 3D vocal tract models is of Artisynth, a project spear-headed by the Department of Electrical and Computer Engineering at the University of British Columbia. Artisynth is, perhaps, the most elaborate vocal tract modeling project in existence today. The work is collaborative, involving individuals from the fields of medicine, linguistics, and engineering, which allows the scope of the model to include both biomechanical and linguistic aspects of vocal tract anatomy and physiology. Importantly, there is attention paid to the cartilaginous and (minimally) the extrinsic muscular framework of the larynx. This allows for a representation of global laryngeal movements and interactions with other vocal tract structures such as the tongue. However, the extent of intrinsic laryngeal details is more limited. The airway is modeled to represent the vocal tract lumen and is coupled with and deformed by the oral articulators and laryngeal positioning (Fels et al. 2006: 4-5). A two mass model of the vocal folds is used to provide a vocal source. Thus, this model can be used to generate a comprehensive simulation of vocal tract acoustics.

1.3.4 *Classification of the 3D Laryngeal Constrictor Model*

The model discussed in this thesis, named the 3D Laryngeal Constrictor Model (3D LCM), is uniquely situated within the field of vocal tract modeling and speech synthesis. The present model uses histological data to construct representations of laryngeal structures, while laryngeal state information is derived from observation of laryngoscopic video. Both aspects are uncommon in the field of articulatory synthesis (see Palo 2006). Thus, the present work contributes to the field at large by describing how laryngoscopic observation can be fruitfully used to characterize model articulatory posturing, and the use of histology to create 3D

[10] Their work is affiliated with the Visible Human Data developed by the National Institutes of Health (NIH).

representations of laryngeal structures. On the level of structural detail, the model is without equal concerning its focus on laryngeal function in relation to the laryngeal constrictor mechanism, and its biomechanical model of aryepiglottic trilling. Despite these distinguishing characteristics, it still does fit within the overall compass of speech synthesis, as Figure 1.1 diagrams.

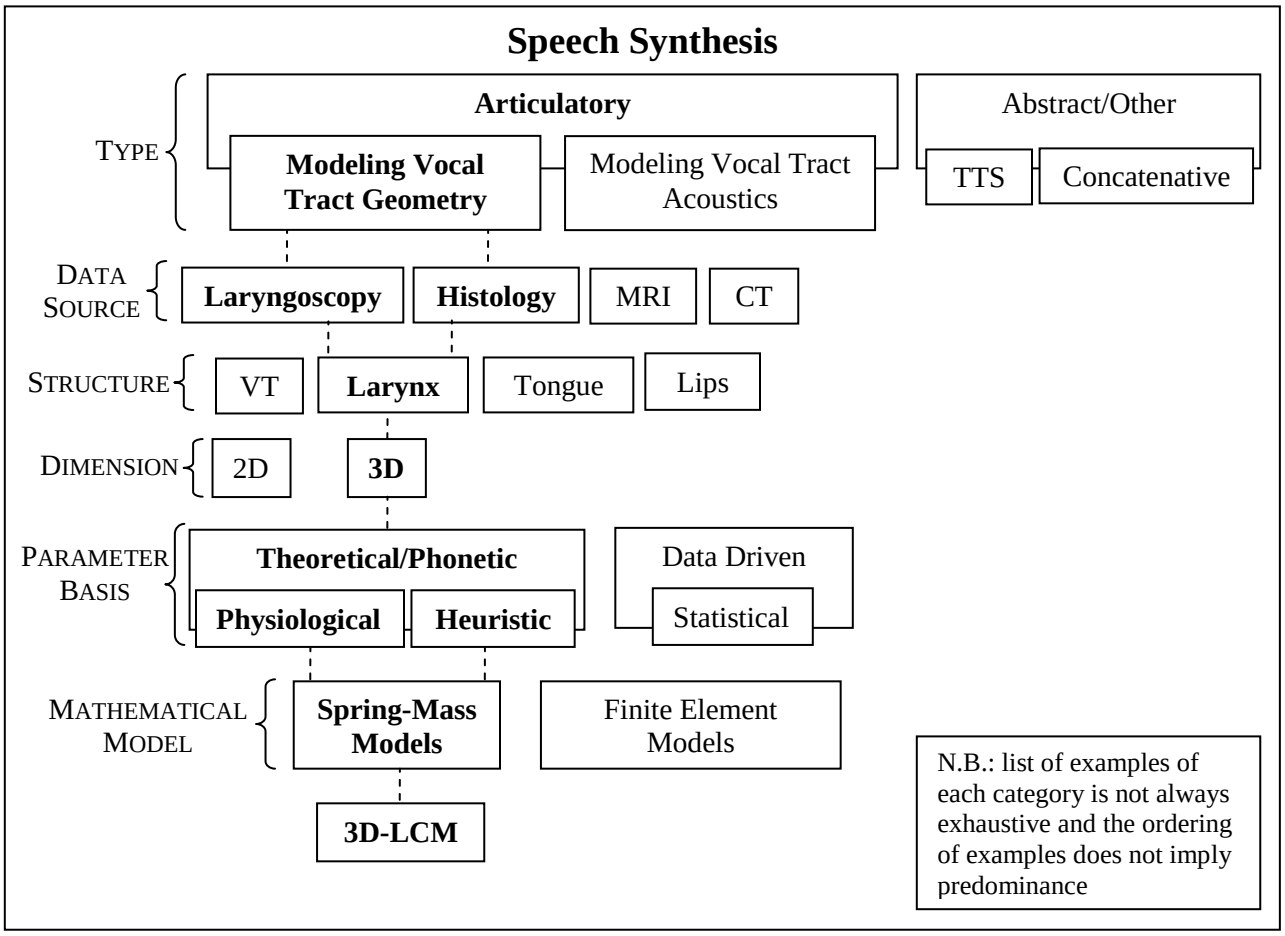


Figure 1.1 Classification of the 3D Laryngeal Constrictor Model (3D LCM). Only boxes that envelope other boxes, either entirely or partially, are intended to convey dependency. Categories organizing the various features of a model are provided in small caps. The diagram uses a dotted line and bold text to show the relationship of the 3D-LCM to general aspects of speech synthesis. TTS refers to text-to-speech synthesis, MRI to magnetic resonance imaging, CT to computed tomography, VT to vocal tract, and *concatenative* describes a type of speech synthesis that uses a database of pre-recorded utterances (commonly at the word level, but also segmental or syllabic levels) to carry out the synthesis.

The diagram is meant to be interpreted as a constellation of features that characterize speech synthesizers. It is frequently the case that researchers modeling the vocal tract will combine an articulatory model with an acoustic model to obtain a more accurate synthesis of vocal tract acoustics

(e.g. Birkholz & Kröger 2007). The 3D LCM does not model acoustics in the current phase of development described here, as the association lines in the diagram indicate. Rather, the model is primarily an articulatory synthesizer, and hence visual in nature. Control of the model is obtained through a theoretically derived parameter basis with a physiological focus. However, muscle function is abstract; it is based primarily on simple transformations of laryngeal structures, while articulatory movements are based on estimation of movement using laryngoscopic observation. Finally, all biomechanical elements of the model are controlled by means of a spring-mass formulation of tissue function, rather than a finite element approach (for example).

1.4 Thesis Outline

The following chapters of this thesis describe the creation of the 3D Laryngeal Constrictor Model. Chapter 2 builds the foundation for modeling the laryngeal constrictor mechanism by discussing the linguistic motivation and theory behind the model, followed by a detailed look at anatomical and physiological issues, and an examination of biomechanical models of the vocal folds, with particular attention paid to Titze's Sixteen-mass model (1973, 1974). Chapter 3 describes the method used to create representations of laryngeal structures and describes basic aspects of how the structures move relative to one another, much of which is based on the observations reported in Chapter 2. Chapter 4 discusses three image analysis techniques used to extract useable information from laryngoscopic videos to visually synthesize various phonation types and articulations. Chapter 5 deals with the biomechanical simulations that are used in the model, and focuses primarily on the simulation of aryepiglottic trilling. Chapter 6 represents the culmination of the preceding chapters by illustrating the model itself, focusing on the architecture and movement of the model. Finally, chapter 7 provides a summary of the thesis and outlines areas of further development for future phases of the model.

Chapter Two

LINGUISTIC, ANATOMICAL, AND BIOMECHANICAL CONSIDERATIONS

2.1 Theoretical and Linguistic Foundations for Modeling

In the following sections, I delineate the theoretical and linguistic considerations that the laryngeal constrictor model will be built upon. First, I present a review of the literature on pharyngeal and laryngeal articulations, focusing on the diversity of conceptualizations that have been put forth concerning laryngeal and pharyngeal articulations (section 2.1.1). This discussion is then followed by an introduction to Esling's (1996) innovative and comprehensive approach to laryngeal and pharyngeal articulations. Section 2.1.3 discusses voice quality/phonation type with respect to the role of the aryepiglottic folds as an additional vocal tract source. The next two sections address two of the more amorphous phonetic categories, and the extent of involvement of the laryngeal constrictor: laryngealization (section 2.1.4), and pharyngealization (section 2.1.5). The next topic (section 2.1.6) introduces the theoretical model proposed by Edmondson and Esling (2005) that is used as the linguistic parameter set for the 3D model of LCM. The final section (2.1.7) concludes with a summary of the issues discussed and the linguistic scope for this phase of the model.

2.1.1 Articulations of and in the pharyngeal tube: a variegated picture

The received view of what constitutes a pharyngeal articulation is stated unequivocally by Delattre (1971: 129):

> A pharyngeal articulation is one in which the root of the tongue assumes the shape of a bulge and is drawn back towards the vertical back wall of the pharynx to form a stricture. This radical bulge generally divides the vocal tract into 2 cavities, one below extending from the stricture to the glottis, the other above extending from the stricture to the lips.

Despite the simplicity of this statement, phoneticians have, until recently, been unable to provide a more precise formulation of what pharyngeal articulations are or how they work. The view that the tongue root is ultimately responsible for generating constrictions in the pharynx is pervasive, likely due to the ease with which it can be viewed in x-ray, MRI, ultrasound, and other imaging modalities. In Ladefoged & Maddieson's seminal cross-linguistic assessment of articulatory phonetics, *Sounds of the World's Languages* (1996), two possible places of articulation in the pharynx are described: upper pharyngeal and epiglottal. The former term, being the more commonly used, is actually often a mislabeling for fricative sounds that ought to be considered

epiglottal (1996: 37). This fact was recognized by the IPA in 1993 (IPA 1993; Pullum & Ladusaw 1996) in response to Laufer's advocacy to use the symbol [ʔ] (voiceless epiglottal plosive) to describe an allophonic variant of the voiced pharyngeal approximant of Arabic and Hebrew (Laufer & Condax 1979, 1981; Laufer 1991). In terms of manner, Ladefoged & Maddieson confirm that epiglottal stop is an articulatory possibility involving the epiglotto-arytenoidal occlusion of the laryngeal vestibule. Nevertheless, in their proposal for a phonetically grounded phonological feature set, pharyngeals and epiglottals are interpreted as primarily tongue root/radical articulations (1996: 372). Apart from primary articulations, the pharyngeal tube is also implicated in the production of various secondary articulations to modify vowel quality (1996: 299-313). One of the most common adjustments that can be made is the positioning of the tongue root, often referred to as Advanced/Retracted Tongue Root [±ATR] (symbolized with the diacritics, [̟] and [̠], respectively). Ladefoged & Maddieson argue that this adjustment is distinct from tongue height adjustments that characterize the tense/lax distinction (e.g. many Germanic languages like English). The antero-posterior positioning of the tongue root is viewed as directly affecting the volume of the pharyngeal lumen, which often is accompanied by a change in larynx height (e.g. Tiede 1996: 400). In the case of tongue root advancement, the pharyngeal lumen expands and the larynx tends to lower. Active retraction, on the other hand, is used in vowel and consonant modifications to produce pharyngealization, stridency, and rhoticization. Perhaps the most distinct of these three modifications is stridency of vowels (found in Khoisan languages such as !Xóõ: Traill 1985, 1986). Stridency is described as a sphincteric constriction of "the part of the tongue below the epiglottis and the tips of the arytenoid cartilages" (Ladefoged & Maddieson 1996: 311). In addition, these sounds are noted to engage the epiglottis or arytenoid cartilages in trilling movement (Traill 1985).
Pharyngealization involves narrowing of the pharyngeal cavity through tongue root retraction and raising of the larynx, which raises F1 and lowers F2 (Delattre 1971: 131). Rhoticization is noted to involve tongue root retraction without concomitant larynx raising, identifiable acoustically by the characteristic lowering of the third formant (Lindau 1978: 554-555). Ladefoged and Maddieson note that rhoticization is cross-linguistically rare, but found in Mandarin Chinese and English (two of the most widely spoken languages[11]).

[11] Of course, not all dialects of English have rhoticization, but that is an entirely different issue.

Despite being a contemporary *locus classicus* of phonetic knowledge concerning cross-linguistic articulatory possibilities, Ladefoged & Maddieson's treatment of pharyngeal articulations is insufficient to provide a unified account of the behaviour of the structures of the pharynx and larynx in articulatory terms, in light of research within the past fifteen years motivating a revised view of pharyngeal articulation. Esling (1996) provides a comprehensive survey of the literature over the past seventy years on pharyngeal articulations, which documents the diversity in approaches to pharyngeal-laryngeal articulations. From the diversity of perspectives reviewed, it is clear that there is a need for a unified conceptualization. Esling's (1996) review is provided in Table 2.1; I have appended information where I saw fit to do so.

Table 2.1 Esling's (1996) pharyngeal literature review

Primary Articulations					
Place	Manner(s)	Description	Example Language(s)	Symbol(s)	Reference
Pharyngeal	stop	tongue root retraction narrows pharyngeal region and provides seal	'dialects of Arabic'	N/A[12]	Hockett 1958
	stop	N/A	Nahk, Dagestanian, Georgian	N/A	Catford 1977b
	N/A	ventricular fold vibration & larynx raising	Somali	[ʕ] & [ħ]	Jones 1934
	vowel	lower pharyngeal approximation	N/A	[a]	Delattre 1971: 131
	N/A	lower pharyngeal stricture	Arabic	[ʕ] & [ħ]	Delattre 1971: 153
Epiglotto-pharyngeal	fricative; approximant; trill	extreme tongue root retraction - epiglottis approximates posterior pharyngeal wall	N/A	N/A	Catford 1968
Epiglottal	stop	epiglottis occludes laryngeal vestibule; voicing is impossible	some speakers of Arabic & Hebrew in slow speech	[ʡ]	Laufer 1991
	stop	epiglottis actively folds over to meet arytenoids	Chechen	N/A	Catford 1983
	fricative-trill vowel	vibration around the epiglottis	Khoisan	e.g. [a͡ä]	Ladefoged & Maddieson 1996: 311
Epiglotto-arytenoidal[13]	stop; fricative; approximant	articulation between base of epiglottis	Arabic & Hebrew	[ʕ] & [ħ]	Laufer & Condax

[12] N/A is used to designate either that the space is not applicable or information was not available.

		and top of arytendoids			1979, 1981
Faucal or Transverse Pharyngeal	approximant	approximation of faucal pillars, larynx raising	N/A	[ʕ] & [ħ]	Catford 1977a: 163
Uvular-pharyngeal[14]	fricative	extended channel from 'rear velar zone' (uvular to radico-pharyn. zone)	Bzyb	[χ] & [χ^{w}]	Catford 1977a: 195
Linguo-pharyngeal	approximant	tongue root (& epiglottis) retract narrowing pharynx	Danish (pharyngeal 'r')	N/A	Catford 1977a: 163
Mid-pharyngeal	fricative; approximant	below the faucal pillars and above the epiglottis (like a low back vowel)	Danish (pharyngeal 'r')	[ʕ] & [ħ]	Ladefoged & Maddieson 1996: 170
Ventricular	stop; fricative-trill	ventricular fold adduction (concomitant vocal fold closure) & general constriction in upper larynx & pharynx	Caucasian languages (Abkhaz, Adyghe, Kabardian)	[ʡ͡ʡ] & [ʜ͡ʢ][15]	Catford 1977a: 163
Secondary Articulations					
Place	Manner(s)	Description	Example Language(s)	Symbol(s)	Reference
Pharyngealization	N/A	upper pharyngeal constriction accompanies main articulation of 'backed sounds'	English, Spanish, German, French	N/A	Delattre 1971
	'emphatic consonants'	epiglottis and tongue root produce constriction in lower pharynx	Semitic	N/A	Laufer & Baer 1988
Glottalization	stops	ventricular fold adduction in addition to vocal fold adduction	American English	[t͡ʔ]	Fujimura & Sawashima 1971
	stops	occlusion of true vocal folds, false vocal folds, and aryepiglottic folds	N/A	N/A	Roach 1979: 2
Laryngealization	N/A	voicing striations in spectrogram; Esling (1996) speculates that this is trilling at a pharyngeal place	Iraqi Arabic	[ʢ]	Butcher & Ahmad 1987

[13] No explicit term could be found, so I have created a term that reflects their speculation about the nature of these articulations. Aryepiglotto-epiglottal is a likely candidate.

[14] This term is also applied to sounds found in Salish languages (Bird, S., personal communication). Esling, Fraser, and Harris (2005) describe the possibility of a simultaneous uvular and epiglottal closure; in this situation, the oral constriction is defined as primary, while the pharyngeal is secondary.

[15] [ʡ͡ʡ] = 'strong glottal stop/ pharyngealized glottal stop'; [ʜ͡ʢ] = 'breathy-voiced ventricular fricative trill'

		of articulation			
Sphincteric	vowels	vibration of the epiglottis & arytenoids	!Xóõ	[ã̪]	Traill 1985

The primary observation to be made concerning this review of pharyngeal & laryngeal articulations is the degree of disparity among the analyses. Esling himself notes (1996: 67) that there are no fewer than seven terms for characterizing stop articulations in the pharynx: "epiglottopharyngeal stop", "massive glottal stop", "strong glottal stop", "ventricular stop", "pharyngeal stop", "pharyngealized glottal stop" and "epiglottal stop". There is also considerable confusion about whether there are multiple articulators in the pharynx or merely a single one: upper/lower pharyngeal, faucal/linguo-pharyngeal/ventricular, and pharyngeal/epiglottal (Delattre 1971; Catford 1977a; Ladefoged & Maddieson 1996, respectively). The other discrepancy among the accounts concerns the active-passive status of pharyngeal-laryngeal articulators: while the epiglottis (Laufer & Condax 1979, 1981; Catford 1983), tongue root (Hockett 1958; Catford 1968), and the ventricular folds (Catford 1977a: 163) are all claimed to be active pharyngeal articulators, claims are made that the pharynx itself is an active articulator (Catford 1977a: 163). This claim usually attributes articulatory activity to the lateral or anterior-posterior compression of the pharynx (through the pharyngeal constrictor muscles). Furthermore, Catford (1977a: 163) characterizes standard Arabic ʿain [ʕ] as a faucal approximant, suggesting that the faucal pillars are active. Esling (1996) treats lateral compression of the pharynx as dependent on tongue root retraction and rejects faucal compression completely.

2.1.2 The Aryepiglottic Answer: A revised view of pharyngeal articulation

The answer Esling (1996, 1999a, 2005) provides to unify the descriptive disarray that pervades the study of pharyngeal phonetics is that the *aryepiglottic sphincter*[16] works in tandem with tongue-root retraction and larynx elevation to yield the range of articulatory possibilities in the pharyngeal region of the vocal tract. Importantly, Esling regards the aryepiglottic sphincter as the active articulator, while the epiglottis plays a passive role in forming a contact surface for the aryepiglottic folds (AEFs), which move anteriorally. In addition, tongue-root retraction and a raised larynx setting are the unmarked concomitants of aryepiglottic constriction. It is possible,

[16] See Appendix A for a brief description of the aryepiglottic sphincter and some illustrative diagrams.

however, to adjust laryngeal height independently of tongue-root and aryepiglottic fold activity. Moreover, Esling's interpretation also establishes an important link between the pharyngeal articulator and the functionally unrelated series of gestures involved in swallowing, throat clearing, and excretion (or exertion). The sequence of pharyngeal & laryngeal constriction gestures is largely the same for all of these gestures: vocal fold adduction, ventricular fold adduction, cuneiform cartilage & aryepiglottic fold approximation to the base or tubercle of the epiglotis, and epiglottis retraction, which is driven by movement of the tongue root (Painter 1986; Logemann 1993; Esling 1996). A summary table of Esling's reconfiguration of pharyngeal articulations is presented below:

Table 2.2 Esling's (1996, 1999a, 2005) unified approach to pharyngeals and laryngeals

Pharyngeal and Laryngeal Articulations					
Place	Manner(s)	Description	Example Language(s)	Symbol(s)	OldTerm(s)/Replaces?
Glottal	stop	adduction of vocal folds	N/A	[ʔ]	
	fricative	vocal folds are abducted	N/A	[h]	
Pharyngeal (aryepiglotto-epiglottal)	stop	AE folds press against epiglottis on the retracting tongue; inherently voiceless (Laufer 1991)	Caucasian	[ʡ]	'strong glottal stop'; 'pharyngealized glottal stop'
	fricative	AE fold constriction with narrow triangular formation between AE folds and epiglottis; VFs abducted	Arabic	[ħ]	
	approximant	AE fold constriction and approximation against epiglottis; VFs vibrating	Arabic (Laufer 1996)	[ʕ]	'voiced pharyngeal fricative'
	trill	considerable AE constriction & increased aerodynamic pressure results in AE fold trilling	strident vowels in !Xóõ; Caucasian	[ʜ] & [ʢ]	'ventricular fricative trill'

This view provides a clean distinction between glottal and pharyngeal (aryepiglottal-epiglottal articulations) on the basis of aryepiglottic fold activity. It must be emphasized that, as Esling (2005) asserts, while both tongue root retraction and larynx-raising are possible and indeed frequent concomitants of aryepiglottic stricture, they are not necessary for laryngeal constriction to occur. A notable instance of pharyngeal articulation with lowered larynx setting occurs in Palestinian Arabic pharyngeals (Edmondson & Esling 2005: 166). Both tongue root

retraction and larynx raising make aryepiglottic occlusion more efficient, and indeed modify the shape of the pharyngeal cavity and consequently the acoustic output.

Table 2.2 serves to clarify the cardinal categories of laryngeal and pharyngeal articulations. With the role of the aryepiglottic folds established, further discussion concerning phonation type (Section 2.1.3) and secondary articulations (Sections 2.1.4 & 2.1.5) is now entertained to explore and refine the taxonomic structure and scope of these terms in light of the revised view of laryngeal-pharyngeal articulation. These sections identify that, while the proposal to attribute primary articulatory activity in pharyngeals and epiglottals to the aryepiglottic constrictor, this view has not been, and still is not ubiquitously held by all researchers of phonetics, although the idea is starting to gain currency.

2.1.3 Voice quality / phonation type and aryepiglottic stricture

One of the central themes in acoustic phonetics is the relationship between the glottal source and the vocal tract filter. The contribution of the aryepiglottic sphincter potentially amounts to the addition of two more sources, as in aryepiglottic trilling during harsh voice. Ladefoged & Maddieson (1996: 372) limit phonation type to a matter of glottal characteristics, and vaguely attribute pharyngeal activity as a property of vowels. More troubling still is that glottal height, a function of larynx height, is disassociated from its inherent physiological relationship with laryngeal constriction, which engages the aryepiglottic folds and reshapes the epilaryngeal lumen. Laver's discussion of voice quality, reinterpreted into matrix format in Table 2.3, is more suggestive of the combinatory possibilities of mixing longitudinal tension, medial compression, glottal shape, and aryepiglottic stricture[17]. According to Laver (1980: 93), voice qualities can be superimposed to yield compound phonation types. Superimposition may be a matter of mixing glottal parameters, or, as in the case of harsh whispery voice (HWV), potentially adding aryepiglottic sources, particularly if the harshness occurs at lower pitches. For Laver, and contemporaries, harshness is interpreted not as an independent parameter, but rather a modification to other parameters involved in any particular phonation type or mode, particularly glottal period. Harshness is described as emergent with increasing aperiodicity of the fundamental frequency, or glottal *jitter*, which notably impacts glottal quality rather than glottal

[17] He does not explicitly discuss aryepiglottic stricture as a parameter, but the inclusion of harsh voice into the classification scheme allows for the interpretation that the aryepiglottic folds play a role in determining phonation type.

pitch (1980: 126-132). Interference with vocal fold function producing harsh voice is suspected to be a result of excessive tension of the vocal folds or a result of the ventricular folds impeding normal vocal fold movement.

Table 2.3 Laver's (1980: 111-112) Classification of Phonation Types

Class[18]		I		II		III	
	Type	Modal (V)	Falsetto (F)	Whisper (W)	Creak (C)	Harsh (H)	Breath (B)
I	Modal (V)	V	-	WV	CV	HV	BV
	Falsetto (F)		F	WF	CF	HF	-
II	Whisper (W)			W	WC	HWV / HWF	-
	Creak (C)				C	HCV / HCF	-
III	Harsh (H)			HWCV / HWCF		-	-
	Breath (B)						-

Catford (1964, 1977a) provides a similar interpretation on the phonation type taxonomy: he asserts that the ventricular folds give rise to what is effectively the harsh category in Laver's classification. The possibility for a second laryngeal source is implicit in his term 'double voice' (see Table 2.4: ventricular + glottal). Additionally, an *anterior* parameter is introduced that describes the potential for vibration of the ligamentous glottis exclusively, while the cartilaginous glottis remains forcibly static due to strong arytenoid adduction. The perceptual terms 'tight' and 'hard' are used to describe laryngeal settings with anterior quality (1977a: 103). Catford's phonation type classification is presented in Table 2.4.

Table 2.4 Catford's (1977a: 111-112) Classification of Phonation Types[19]

Stricture Type	Location		
	glottal	anterior	ventricular + glottal
a: wide open	voiceless		
b: narrowed	whisper	anterior whisper[20]	ventricular whisper
c: vibrating freely	voice	anterior voice	double voice
d: low frequency taps	creak	anterior creak	ventricular creak
a + c: open vibrating	breathy voice		
b + c: narrowed vib.	whispery voice	anterior whispery voice	
c + d: vibrating & taps	creaky voice	anterior creaky voice	

[18] Class groups the phonation types according to how pattern with regards to other types. Class I can combine with other types but not other types from class I. Class II can occur alone and as compound types, while Class III must be combined. One might ponder whether voiceless harshness is not a possibilty (as in voiceless aryepiglottic trilling), or simple breath.

[19] Catford does not include creaky voice in his classification – it has been added here. Other gaps are surprising as well, due to the absence of any physiological motivation for a phonation type not to be listed, for example: ventricular whispery voice, ventricular creaky voice.

[20] Anterior whisper and anterior whispery voice appear to be improbable in light of laryngoscopic evidence showing whispery voice to involve a patent posterior glottis (see section 4.2). Catford's understanding of whisper is that it is the narrowing of the glottis which yields turbulent flow, rather than turbulent glottal flow through the aryepiglottic folds and epiglottis.

These two classic taxonomies of phonation type discussed above do not explicitly describe the aryepiglottic folds as source generating. However, laryngeal constriction is implicit in the inclusion of ventricular (i.e. harsh) into the taxonomy. Catford's "double voice" may not be directly interpretable as referring to aryepiglottic trilling: Catford exemplifies this phonation type with the singing style of Tibetan monks and Jazz/Scat singer Cab Calloway[21] (1907-1994). As Esling (2002) describes, of the range of phonation types investigated, including Tibetan chanting, and the 'tense/lax' register tone contrast found in three Sino-Tibetan languages (Yi, Bai, and Tibetan), only Bai employs aryepiglottic trilling in its tense register when the tone level is at its lowest (tone 21). For the most part, what is observed is ventricular stricture where the "arytenoids, vocal folds, ventricular folds, and lower muscular components of the laryngeal sphincter mechanism are tightly adducted to form a channel of substantial height with a distinct ring of mucous at its upper border" (Esling 2002: 1082). Similar vocal-ventricular sources are used in the Mongolian singing styles of the Altai Mountains described and investigated in Sakakibara et al. (2004) and Lindestad et al. (2001). Two styles are attested: *drone voice*, which involves considerable ventricular stricture and vibration, with the possibility for whistle-like overtones, and *kargyraa voice*, which is less constricted than drone with ventricular vibration that apparently occurs at $f_0/2$, which may correspond also with Catford's double voice. In Lindestad et al. (2001: 83), high-speed laryngoscopic video clearly shows kargyraa engaging the ventricular folds in oscillation that is half the frequency of vocal fold vibration.

From consideration of Catford's and Laver's taxonomies, it would seem that aryepiglottic fold vibration was left out, despite being a anthropophonic possibility, and even attested as a phonetic realization of pharyngeals in Iraqi Arabic (Edmondson et al. 2007) and tense or constricted registers in languages such as Bai. One only needs to go as far as the voice of Louis Armstrong, famous American jazz trumpeter and singer (1901-1971), which is frequently stylized by aryepiglottic trilling and raised larynx voice.

2.1.4 Laryngealization

Laryngealization has proven difficult to incorporate into phonation type taxonomies (e.g. Gerratt & Kreiman 2001). The nomenclature used to describe laryngealization is notably diverse

[21] Upon listening to a vast range of Cab Calloway's music, I could not detect aryepiglottic trilling or ventricular voice/double voice (harsh voice) in his singing. Rather, the style tended towards a more open, modal style with infrequent use of harsh voice at high pitch and an emphasis on the rhythmic aspects of scat singing.

and at times ambiguous (see Esling, Fraser, & Harris 2005). The term is often used as a broad label, and is either equated to or encompasses other related labels. Some of the more prominent labels include: creak and creaky voice, glottal fry or vocal fry (particularly in the Americanist tradition), glottalization (see Laver 1980), pulse register phonation, pressed phonation (e.g. Stevens 1988; Titze 1995), or phonation in a "braced configuration" (Pierrehumbert & Talkin 1992; also see Henton & Bladon 1988 for a complete survey). Laryngealization is also sometimes interpreted as a secondary articulation where glottal constriction accompanies an oral articulation, which is a common allophonic variant of the production of English stops in word final position, such as *pack* [$p^{h}æ\widehat{ʔk}$] (Ladefoged & Maddieson 1996: 73-74). This phenomenon is also referred to as glottal reinforcement and occurs phonemically in Thai. Glottalized resonants are a common feature of many aboriginal languages of North America (including: Chumashan, Yokutsan, Salinan, Palaihnihan, Sahaptian, Salishan, and Wakashan families, and a handful of isolate languages, such as Klamath, Kutenai, Wappo, and Masset Haida; Mithun 1999: 19; Bird et al. 2008), Otomanguean, Chadic, Niger-Kordofanian, Khoisan, Mon-Khmer, Kam-Sui, and Caucasian (see Gordon & Ladefoged 2001; Carlson, Esling, & Harris 2004; Esling, Fraser, & Harris 2005).

Ladefoged and Maddieson (1996: 55) allude to a continuum of laryngealization, with the degree of glottal constriction as its prime physiological correlate. In any case, the idea of vocal fold stiffness is pervasively used to characterize laryngealized states (e.g. Blankenship 1997: 1). Stevens (1988: 361-364) describes laryngealization (under the terms pressed or glottalized) as a state where the vocal folds are "pushed strongly together" while at rest. Notably, the compliance of the vocal fold cover is increased due to strong activity of the thyroarytenoid muscle. Unlike model phonation, vibration in this state is not characterized by lateral displacement of the folds, but rather a change of shape of the surface layers or cover of the fold. Hirose (1995) reports that laryngealization, as it occurs in Korean fortis stops and Danish stød, involves increased medial compression (in the form of lateral cricoarytenoid activity) with concomitant thyroarytenoid activity. Ladefoged and Maddieson (1996: 55) classify both of these phenomena as stiff phonation, which only involves "a slight degree of laryngealization" as opposed to creaky phonation, although Esling (1996) suspects aryepiglottic activity for the stød phenomenon.

An almost universal observation about laryngealized phonation types is that they tend to be produced at an extremely low pitch (Hollien 1974), ranging from 24-52 Hz for male speakers

(Laver 1980: 122). Both a low subglottal pressure and volume-velocity (12-20 cm^3/s) contribute to the decreased fundamental frequency for laryngealized phonation (see Catford 1977a: 98). Depending on the areodynamic conditions, the action of the thyroarytenoids and the interarytenoids can both yield laryngealized quality. Typically, the vibratory characteristics of the vocal folds must be altered to produce either a horizontal phase difference between the anterior and posterior portions of the glottis (Ladefoged and Maddieson 1996: 53) or exclusive vibration of the anterior most portion of the ligamentous glottis (Catford 1964: 32). A number of researchers point out that laryngealization may also be accompanied ventricular fold adduction, which increases the functional mass of the vocal folds (e.g. Moore 1971: 72). Under these conditions, the vibratory unit becomes massive, although not very tense, and vocal fold vibration is damped, yet responsive to low subglottal pressures (Hollien, Moore, Wendahl & Michel 1966: 247). The degree to which the activity of the ventricular folds is physiologically essential to the production of laryngealized phonation remains controversial pending further research. Nevertheless, many researchers identify laryngealized phonation as the last phonatory state before complete glottal closure as in glottal stop[22] (Ladefoged 1971; Klatt & Klatt 1990: 823; Gordon & Ladefoged 2001: 384). The interest of this relationship is that the ventricular folds have been identified as damping the vocal folds during glottal stop production in a movement described as ventricular incursion (Esling & Harris 2005; Edmondson & Esling 2006: 169). There is evidence that glottalization should also be interpreted as involving ventricular fold damping: Nuu-chah-nuulth (Wakashan) and Nlaka'pamux (Salishan) have been demonstrated to use ventricular incursion and a moderate amount of aryepiglottic constriction (e.g. Esling, Carlson, & Harris 2002; Esling, Fraser, & Harris 2005).

Overall, laryngealization and glottalization are interpretable as more than simply involving the manipulation or cessation of glottal vibration. Some form of laryngeal constriction is necessary if the ventricular folds are used to form these gestures, with the potential for increased aryepiglottic fold tension or even moderate degrees of sphincteringg, as is the case for harsher varieties of laryngealization (Edmondson & Esling 2005: 162). Physiological and morphological aspects of the ventricular folds will be reviewed later in Section 2.2.

[22] Phonological laryngealization itself is typically expressed as a sound occurring somewhere along the end of this continuum: i.e. it ranges from being produced as creakiness to complete glottal stop.

2.1.5 Pharyngealization

Pharyngealization is perhaps most widely exemplified by the secondary articulation distinguishing emphatic consonants from plain ones in nearly all[23] dialects of Arabic (e.g. Cairo, Iraqi, Jordanian, Libyan, Moroccan, Qatari, Tunisian). The occurrence of pharyngealization in Arabic has been known since early Arab grammarians (such as Sibawayh; see Al-Nassir 1993) identified its use as a contrastive feature, which they described as "*ʔiṭbāq* 'spreading and raising of the tongue', *ʔistiʕlāʔ* 'elevation of the dorsum', and *taʃxīm* 'thickness, heaviness'" (Lehn 1963: 29). While the above descriptions do not necessarily entail pharyngeal activity, modern phoneticians agree that emphatics are achieved by retracting the root of the tongue, effectively constricting or reducing the volume of the pharyngeal cavity, in addition to potential tongue dorsum spreading and concavity, vocal tract tension, and lip rounding (Lehn 1963: 30-31; Ladefoged & Maddieson 1996: 365). Furthermore, Ladefoged and Maddieson suggest that pharyngealization can be broken down into upper and lower pharyngeal cavity variants based on their cross-linguistic comparison of pharyngealization. Numerous investigations into the exact articulatory properties of emphatics in the dialects of Arabic have been undertaken using x-ray (Bukshaisha 1985; Al-Nassir 1993), cine-/videofluorographic (Ali & Daniloff 1972; Ghazeli 1977; Laradi 1983), and endoscopic (Laufer & Baer 1988) imaging modalities. The common finding is that emphatics involve rearward movement of the tongue root approximating or contacting the posterior pharyngeal wall. However, Hassan & Esling (2007: 1755) assert, based on laryngoscopic evidence from Iraqi Arabic, that emphatics also involve aryepiglottic sphincteringand larynx raising, in addition to the ubiquitously observed tongue root retraction. Aryepiglottic sphincteringis also reported to occur in Moroccan (Zeroual 1999: 997) and Jordanian Arabic emphatic consonants (Heselwood & Al-Tamimi 2006; Heselwood 2007: 9). Turning to non-Arabic languages, many languages of the Caucauses (such as Tsakhur and Chechen) employ pharyngealization contrasts on consonants. Kingston & Nichols (1986) report that Chechen pharyngealized stops show F1 raising and F2 lowering in neighbouring vowel formant structure, indicative of tongue root retraction according to Delattre (1971: 131). Many of the languages of the Northwest American coast, from both Wakashan and Salishan families, also

[23] This is according to Lehn (1963: 29), who reports that Maltese Arabic is the only dialect without the emphatic feature.

have pharyngealization; Colarusso (1985: 366-367) describes pharyngealized consonants in these languages as also made with the tongue root, and possibly involving palatalization, yet is aware of aryepiglottic activity in pharyngeal consonants, which he describes as "adytals" (involving epiglottal covering of the laryngeal aditus). Using ultrasound, Namdaran (2006) reports that retracted consonants in St'át'imcets (Interior Salish) involve tongue root activity, although nothing is reported concerning the articulatory status of the aryepiglottic sphincter. Miller (2007) reports that 'epiglottalized' or pharyngealized vowels and consonants in Ju|'hoansi, a Khoisan language of Namibia and Botswana, contain aryepiglottic sphinctering. In the literature documenting cases of pharyngealization as a secondary feature of vowels, the role of the tongue root in the production of these articulations is typically highlighted (Delattre 1971; Gick 2002).

Overwhelmingly researchers have reported that the tongue root is primary in generating pharyngealization. This observation is certainly plausible as the tongue root can be used to manipulate the volume of the pharyngeal cavity, and hence its acoustic characteristics. However, tongue root retraction is a peripheral mechanic in the process of laryngeal constriction. It works synergistically with larynx raising (which is not as frequently noted in discussions of pharyngealization) to make the sphinctering of the larynx more efficient (Esling 2005: 26). This functional relationship is expressed physiologically by the coupling of the larynx and the tongue root across the hyoid bone, by the hyoglossus and the thyrohyoid. Thus, it is suspected that many of the preceding descriptions of pharyngealization have omitted a description of the action of the aryepiglottic folds. This may be on account of the difficulty of visualizing aryepiglottic compression in laryngoscopy as the tongue root (and epiglottis) retracts, and the poor imaging of the folds in midsagittal x-rays.

2.1.6 The valve conception of the laryngeal constrictor mechanism: a unified picture

Esling (1996), discussed in Section 2.1.2, is the theoretical precursor to the conceptual framework for laryngeal and pharyngeal articulations presented in Edmondson & Esling (2005). Simply put, their conceptual model of laryngeal and pharyngeal articulation is a system of interrelated articulatory valves. There are six valves[24]; they are: the vocal folds (V1), the ventricular folds (V2), the aryepiglottic sphincter (V3), tongue-root retraction (with dependent movement of the epiglottis: V4), larynx elevation (V5), and pharyngeal narrowing (V6). Each

[24] Each one is referred to by <V>*n*, where *n* denotes the valve number.

valve has a strict anatomical definition and is characterized physiologically by a set of movement vectors and mechanical properties. Most valves in the model have several different parameters which change the functionality of the valve. These anatomical and physiological characteristics are summarized in the following table:

Table 2.5 The Valves of the Throat (Edmondson & Esling 2005)[25]

Valve	Structure	Parameters	Effect	Muscles (Forces)
V1*	Vocal Folds	Medial Compression: *Lateral-Medial*	abduction, adduction, glottal aperture shape	LCAs, IAs, TAs, PCAs
		Longitudinal Tension: *Antero-Posterior*	glottal pitch, vertical contact & mass	CTs, TAs
		Vibration: *Lateral-Medial*	phonation/voice	(Aerodynamic)
V2	Ventricular Folds	Medial Contact: *Lateral-Medial*	abduction, adduction	Ext.TA, TEs, AE
		Ventricular Incursion: *Superior-Inferior*	vocal fold suppression	Ventricularis m.
		Vibration: *Lateral-Medial*	diplophonia, throat singing voice	(Aerodynamic)
V3*	Aryepiglottic Folds	Laryngeal Constriction: *Antero-Posterior*	epilaryngeal aperture modification	TAs, AEs, TEs
		Dual Vibration: *Antero-Posterior*	trilling/growl voice	(Aerodynamic)
V4	Tongue Root & Epiglottis	Advancement/Retraction: *Antero-Posterior*	expansion/reduction of pharyngeal volume	hyoglossus
		Epiglottal Vibration: *Antero-Posterior*	epiglottal trill	(Aerodynamic)
V5*	Larynx	Height: *Superior-Inferior*	raised larynx voice, lowered larynx voice	supra- & infrahyoids[26]
V6	Pharynx	Sphinctering: *Radial*	reduction of pharyngeal volume	superior, medial, inferior PCs

N.B.: AE = Aryepiglottic m.; CT = cricothyroid m.; IA = interarytenoid m.; LCA = lateral cricoarytenoid m.; PC = pharyngeal constrictor m.; PCA = posterior cricoarytenoid m.; TA = thyroarytenoid m. (external & internal); TEs = Thyroepiglottic m.; * = dominant/independant valve

The valves should not be viewed as entirely independent of one another: there are numerous physiological dependencies which characterize the system. There are three valves that exhibit the greatest degree of independence: V1, V3, and V5. Each of these valves is freely configurable regardless of other valve settings. For example, V1, the vocal folds, can be abducted or adducted when V3 is set to a constricted posture. While the unmarked setting for V3 constriction involves V5 raising (Edmondson & Esling 2005: 166), it is possible to constrict V3

[25] This chart has been adapted from the discussion found in the original article.

[26] The muscles from the suprahyoid group (digastic, stylohyoid, mylohyoid, geniohyoid, hyoglossus, & genioglossus) elevate the larynx; the infrahyoids (sternohyoid, omohyoid, & sternothyroid) serve to lower the larynx.

while the V5 is in a lowered position. V3 may also be actively constricting whether or not V4 is in its facilitating, retracted position. Thus, in this thesis, valves 1, 3, & 5 are thus interpreted as constituting a dominant sub-grouping, given their independence (these valves are marked with a * in Table 2.5).

On the other hand, V2, V4, and V6, are in some way physiologically dependent upon or bound to the actions of the dominant valves. V2, the ventricular folds, are most dependent upon the action of V1, and they lack musculature distinct from the musculature controlling vocal folds (V1). In evidence of this dependency, ventricular vibration or adduction without vocal fold adduction is unattested. V2 is also likely dependent upon the action of V3, which masses the ventricular folds together as the arytenoids, cuneiforms, and aryepiglottic folds approach the epiglottis. It is likely that the folds are forced to bulge in both an inferior and superior direction[27] as this happens, as they are displaced by the actions of V3. Turning now to V4; the key independent linguistic contribution of V4 is the production of the linguo-epiglotto-pharyngeal stop, in which the tongue root forces the epiglottis to seal off the pharynx by contacting back to the posterior pharyngeal wall. This manoeuvre *cannot* be accomplished, however, without initially constricting the lower valves, and most importantly V3, the aryepiglottic sphincter (Edmondson & Esling 2005: 164). Finally, V6 is the most 'degenerate' of all the valves. Its engagement entails the constriction of V3, and also requires that V5 is set to a raised position. Edmondson & Esling (2005: 187) present evidence of distinctive V6 activity in Somali during pharyngeal consonant production. Narrowing of the pharynx is observed during the vowel part of the word [ʡ͡qi:ʡ͡q] 'to emit smoke', which is evidently a coarticulatory effect: the vowel [i:] precedes a linguo-epiglotto-pharyngeal stop (with a simultaneous uvular stop). Despite its status as dependent, pharyngeal narrowing imparts an important acoustic effect on the resonance characteristics of the vocal tract. Stevens (1999) claims that pharyngeal narrowing "accelerates the buildup of pressure during the constricted interval" and influences the "bandwidth of spectral prominences" (1999: 251, 271) yielding a perceptually "squeezed" and "fronted (advanced)" quality (Catford 1977a: 182).

Table 2.6 summarizes valve activity across the cardinal articulatory and phonatory categories of a phonetic/phonological and anthropophonic nature. The list is intended to

[27] The superior buldging is suspected to be key to producing the rounded shape of the laryngeal vestibule used in some throat singing styles.

incorporate much of the laryngeal and pharyngeal phenomena discussed up to this point. It should be remembered that each valve operates in an organic, continuous fashion, as opposed to a binary, categorical one. These valves exhibit gradient engagement, which is analogous to the gradient settings for valves found in household faucets.

Table 2.6 The Linguistic and Anthropophonic Deployment of the Valves[28]

Linguistic Feature		Valves Involved						Parameter Notes
		V1	V2	V3	V4	V5	V6	
Phonation Type	modal voice	●						
	falsetto voice	●						V1 phonation at high pitch V1 long. tension = limited vertical contact
	breathy voice	●				○		V1 open posteriorally with anterior vibration V5 may be lowered
	creaky voice	●	○	○				V1 phonation a low pitch V3 mild to extreme
	laryngealization	●	○	○				V2, V3 possible/optional
	Tibetan Chanting: high chant	●	○	●	●	●		V5 raised
	Tibetan Chanting: deep chant	●	●	○		●		V2 activity is extreme (narrow channel formed) V5 lowered
	Drone voice[29]	●	●					V2 activity is extreme (narrow channel formed) V3-V6 information is unavailable
	Kargyraa[30]	●	●	○		●		Similar to Tibetan 'deep chanting'
	harsh voice (low, mid, high)	●	●	●	○	○	○	V1 at various pitch levels; V1 opposes V3 at high pitch V3 trilling possible at lower V1 pitch V5 raising possible
	pharyngealized voice	●	○	●	○	●	○	V1 phonation at low pitch; V3 constricted V5 raised
	raised larynx voice	●	○	○	○	●	○	V1 phonation at high pitch V5 raised
	lowered larynx voice	●				●		V1 phonation at low pitch V5 lowered
	faucalized voice	●				●		V1 phonation at high pitch (due to V5 setting) V5 lowered; V6 inactive (pharynx expansion)
Primary Articulation	glottal stop continuum	●	●					various degrees of V2 engagement
	[ħ ʕ][31]	●	○	●	○	○	○	V1 ab- or adducted for voiceless/voice V3 constricted for [ħ ʕ]
	[ʜ ʢ]	●	○	●	○	○	○	V1 ab- or adducted for voiceless/voice V3 constricted & trilling
	aryepiglotto-epiglottal stop [ʡ]	●	●	●	○	○	○	V1 phonation is not possible V4, V5 possible/optional
	linguo-epiglotto-pharyngeal stop continuum: [ʡ][32]	●	●	●	●	○	○	V1 phonation is not possible V4 between moderate or extreme

[28] See Edmondson & Esling (2005) for a more elaborated discussion.

[29] Drone voice is described as involving strong ventricular fold adduction and may be produced with "whistle-like overtones" (Sakakibara et al. 2004a).

[30] Dzo-ke is the Tibetan equivalent of Mongolian Kargyraa (Lindstad et al. 2001: 78), and termed "deep" chant in Esling (2002). To avoid possible confusion, the Tibetan chants have been kept distinct from the Mongolian ones in Table 2.6.

[31] The arypiglotto-epiglottal consonants may have optional V5 raising or lowering as is evinced by Palestinian and Tigrinyan Pharyngeals respectively (Edmondson & Esling 2005: 166).

[32] Currently, there is no symbol or diacritic in the IPA to distinguish between aryepiglotto-epiglottal stop and linguo-epiglotto-pharyngeal stop. The necessity of a different symbol for this sound, however, is dubious as both aryepiglotto-epiglottal stops are auditorily indistiguishable (Esling, personal communication). This is a situation that is parallel to the absence of a different symbol for American English /ɹ/, whether molar or tongue-tip in its nature.

Secondary Articulation	glottalization	●	●					various degrees of V2 engagement
	pharyngealization	●	○	●	●	○	○	

N.B.: ● = necessary valve activity; ○ = potential/optional valve activity

2.1.7 Summary of the linguistic context for the 3D Laryngeal Constrictor Model

Sections 2.1.1 to 2.1.6 have introduced the issues, contentions, and scope of laryngeal and pharyngeal articulatory phonetics. The most contemporary perspective offered is that these sounds are a function of what is called, in physiological terms, the *laryngeal constrictor mechanism* (LCM), and the goal of the 3D Laryngeal Constrictor Model is to explore and demonstrate its function through visual analysis by synthesis. This final section is used to summarize the linguistic context.

The laryngeal constrictor mechanism is the principle physiological force behind deglutition (swallowing) and is accomplished by drawing the aryepiglottic folds anteriorally towards the epiglottis and tongue root. This potentially results in a hermetic sphincteringof the epilaryngeal tube. By default, the vocal folds and ventricular folds adduct during laryngeal constriction to protect the larynx, but are not necessarily engaged when the aryepiglottic folds are. Linguistically, laryngeal constriction plays a central role in pharyngeal articulations, both primary and secondary (Esling 1996, 1999a, 1999b, 2005; Edmondson & Esling 2006; Edmondson, Padayodi, Hassan, & Esling 2007; Esling, Zeroual, & Crevier-Buchman 2007; Hassan & Esling 2007). Depending on the degree of constriction, action of the vocal folds below, and activity of the oral cavity numerous sound production possibilities arise. Among them are the entire set of what are identified as pharyngeals and epiglottals in the IPA, [ʕ, ħ, ʢ, ʜ, ʡ], which includes approximant, fricative, trill, and also stop manners of articulation. All of these gestures can occur with a superimposed laryngeal height setting (Esling 1999a: 369). The standard case of laryngeal constriction involves raised larynx position, which yields canonical aryepiglotto-epiglottal articulations. This set of cardinal categories of pharyngeal articulations is paramount to demonstrating laryngeal constrictor function and thus it is demonstrated in the model (see Section 6.3). Additionally, the laryngeal constrictor mechanism contributes both a source and resonator (Edmondson & Esling 2006; Edmondson et al. 2007) to voice production,

and it can be recruited into various phonological vocal register contrasts found across languages. For example, Somali, Akan, and Kabiye, traditionally classed as [± ATR] harmony languages (Stewart 1967; Saeed 1999: 11-16), are reinterpreted as exhibiting a [± constricted] harmony (Edmondson et al. 2007) based on laryngoscopic evidence. In the [+constricted] set, a single unifying action of the laryngeal constrictor mechanism is attested, altering the vocal tract resonance and yielding the harmony effect. In languages such as Jianchuan Bai and Yi the laryngeal constrictor combines with glottal adjustments to produce register contrasts at varying tone levels and contours (Edmondson et al. 2001). The constricted vocal register in this language is characterized by trilling at the aryepiglottic folds yielding varying degrees of harsh voice, which can be considered a phonatory setting of two sources (vocal and aryepiglottic). The laryngeal constrictor mechanism also appears to be active in infant vocalization. As Benner, Grenon, & Esling (2007) and Grenon, Benner, & Esling (2007) demonstrate, the preponderance of vocal output by pre-linguistic infants is constricted (harsh, creaky, and glottalized). This pattern changes as the infants proceed into the babbling phase of linguistic production. The hypothesis is that the use of constriction in babbling is influenced by the degree to which constriction is an active phonological feature of the ambient language. In evidence of this, languages such as Moroccan Arabic, where constriction is a prominent phonological feature, babbling tends to be more persistently constricted than it is for languages where constriction does not play a strong phonological role, such as English or French (Benner et al. 2007: 2075). Given the importance of phonation type, the 3D model will also examine the laryngeal constrictor mechanism as it functions to produce voice quality and phonation modification.

Together, the entirety of the linguistic phenomena generated in the larynx is expressed in a unified manner by the laryngeal articulator model presented in Esling (2005) and further developed in Edmondson & Esling (2005). The fundamental contribution of this model is to divide the vocal tract into two articulatory domains: the oral and the laryngeal vocal tracts. Their action is coupled at the tongue root, which impacts not only the treatment of consonantal articulations, but vocalic ones as well. Vowels are reconceived of as existing in a tripartite space (fronted, raised, retracted) rather than a quadripartite space (back-front, high-low), based on the physiological possibilities of tongue movement in conjunction with those of laryngeal structures (Esling 2005: 19-23). The laryngeal articulator model also underscores the primacy of the aryepiglottic folds as an active laryngeal vocal tract articulator. The aryepiglottic folds figure

into the laryngeal articulator model as one of a series of interacting valves (Edmondson & Esling 2005), each with its own parameters for controlling the vocal aerodynamics and resonances. This includes the addition of a secondary vocal source from the vibration of the aryepiglottic folds (as is the case in harsh voice). The valve conceptualization will serve as the linguistic control context for the 3D Laryngeal Constrictor Model, which will allow for parameters of muscle function to be interpretable within a well developed linguistic framework, which indicates the phonetically and linguistically relevant features operating in the production of laryngeals and pharyngeals.

2.2 Anatomical Issues of the Laryngeal Vocal Tract

Despite the fact that anatomists have been researching the structure of the larynx and particularly the vocal folds for centuries, many aspects of laryngeal anatomy and physiology remain controversial (Pretterkleiber 2003: 251). This controversy deeply impacts the 3D Laryngeal Constrictor Model from a structural and functional perspective. The goal of the model is to provide a generalized anatomy of the larynx and pharynx, abstracting away from individual variation and anatomical aberrations that occur. The larynx is intended to be presented in a manner that is prevailingly congruent with the most widely accepted views of laryngeal anatomy, thus it is necessary to outline what those views are. Much of the controversy concerns muscular attachment, function, and even existence. Some structures, particularly the ventricular folds and the aryepiglottic folds and related structures have not received a great deal of focus from the literature. Thus, one objective is simply to find and report on whatever material exists that enhances our understanding of the form and function of these structures that are so important to the model and speech in general.

Another concern is deciding what to include, and what to exclude from the model. Computing power and memory are finite resources; judicious choice is needed to select a set of features to include in the model. The reality of the larynx is that it is infinitely more morphologically and physiologically complex than what can currently be modeled on computer. Thus, for the present model, many important features, such as nervous, adipose, and glandular tissue will not be featured[33]. The model meshes used to represent laryngeal structures will also be

[33] Later versions of the model may include these features.

considerably simplified representations of their real world counterparts. An excellent example of this is the epiglottis: the structure is notoriously pitted with crevasses to house the myriad glandular structures that are embedded within it; this is visually evident in histology such as that of Hirano and Sato (1993). As will be discussed in the model mesh methodology section, much of this detail is excluded. Exclusion of details is a matter of computational economy: the fewer polygons presented to the screen, the faster the model can be rendered.

On the macroscopic level, it is impractical to represent the entire head from the neck on up, as is common in many 3D models based on CT and MRI data sources (see Section 1.3). Because the prevailing focus is on the structure of the aryepiglottic folds and the laryngeal sphincter, it is necessary to include, minimally, the root of the tongue (V4). However, due to the complexity of creating the first three valves (i.e. the vocal, ventricular, and aryepiglottic folds), this development is postponed for a later phase of the model. Conceptually, the pharynx is also considered an articulatory component in the valve model (see Section 2.1.6). Thus, it must also be included. However, at some point both of these structures need to be truncated. The exact extent of the global anatomical structure that will be chosen for representation will be discussed in this section. Again, however, these delineations of model scope reflect the overall modeling project and *not* this specific phase of model development[34]. Another important matter concerns the sexual dimorphism widely known to differentiate the male and female larynx. The present version of the model uses two female larynges as the basis for developing the model meshes, and measurement data obtained from female larynges. This is a direct consequence of the choice to develop the model based on Hirano and Sato's (1993) histological work, which uses female specimens to illustrate adult laryngeal structure. Future versions of the model may include the option to view a male larynx, but for the present phase of the model, the female larynx is used to illustrate laryngeal-pharyngeal anatomy and articulatory processes.

In this section I will focus on the anatomical detail of the laryngeal constrictor mechanism. The objective is to provide an overview of the types of structures that are planned for representation in the complete 3D model. The first topic is a survey the component tissues of the laryngeal vocal tract (section 2.2.1). The following three sections (2.2.2 – 2.2.4) discuss specific anatomical and physiological issues pertaining to the vocal, ventricular, and

[34] The results of which are presented in Chapter 6. Exclusion of the upper valves (V4-6) in the present phase, which this thesis documents, is due to lack of appropriate materials to create the tongue root and epiglottis (as discussed in Section 3.2)

aryepiglottic folds (in that sequence). These are the valves (V1-V3) that will be represented in the present phase of the model, and thus an in depth treatment is warrented. Finally, Section 2.2.5 provides an outline of the specific details of the present phase of the 3D Laryngeal Constrictor Model.

2.2.1 Overview of the morphology of the laryngeal vocal tract[35]

This section provides an outline of the basic anatomy and physiology of the laryngeal vocal tract. It lays down the ultimate scope for the 3D model in its complete form. The reader should note carefully, however, that the present model does not include all of these details. Due to time constraints, only the first three valve structures are rendered. The global details of the laryngeal vocal tract are discussed, however, for completeness of the anatomical description.

The distinctive shape of the larynx and the laryngeal collar is due, in part, to the cartilages that comprise it[36]. There are three distinct types of cartilage, each of which exhibits different mechanical properties. The two that presently concern us are hyaline and fibroelastic as the laryngeal cartilages are solely of these two types. The thryroid, cricoid, and arytenoid cartilages are hyaline in nature and consequently are relatively rigid structures that exhibit a tendency towards calcification or ossification in old age (Zemlin 1997: 12, 104). The elastic cartilages of the larynx are the epiglottis, corniculate, and cuneiform; the latter two are paired structures. These are noted for their flexibility.

The hyoid bone, the sole bone in the laryngeal tube, fulfills two roles: it is a supportive structure for the tongue and a rigid frame from which the larynx is suspended and the muscles responsible for laryngeal elevation are attached (Palmer 1993: 110). The distance between the caudal edge of the thyroid notch to the lower edge of the hyoid is approximately 18.7mm in males (n = 28, SD = 5.3mm) and 13.1mm in females (n = 25, SD = 2.32mm).

The larynx is comprised of two layers of muscles: intrinsic and extrinsic. The extrinsic muscles are subdivided into suprahyoid and infrahyoid groups; contraction of either set results in a vertical translation of the hyoid bone and consequently of the larynx (Palmer 1993: 123). Thus, the terms laryngeal *elevators* and *depressors* are sometimes used (Zemlin 1997: 121) to describe the suprahyoid and infrahyoid groups respectively. Raising of the larynx is accomplished using

[35] All measurements presented in this section were obtained from Eckel et al. (1994).
[36] For a description of the aryepiglottic folds and diagrams of the laryngeal collar and its associate cartilages and tissues see Appendix A.

three means: if the hyoid bone is immobilized, contraction of the thyrohyoid muscle group will draw the larynx towards the hyoid; another possibility is to elevate the hyoid and larynx simultaneously using the suprahyoid muscles (digastric, stylohyoid, mylohyoid, geniohyoid, hyoglossus, genioglossus). A final possibility is the use of the stylopharyngeus muscle, which primarily serves to elevate the pharynx. On account of a structural link between the muscles of the pharynx and the laryngeal structures, contraction of this muscle also raises the larynx (Zemlin 1997: 277). There are still, however, subcategories of larynx-raising available depending on which suprahyoid muscles are constricted. Anterior raising movement is possible with the geniohyoid, genioglossus, and the anterior belly of the digastric (Zemlin 1997: 123). Contraction of the posterior belly of the digastric and the stylohyoid will induce a backward raising motion. Lowering, on the other hand, results from contraction of the sternohyoid, omohyoid, or sternothyroid muscles. As was discussed in Section 2.1.6, V5 of the laryngeal constrictor mechanism involves adjusting laryngeal height. The extrinsic muscles described here define the movement parameter of this valve as well as give shape to the larynx and, therefore, are an essential component requiring representation in the *complete* version of the 3D model.

The intrinsic muscles of the larynx serve to primarily characterize V1 in the laryngeal constrictor model. The vocal folds are a complex bundle of muscles collectively referred to as the thyroarytenoids. Contraction of the thyroarytenoids acts antagonistic to the cricothyroids resulting in thickened and slackened vocal folds (critical to produce creaky voice). Importantly it also results in an anterior translational movement of the arytenoids themselves (Palmer 1993: 134), which partially explains the narrowing of V3 during creaky voice. Longitudinal tension is regulated by the criocothyroid, while adduction involves the lateral cricoarytenoids and interarytenoids. These two muscles can contract independent of one another to yield varying medial tension (illustrated in Berg et al. 1960). Finally, abduction proper[37] is solely a function of the posterior cricoarytenoids, although contraction of the interarytenoids without concomitant lateral cricoarytenoid activity can bring about moderate abduction of the vocal processes of the arytenoids (Zemlin 1997: 148). The muscles of the ventricular folds and aryepiglottic folds are considered in further detail in section 2.2.3. All together, the vocal, ventricular, and aryepiglottic folds are featured in the present version of the 3D model.

[37] This involves separating both the vocal processes and bodies of the arytenoid cartilages. The interarytenoids are also capable of causing the vocal processes to be abducted when unopposed (e.g. Zemlin 1997: 148; Titze 2006: 133).

The central question concerning the structure of the tongue in the final version of the 3D Laryngeal Constrictor Model is: how does the tongue achieve a retracted state? Knowledge of these muscles will be central to configuring the 3D model to accurately portray how V4 operates. Of the extrinsic tongue muscles, only the genioglossus (the anterior fibers) and hyoglossus yield a backward and downward movement of the tongue; the tongue root can, however, be depressed and retracted by immobilizing the larynx (for example, with the infrahyoid muscles) and contracting the thyrohyoid muscle (Zemlin 1997: 122, 256). Kokawa et al. (2005) report that the superior pharyngeal constrictor muscle might also be responsible for 'retrusive' movements of the tongue, with accompanying constriction in the mid-pharyngeal cavity.

The final muscles relevant to the complete version[38] of 3D Laryngeal Constrictor Model are those of the pharyngeal constrictor group. The number of muscles involved in this function is considerable: the constrictor group is divided into three units, superior, medial, and inferior, each of which has a number of constituent muscles (Zemlin 1997: 275-4). In addition, the pharynx is also controlled by the stylopharyneus and the salpingopharyngeus which produce upward movement of the pharynx and medial movement of the pharyngeal walls. Functionally distinct muscles also occupy the pharynx; these are the muscles of the faucal pillars (palatoglossus and palatopharyngeal) and the sphincteric muscles of the esophagus and velopharyngeal port (the cricopharyngeal and velopharyngeal sphincters). The *complete model* will focus on representing the pharyngeal constrictor muscles as they constitute the main components of V6. However, only the three main muscles of this set will be represented, i.e. the superior, medial, and inferior constrictors. The faucal pillars will also need to be represented as they are stretched by the action of V5 lowering.

There are two other important tissue types that merit representation in the *complete model*. The first is the set of non-cartilaginous connective tissues (comprised of membranes and ligaments) that interconnect the various structures of the larynx and constrain the movement of cartilages that articulate together. The following membranes and ligaments will require representation: the hyothyroid membrane & ligaments, the hyoepiglottic ligament, the conus elasticus, the lateral cricothyroid membranes (terminating in the vocal ligaments), quadrangular membrane, and the pharyngeal aponeuroses. Less important to the 3D Laryngeal Constrictor

[38] The reader is reminded again that model demonstrated at the end of this thesis only represents the first three valves. Details not included in the present phase are planned to be included at a later date.

Model is the cricotracheal membrane which serves to link the cricoid cartilage to the superior edge of the first tracheal ring. The second tissue type is the mucous membrane which lines virtually the entire vocal tract (including the tongue). There are local variations in the cellular structure of the mucous (squamous and ciliated columnar epithelium),. Regional differences are most pronounced for the vibrating structures of the larynx (the vocal, ventricular, and aryepiglottic folds). The contact edge of these folds consists of stratified squamous epithelium, while the remainder of the mucosal lining is columnar epithelium. Representation of regional differences of the laryngeal epithelia, however, will not be a feature of the 3D model.

2.2.2 Specific issues: vocal fold (V1) anatomy and physiology

There are three issues[39] that are most pertinent to the development of the 3D model concerning the morphology and physiology of the vocal folds and associated structures (V1). The first is the anatomical nature of the thyroarytenoid muscle; the second concerns the relationship of the thyroarytenoid muscle with the vocal ligament; the third issue concerns the impact of the cricothyroid muscle on thyroid and cricoid cartilage orientation, which impacts the longitudinal tension of the vocal folds.

The first issue reviewed concerns whether the thyroarytenoid muscle can be subdived into smaller muscular units. Anatomists have suspected that there are distinct muscular compartments of the thyroarytenoid muscle for a considerable amount of time. The classical anatomist Vesalius remarks that the muscle seems to "triple" in its course (Garrison & Hast 1993: 24). Most frequently two divisions are posited: a medial or internal portion identified as the thyrovocalis (or simply vocalis), which inserts into the vocal process of the arytenoid cartilage, and a lateral or external potion identified as the thyromuscularis (Wustrow 1952; Berg & Moll 1955), which inserts into the oblong fovea and base of the arytenoid cartilage. Elze (1925), Mayet (1955), Sonesson (1960: 19-20) and more recently Zemlin (1997) all support the view that the division is dubious due to lack of evidence for muscle fascia or connective tissue

[39] There is actually a fourth issue: the existence and utility of the superior thyroarytenoid muscle. According to numerous researchers (Kotby et al. 1991; Harrison 1995: 134; Zemlin 1997: 131) it is vestigial. It may play a role in relaxing or reducing tension of the vocal folds, possibly drawing the arytenoids antero-medially (Zemlin 1997: 131), however, it is not considered at this stage in the development of the 3D model.

separating them. Furthermore, the muscle fibres interdigitate considerably (Sonesson 1960: 19), which provides evidence for the case that the thyroarytenoid is a single complex muscle.

The second issue concerning the thyroarytenoid muscle is about the degree to which its muscle fibres insert into the vocal ligament. Goerttler (1950) and his students asserted that the thyroarytenoid muscle inserts into the vocal ligament. They argued that two compartments of the thyroarytenoid muscle could be identified, an anterior and a posterior one, each projecting its insertions obliquely into the vocal ligament. The possibility that muscle fibres insert into, and by extension, controlled the horizontal positioning of the vocal ligaments gave support to Husson's (1950) neuro-chronaxic hypothesis of vocal fold vibration, which attributed vibratory behavior of the vocal folds to cyclic contraction of the thyroarytenoid muscle. The neuro-chronaxic hypothesis has since been refuted by numerous physiological studies, and ultimately eclipsed by the (earlier) myoelastic-areodynamic theory of vocal fold vibration (Berg 1954; Laver 1980: 98; Zemlin 1997: 182). However, the question of how much the thyroarytenoid muscles insert into the vocal ligament still remains relevant. In reviewing the research on this issue that occurred during the 1950's, Sonesson reports that studies in criticism of Goerttler's findings "stressed that such muscle fibres were extremely rare and, when they occurred, they were only few and slender" (1960: 14). Furthermore, Wustrow (1952) observes that many of the fibres Goerttler identified as inserting into the vocal ligament, were not fibre ends, but rather continued inferiorily, likely inserting into the conus elasticus, which has been reported to carry many thyroarytenoid muscle fibre insertions (Sonesson 1960: 14). Hirano & Sato (1993: 45) report that the muscle fibres run in a parallel course to the vocal ligament, based on coronal histological sections taken of the vocal fold. Sonesson's own histological examinations of twenty-four adult and twenty-three new-born excised larynges found that the muscle fibres adjacent to the vocal ligament run longitudinally and parallel to it (1960: 19, 26). Sonesson also notes that the coupling between the ligament and the muscle fibres is strong enough to make removal of the ligament impossible without tearing the adjacent muscle fibres; however, none of these fibres, in any of the specimens, were obliquely directed. The strength of the vocalis-ligament coupling may be because the connective tissue of the vocal ligament blends into the perimysium[40] of the adjacent muscle fibres. Zemlin (1997: 131) reports that his own investigations corroborate the Sonesson study.

[40] A connective tissue sheath the bundles individual muscle fibres into fascicles.

The third issue concerns the action of the cricothyroid muscle, which regulates longitudinal tension of the vocal folds. It is unanimously agreed upon by researchers that the cricothyroid muscle is the primary muscle for stretching the vocal folds along their long axis. However, the impact that this muscle and its corresponding joint have on the cricoid and thyroid cartilages is not fully certain. Generally, researchers concede that it is a rocking motion that increases the distance between the arytenoids (which ride on the cricoid cartilage) and the thyroid notch, where the vocal ligaments and thyroarytenoid muscles originate. It is possible that rotation only occurs for one of these two cartilages, or that both rotate simultaneously (Zemlin 1997: 113). Mayet & Muendnich (1958) are proponents of the view that only the cricoid cartilage rotates, implying that the thyroid cartilage is the more fixed of the two. Vilkman, Pitkanen, & Suominen (1987) attribute different functions to the muscle's subdivisions: the *pars recta* rotates the thyroid down through an angular range of 18°, while the *pars oblique* translates the cricoid cartilage backward and the thyroid cartilage forward. Translational and rotational movements of the thyroid cartilage are suspected by Hong, Kim & Kim (2001: 517) and attested in MRI by Honda et al. (2004), along with differential activation of the muscle's two divisions during speech. Fujisaki, Ohno, & Gu (2004) present a model of the cricothyroid joint and muscle that specify only movement of the thyroid cartilage to simulate tonal contours in Mandarin, Thai, and Cantonese. Titze (2006: 117), on the other hand, mathematically models the activity of the cricothyroid muscle as generating cricoid rotation and translation exclusively. In light of the indeterminancy of exactly how the cartilages are moving with respect to one another, Zemlin (1997: 113-114) concurs with Vennard's (1967) earlier work that both cartilages rotate during contraction of the cricothyroid muscle.

2.2.3 Specific issues: ventricular fold (V2) anatomy and physiology

The ventricular folds are capable of adducting. This is fully evident in research on ethnic "throat singing" styles (e.g. Lindstad 2001). Zemlin asserts, however, that they cannot become tense and their movement is entirely dependent on the arytenoids, to which they attach, in the region of the triangular fovea (1997: 117). Sakakibara et al. (2004a) note that the ventricular folds do not always simultaneously adduct with the vocal folds, which suggests that the ventricular folds might be functionally independent of the vocal folds. Numerous researchers have reported that the folds do carry a few muscle fibres (Saunders 1964; Kaplan 1971; Kotby et

al. 1991; Guida & Zorzetto 2007) and that projections of the thyroarytenoid muscle and thyroepiglottic muscle may facilitate both medial and downward movement of the folds (Agarwal, Scherer & Hollien 2003: 98). From an examination of thirty-two adult human larynges, Reidenbach (1998) reports that there is a distinct muscle sheet, identified as the ventricularis by previous researchers; the muscle is described as a cranioventral extension of the thyroarytenoid muscle containing an anterolateral bundle and isolated anteromedial muscle fibre projections. Reidenbach posits that contraction of the lateral component leads to ventricular fold adduction, while the medial portion may facilitate downward movement of the folds[41]. Sakakibara et al. (2004a) suspect that abduction may be functionally linked to supraglottic constriction, which involves the thyroepiglottic, thyroarytenoid and aryepiglottic muscles exerting a medial force on the folds, resulting in their partial or complete adduction. Sakakibara et al.'s electromyographic research provides evidence that there is strong activation of the thyroepiglottic and vocalis muscles during 'pressed' or ventricular phonation types found in ethnic singing styles (throat singing or drone phonation), aryepiglottic trilling, creaky voice, swallowing, and breath-suspension, when compared with modal or falsetto phonation.

2.2.4 Specific issues: aryepiglottic fold (V3) anatomy and physiology

At the heart of the laryngeal constrictor mechanism are the aryepiglottic folds (V3); they have been identified as the active articulator in the production of pharyngeals and constricted vocal qualities such as harsh voice. Thus, it is worth considering more carefully what the aryepiglottic folds are anatomically and how they are controlled physiologically. Both of these factors will play a role in the 3D model.

The folds themselves are composite structures emerging from the integration of several anatomically distinct tissues that converge at the arytenoids. The body of the aryepiglottic folds are formed by the superior margins of the quadrangular membranes, which originate from the lateral borders of the epiglottis (and the inferior part of the internal angle of the thyroid) and attaching to the anterior surfaces of the corniculate and arytenoid cartilages. Interestingly, these membranes not only form the aryepiglottic folds, but also contribute to the structure of the ventricular folds on their inferior margin (Zemlin 1997: 116). The aryepiglottic folds are invested

[41] This depends on whether the lateral periepiglottic tissue pushes the medial bundles into a horizontal configuration (Reidenbach 1998: 367).

by the cuneiform cartilages. Krmpotic (1957) describes that the posterior face of the cartilage attaches to the colliculus[42] of the arytenoid, which is a protuberance immediately superior to the triangular fovea. Their length ranges from 5 to 9mm, while their girth ranges from 0.5 to 1mm (Poirier, Charpy, & Nicolas 1903). Variation among individuals is considerable; apart from the dimensions, the cartilage can take the form of a single piece or many individual sections. Some individuals have such well developed cuneiforms that they give "l'impression que l'aryténoïde était double" (Krmpotic 1957: 945). The apparent link between the cuneiforms and the ventricular folds supports the functional dependency expressed in the valve model of the laryngeal constrictor mechanism (see Section 2.1.6). These cartilages serve a structural role. They stiffen the aryepiglottic folds "like whale bone in a corset" (Painter 1986: 330) and aid in maintaining a wide epilaryngeal tube aperture (Zemlin 1997: 110), by providing elastic recoil. Negus (1949: 83) also asserts that the cuneiform cartilages play a supportive role in maintaining the shape, stiffness, and erect posture of the aryepiglottic folds. Based on cross-species morphological comparison, he notes that animals with arytenoids large enough to support the aryepiglottic folds tend to lack cuneiform cartilages. Negus also points out that the morphogenesis of the cuneiform, in evolutionary terms, is based on the (lateral margins of the) epiglottis, and in some species the cuneiform is continuous with the epiglottis. Pracy (1984) notes that, often, for the immature human, the cuneiform tubercle is a swelling of glandular tissue that is roughly "sausage" shaped. Their position is roughly 9.5 to 13mm posterior to the thyroid notch (Painter 1986: 330). The aryepiglottic folds also house the aryepiglottic muscle, which is actually a continuation of the oblique fibres of the interarytenoid muscle and lateral fibres of the thyroarytenoid muscle (Negus 1949; Lufkin 1986: 748; Zemlin 1997: 134; Pretterklieber 2003: 250, 253).

Not much is known about the function of the aryepiglottic folds, or how laryngeal sphinctering is accomplished. Harrison (1995: 31) notes based on comparison of mammalian larynges, the aryepiglottic folds and arytenoid complex are considerably elevated, likely due to the absence of a supraglottal tube inserting into the nasopharynx. Thus, its biological function is clearly protective. It is not clear, however, whether the aryepiglottic folds are responsible for actively controlling laryngeal sphinctering. The aryepiglottic muscle fibres are claimed to not be

[42] Latin for "little hill". This is also a possible location of ventricular ligament attachment to the arytenoid cartilage. It also may attach to the triangular fovea or the cuneiform cartilage itself (Krmpotic 1957: 945).

ubiquitous in humans: Van Daele, Perlman & Cassell (1995: 6) report that only two of twenty specimens they examined contained them. While the status of the muscle fibres may be controversial, several researchers agree they exist and even play a role in depressing the epiglottis (e.g. Negus 1949; Ekberg & Sigurjónsson 1982; Meller 1984; Palmer 1993; Ekberg & Pokieser 1997; Zemlin 1997; Pretterklieber 2003; Sakakibara 2004a, 2004b). The movement of the epiglottis, which is related to aryepiglottic function, is a much more thoroughly researched aspect of sphinctering, and its function offers insight into aryepiglottic function. Ekberg & Sigurjónsson (1982: 106) posit that three independent muscles might actively depress the epiglottis during swallowing: the stylopharyngeus, thyroepiglottic, and aryepiglottic muscles. Zemlin (1997: 116), however, claims that the aryepiglottic muscles are "poorly developed" and do not play a major role in protective closure of the laryngeal tube during swallowing. Ekberg & Pokieser (1997: 8) note that the aryepiglottic folds may not be responsible for approximation of the epiglottis[43] to the arytenoid region, although not as far as the esophageal inlet. Petterklieber (2003: 254) rules out the thyroepiglottic muscles, which he claims are the antagonists of the aryepiglottic muscles, and serve to broaden the aperature of the laryngeal collar/inlet. Electromyographic evidence collected by Sakakibara et al. (2004a) shows that the thyroepiglottic and thyroarytenoid muscles (the vocalis portion) are in fact highly activated during swallow, breath holding, and to a lesser extent in creaky voice, and growl (i.e. aryepiglottic trilling), which runs contrary to Pretterklieber's observation, but supports an earlier study of epiglottal function done by Ekberg & Sigurjónsson (1982: 105-106). In the Ekberg & Sigurjónsson study, the thyroepiglottic muscle is described as inserting into the lateral margin of the epiglottis and into the aryepiglottic fold itself; they assert that contraction of the muscle produces an antero-medial force, possibly drawing the aryepiglottic fold and cuneiform cartilages forward. Sakakibara et al. (2004a) posit that the aryepiglottic muscles can approximate the arytenoid cartilages and epiglottis during aryepiglottic constriction, however, their electromyographic investigation of aryepiglottic fold activity failed due to the difficulty of inserting a hooked wire electrode into the aryepiglottic muscle. Esling, Zeroual, & Crevier-Buchman (2007) suspect that the thyroarytenoids are the primary agonists of laryngeal constriction, given that laryngealized or creaky voice, which is achieved through thyroarytenoid contraction, and constricted articulations, results in aryepiglottic constriction. Ultimately, as the thyroepiglottic muscle is in

[43] Of course, one cannot forget that the epiglottis is also dependent on tongue root position as well.

some regards an extension of the thyroarytenoid muscle, it is likely that the burden of laryngeal sphincter during deglutition and constricted articulations is a result of synergistic activity of the muscles.

The overall dimensions of the aryepiglottic folds could not be obtained for adult speakers; however, infant aryepiglottic fold widths are reported at three locations for 100 youths[44]: thickness of the aryepiglottic folds at their midpoint, thickness of the aryepiglottic folds just behind the epiglottis, and thickness at the base of the folds (slightly above the level of the ventricular folds). Respectively, these measurements are as follows: 1.80mm, 1.70mm, and 6.95mm.

2.2.5 Anatomical scope of the 3D LCM

The goal of this section has been to describe the anatomical scope of the model and review controversial aspects of the laryngeal architecture that most deeply impact how the model is designed. It has been necessary in some instances to make decisions about issues that are not yet resolved in the literature in order to build the model. The model is, however, flexible and extensible: as more research is done that fine tunes our understanding of laryngeal morphology and physiology, it is possible to integrate new findings into the model. In this section, the morphological and physiological issues reviewed in this section are summarized and the specific details concerning implementation of anatomical research are presented. The model is also intended to be developed in phases, with the version presented in this thesis being the first phase. Later phases will expand on the anatomical scope.

The first phase of the model starts at the logical beginning of the laryngeal vocal tract, the larynx. By representing the cartilaginous and muscular framework of the larynx, it is possible to present pharyngeal and epiglottal articulations in the most coherent and contextualized way, and to illustrate the relationship between glottal configurations and supraglottal ones. While the pharynx proper and the tongue root play a role in the formation of pharyngeal articulations, as has been discussed in Section 2.1.2, the primary articulator in pharyngeals and epiglottals, is the aryepiglottic sphincter mechanism. Thus, for this phase, the overall model will be of the laryngeal architecture, with a focus on the aryepiglottic folds. A listing of structures represented in the model is provided in Table 2.7.

[44] 6 months to 12 years with a mean age of 2.5 years

Table 2.7 Summary of anatomical details of the 3D Laryngeal Constrictor Model (Phase 1)

<table>
<tr><th rowspan="2">Gross Structure</th><th colspan="4">Tissue Type</th></tr>
<tr><th colspan="2">Connective</th><th colspan="2">Muscle</th><th>Epithelia</th></tr>
<tr><td rowspan="4">Larynx</td><td>bones</td><td>hyoid</td><td rowspan="2">intrinsic</td><td rowspan="2">thyroarytenoid
cricothyroid
post. cricoarytenoid
lat. cricoarytenoid
interarytenoid
aryepiglottic
thyroepiglottic</td><td rowspan="4">aryepiglottic folds
vocal folds
ventricular folds
epiglottis</td></tr>
<tr><td>cartilages</td><td>thyroid
cricoid
arytenoids
corniculates
cuneiforms
epiglottal</td></tr>
<tr><td>ligaments</td><td>ceratocricoid
post. crioarytenoid
vocal
thyroepiglottic</td><td rowspan="2">extrinsic</td><td rowspan="2">-</td></tr>
<tr><td>membranes</td><td>lateral cricothyroid
conus elasticus
quadrangular</td></tr>
</table>

Concerning the vocal folds (V1), the simplest approach to representing the thyroarytenoid muscle is to design it as a single muscular unit. The literature on the putative divisibility of this muscle indicates that the simple approach is justified, as no conclusive evidence can be levied to argue for a distinct division of the muscle. The terms vocalis (internal thyroarytenoid) and thyromuscularis (external thyroarytenoid) are interpreted as referring to regions of the muscles, instead of distinct subsections. The thyroarytenoid also has been shown to run in a parallel course with the vocal ligament, but there is some measure of coupling between the two. In the 3D model, the ligament will be shown to run parallel to the muscle, but directly influence the position of the muscle to capture this observation. As no conclusive decision about how the cricoid cartilage rotates with respect to the thyroid cartilage under cricothyroid activity, the model will represent the effect of this muscle as equally rotating both cartilages.

The ventricular folds (V2) are likely infused with muscle fibres, constituting a ventricularis muscle. Despite the persistent claims for the existence of this muscle, the model will not include a representation of it, as the research indicates that the muscle is sparse and closely related to the thyroepiglottic muscle. A plausible effect that the muscle might have is to

cause the ventricular folds to bunch together with the vocal folds, increasing the effective mass of the latter. This might play a role in facilitating the damping effect they have on the vocal folds during the movement referred to as ventricular incursion (see Section 2.1.4). Medial movement of the ventricular folds is assumed to be a result of epilarygeal constriction of the thyroepiglottic, thyroarytenoid, and aryepiglottic muscles.

Finally, the aryepiglottic folds (V3) are assumed to be invested with an aryepiglottic muscle that is responsible for causing the folds to bunch together with base of the epiglottis, or "tighten the laryngeal inlet in the same manner as the string in a tobacco pouch" (Ekberg & Sigurjónsson 1982: 106) which they facilitate in drawing in a posterodorsal direction. It is clear that they are not the sole force responsible for drawing the epiglottis down over the laryngeal aditus, but this aspect is not considered in this phase of the model. The antero-posterior positioning of the aryepiglottic folds is suspected to be influenced by the activity of the thyroepiglottic and thyroarytenoid muscles as well. The thyroepiglottic muscle is assumed to aid in pulling the aryepiglottic folds in an anteromedial direction, impacting the location of the cuneiform cartilages as well. The thyroarytenoid muscle, on the other hand, is capable of generating a moderate anterior translation of the arytenoid cartilages, which will also assist in sealing the epilaryngeal tube. Together, with the activity of the aryepiglottic muscle, complete closure is represented as being achieved by these muscles in the 3D model pending future research that runs contrary to these assumptions.

A final note must be made concerning details identified as critical, but *absent* from the current phase of model development: the tongue root (V4), the physiology of laryngeal height (V5), and the pharyngeal tube (V6). All of these features are planned for implementation in the complete model, but the primary focus is on the aryepiglottic folds, and because the modeling of any one component is a very involved task, time does not permit including the tongue root, pharyngeal tube, or larynx height as features in the current phase of model development.

This concludes the review of laryngeal anatomy as it pertains to design of the model both in terms of morphology and physiology. The actual implementation of these details is outlined in Chapter 3, which documents how the model is constructed and the basic movement properties of the laryngeal model.

2.3 Biomechanical Models

An important part of tissue function in articulatory processes is its movement characteristics when subjected to an external driving force. Under aerodynamic conditions within the vocal tract, tissue response is typically oscillatory. The specific physical properties of the tissue considered are what give rise to variation of the oscillation that occurs. The prime example of this phenomenon is the vibratory movement of the vocal folds when they are subjected to sufficient subglottal pressure. It is in this sense that the biomechanical models discussed here are referred to as *self-oscillating*, which distinguishes them from theoretical models that propose neurological control of the oscillating function, such as the neuro-chronaxic theory (Titze 2006: 339). Since the 1960's, it has been firmly established that vocal fold oscillation is due to biomechanical and aerodynamic processes rather than neurological and muscular ones, and this view is encapsulated within the Myoelastic Aerodynamic Theory of Phonation, which is the most widely recognized theory of vocal fold function.

Biomechanical models of the vocal folds are used as a means to synthesize, and thereby analyze the properties of vocal fold function under varying conditions. The objective of creating biomechanical models is to approximate vocal fold physical characteristics using finite models. Most of these models are discrete or lumped combinations of physical elements or model primitives, whose mechanical properties are well understood (Shabana 1991). In the context of a vibratory system, the elements are typically springs, masses, and dampers coupled together in ways that reflect the state of the system being modeled, with boundary conditions to restrict/constrain their movement (see Figure 2.1). Continuous models are also a possibility, drawing on the principles of continuum mechanics, but require considerably more complexity to model, which is not always desirable from a computational perspective.

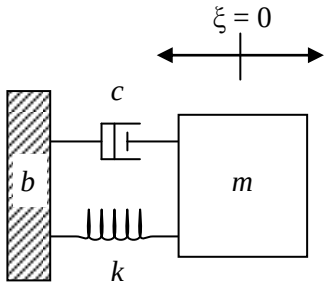


Figure 2.1 An illustration of the basic physical elements typically used in biomechanical models: springs, masses, and dampers. The spring represents a coupling between two bodies in the model (in this case between the mass *m* and the boundary *b*) with spring/stiffness

coefficient *k*. Displacement of the mass will either stretch or compress the spring, generating a restoring force in the opposite direction. Element *c* is a viscous damper or dashpot (mathematically: the coefficient of viscous damping), used to dissipate energy within the vibrating system by introducing force that is proportional to the velocity of the mass.

The goal of this section is to provide a basic outline of the self-oscillating mechanical model of the vocal folds, by reviewing a handful of classic examples. Of particular importance is the Sixteen-Mass Model proposed by Titze (1973, 1974), which is used as the basis for the physical model of the vocal folds in the 3D LCM. Furthermore, Titze's model is used as conceptual starting point for the construction of a parallel model of aryepiglottic fold function during trilling (discussed in Section 5.3). The models reviewed here reflect the level of detail sought after for the development of the aryepiglottic model. That is, while recent models are employing increasingly sophisticated model formats (such as deformable body and continuum models[45]), a simpler model has the benefit of being computationally inexpensive and easier to construct and thus is the approach adopted here.

2.3.1 Simple biomechanical models

The most basic observation about vocal fold movement is that they move along a latero-medial dimension. Among the first self-oscillating models described are those by Flanagan & Landgraf (1968) and Ishizaka & Flanagan (1972), which constitute 1-mass and 2-mass models, respectively. The masses are attached to compressible springs (or mechanical compliances) that characterize the coupling stiffness responsible for generating a restoring force if one of the masses is displaced from its equilibrium position. Movement is limited to the latero-medial direction and driven by applying forcing functions that are sensitive to the glottal state. Figure 2.2 provides an illustration of the spring-mass type of model. The Flanagan & Landgraf model accounts for contact mechanics of the folds. Upon medial collision, the model implements two different constraints on vocal fold momentum depending on the prevailing conditions of the vocal folds at the time. Modal phonation implements a viscous vocal fold collision, which results in a loss of fold momentum and a momentary state of closure. When the folds are under

[45] For the most modern approach to biomechanical modeling of laryngeal structures, the reader is encouraged to consult Titze's (2007) book *The Myoelastic Aerodynamic Theory of Phonation*, which provides a comprehensive treatment of glottal biomechanics using sophisticated models that go beyond low-dimension body-cover models of the vocal folds.

considerable tension, as is the case during high pitched phonation and falsetto, the fold collision is considered hard and there is a rebounding force generated.

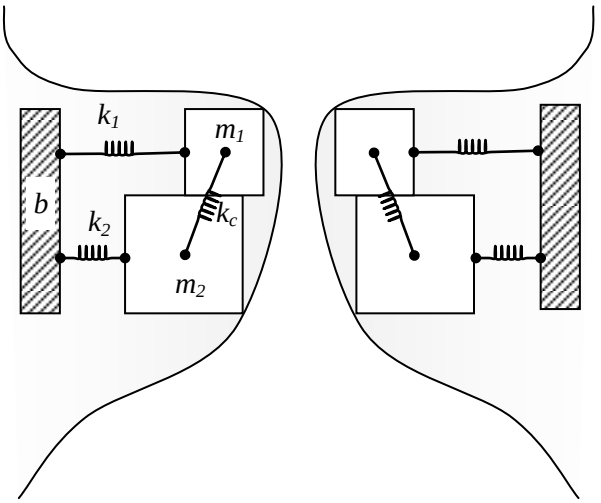


Figure 2.2: A two-mass biomechanical representation of the vocal folds (adapted from Ishizaka & Flanagan 1972; Stevens 1977, 1999; Titze 2006). The upper mass m_1 represents the mucosa of the vocal folds, couple to the lower mass m_2, representing the body (ligament and vocalis muscle), and the boundary (the thyroid cartilage).

Hirano (1974) advanced the important idea that the vocal fold mucosa participates in the oscillating process by transmitting a surface or mucosal wave across the vocal folds for every cycle of vibration. The Flanagan & Landgraf model does not include a representation of this body-cover distinction, rendering it inadequate to describe more complex modes of vibration (Stevens 1977: 266-267; Titze 2006: 150). The Ishizaka & Flanagan model, however, allows for a differentiation in size between the two mass elements that comprise the vocal folds, which reflects the body-mass distinction. Consequently, vertical phase differences, which characterize modal phonation, can be represented (Stevens 1999: 56-58).

2.3.2 Titze's 1973-1974 biomechanical model of the vocal folds

The representative spring-mass models presented thus far are limited to one-dimensional motion and fail to capture some of the more complex behaviours of the vocal folds, particularly with respect to the modes of vibration and the horizontal phase characteristics of vibration (Stevens 1977, 1999; Titze 1973, 2006). To elaborate on the spring-mass model, Titze (1973, 1974) describes a sixteen-mass model that represents the vertical and horizontal phase

characteristics, with particular focus on differentiating the vocal fold body from the cover. The sixteen masses are arranged in two longitudinally oriented rows[46], one for the vocal fold body (ligament and muscle) and one for its cover or mucosa, which are coupled together along their entire length. A lateral boundary serves as an anchoring point for the string of particles representing the body. The particles are confined to transverse planar motion, one direction normal to the direction of flow, and one perpendicular to it, which means that each particle has two degrees of freedom. Particle masses are used to obviate the need for rotational degrees of freedom and a model of tissue deformation; however, dimensions are specified for each tissue when aerodynamic conditions are considered. Figure 2.3 illustrates the overall layout of the sixteen-mass model.

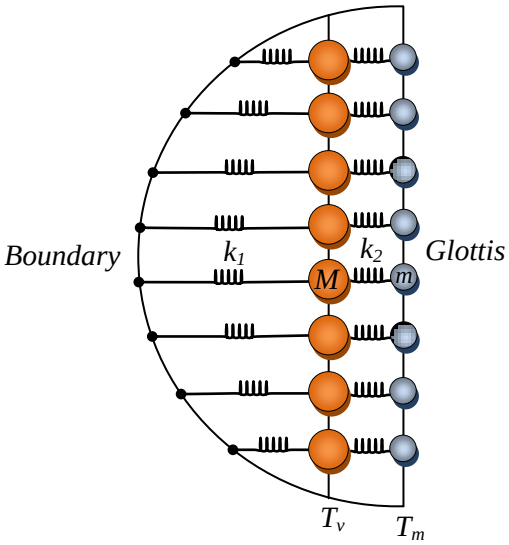


Figure 2.3: Titze's (1973, 1974) 16-mass model of the vocal folds (adapted from Titze 1973: 140). The representation is symmetric, thus, only one vocal fold is shown. The eight large orange circles, designated *M*, are particle masses of the vocal fold body, while the eight small blue circles, designated by *m*, are of the vocal fold mucosa. The mucosa and body are coupled together with spring coefficient k_2 and the body particles are coupled to the boundary with spring coefficient k_1. Longitudinal tension of the particle sets/strings is specified by T_v for the body and T_m for the mucosa.

Considering only one vocal fold, movement of the particles is determined by numerically solving 32 differential equations, one for each degree of freedom, that specify the dynamics of

[46] Eight masses per tissue "string" is considered an optimal number of particles because it is sufficient to represent the eigenmodes (normal modes) of vocal fold vibration, where "energy carried in modes beyond the eighth will be negligible" (Titze 1973: 141). Additionally, inspection of high speed laryngoscopic videos of tissue vibration at extremely high pitches (ca. 1000Hz) reveals that eight longitudinal degrees of freedom are necessary to represent the standing wave pattern that occurs.

the particle system. Particle motion is determined by the sum of forces (Eq. 2.3) acting on the particle, which include: nearest neighbour restoring forces (Eq. 2.1), damping forces (Eq. 2.2), and external (aerodynamic) forces (discussed next). Maximum strains for the muscle and ligament of the vocal fold are determined based van den Berg's (1960) research on the elastic properties of the two tissues, while the elastic properties for the mucosa are introduced and heuristically adjusted, due to lack of research on the subject. This is required to represent tissue indistensability at maximum strain. Particle constraints are imposed to prevent anomalous behaviour due to finite step sizes in the simulation[47].

Eq. 2.1 Restoring forces for the i^{th} particle (vocal fold model)[48]

$$F_{xi}^{(1)} = T_i\,(x_i - x_{i-1}) \div r_i \qquad F_{xi}^{(3)} = k_1 S_{1i}$$
$$F_{xi}^{(2)} = T_{i+1}\,(x_i - x_{i+1}) \div r_{i+1} \qquad F_{xi}^{(4)} = k_2 S_{2i}$$

k = spring coefficient
r = distance between particles
S = lateral nearest-neighbour strains ($r_n - r_0$)
T = particle string tension

Eq. 2.2 Damping forces for the i^{th} particle (vocal fold model)[49]

$$D_{vi} = 2\zeta\sqrt{M k_{vi}}$$
$$F_{xi}^{(d)} = D_{vi}\dot{x}_{vi}$$

D = damping
ζ = damping factor (≥ 0.16)
$\dot{x}$ = velocity[50]

Eq.2.3 Equations of motion (vocal fold model)

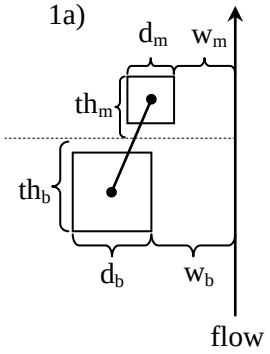
$$M_i\ddot{x}_i = F_{xi}^{(1)} + F_{xi}^{(2)} + F_{xi}^{(3)} + F_{xi}^{(4)} + F_{xi}^{(d)} + F_{xi}^{(e)}$$

$\ddot{x}$ = acceleration

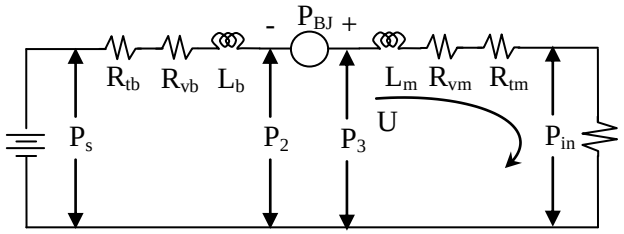
[47] Unlike particle behaviour in real life, the state of the simulation is updated every time step, which are discrete given the finite nature of digital simulation. This leads to the possibility for a particle to be unaffected by forces that would be imposed under a continuous circumstances, which potentially results in particle overlap or other unnatural conditions.
[48] Only the *x* coordinate is shown in equations (1) to (3). Equations $F^{(1)}$ and $F^{(2)}$ apply to both mucosa and body particles. The mucosa is not coupled to the boundary so it lacks a forth restoring force.
[49] Only the body damping equation is shown; the mucosa is the same with the appropriate subscripts and symbols.
[50] The 'dot-over' notation indicates differentiation with respect to time.

Much of the oscillatory behaviour of the vocal folds is determined by aerodynamic conditions, particularly with respect to the characteristic vocal fold tissue positioning at each phase in the phonatory cycle. Titze (1973: 151) notes that aerodynamic forces are in fact largely responsible for determining the frequency of phonation. To determine the aerodynamic forces acting on each tissue region (the body and the mucosa) air pressures through the glottis are analyzed at various locations: subglottal pressure, pressures at the beginning and end of the body as well as the mucosa, and the exiting glottal pressure, which serves as input to the vocal tract. The forces based on the glottal pressures take on unique forms under four different glottal conditions. These are reviewed in figure 2.4 along with the force equations that are applied to the body and mucosa particles.

1a)



1b)



Symbols:
= resistor (R)
= inductor (L)
= source
= indep. source

Subscripts:
b = body
m = mucosa
t = turbulent
v =viscous
BJ = Bernoulli pressure change

Areas:

$A_s = \pi r^2$

$A_b = \left(x_b - \frac{d_b}{2}\right) a$

$A_m = \left(x_m - \frac{d_m}{2}\right) a$

Resistance:

$R_{vb} = 12\mu a^2 \frac{\mathrm{th_b}}{\mathrm{A_b^3}}$

$R_{vm} = 12\mu a^2 \frac{\mathrm{th_m}}{\mathrm{A_m^3}}$

$R_{tb} = 0.875 \frac{\mathrm{P_{bb}}}{|\mathrm{U}|}$

$R_{tm} = 0.875 \frac{\mathrm{P_{bm}}}{|\mathrm{U}|}$

$R_T = R_{vb} + R_{vm} + R_{tb} + R_{tm}$

Inductance:

$L_b = \frac{\mathrm{th_b}}{\mathrm{A_b}}$

$L_m = \frac{\mathrm{th_m}}{\mathrm{A_m}}$

$L_T = \mathrm{L_b} + \mathrm{L_m}$

Bernoulli Pressures:

$P_{bb} = \frac{1}{2}\rho \frac{\mathrm{U^2}}{\mathrm{A_b^2}}$

$P_{bm} = \frac{1}{2}\rho \frac{\mathrm{U^2}}{\mathrm{A_m^2}}$

$P_{BJ} = \mathrm{P_{bb}} - \mathrm{P_{bm}}$

Pressures:

P_s = free variable

$P_1 = \mathrm{P_s} - \mathrm{P_{in}} - 1.37\mathrm{P_{bb}}$

$P_2 = \mathrm{P_1} - (\mathrm{R_{vb}} + \mathrm{R_{tb}})\mathrm{U} - \mathrm{L_b}\dot{\mathrm{U}}$

$P_3 = \mathrm{P_2} + \mathrm{P_{BJ}}$

$P_4 = \mathrm{P_{in}} - \rho|\mathrm{U}|^2 \frac{(1 - \mathrm{A_m}/\mathrm{A_s})}{\mathrm{A_m A_s}}$

$P_{in} = -(\mathrm{RU} + \mathrm{L}\dot{\mathrm{U}} - \mathrm{P_{BJ}} - \mathrm{P_s})$

Subscripts:
b = body (vocalis & ligament)
m = mucosa
x = x-axis
y = y-axis
T = total
BJ = Bernoulli pressure change at the particle junction

Dimensions:
d = depth
r = tracheal radius
th = thickness
w = width of hemiglottis

Quantities:[51]
μ = coefficient of viscosity (0.000186 g/cm - s)
ρ = air density (0.00114 g/cm^3)
U = volume velocity
$\dot{\mathrm{U}}$ = volume acceleration

Pressures:[52]
P_s = subglottal pressure
P_1 = body start pressure
P_2 = body end pressure
P_3 = mucosa start pressure
P_4 = mucosa end pressure of the
P_{in} = pressure input to VT (exit pressure)

[51] Constants are obtained from Titze (2006: 240); for the sake of comparison, Stevens (1999: 27) provides 0.000194 dynes/cm^2.

[52] Pressure between P_1 & P_2 and P_3 & P_4 is assumed to be linear (Titze 1973: 154).

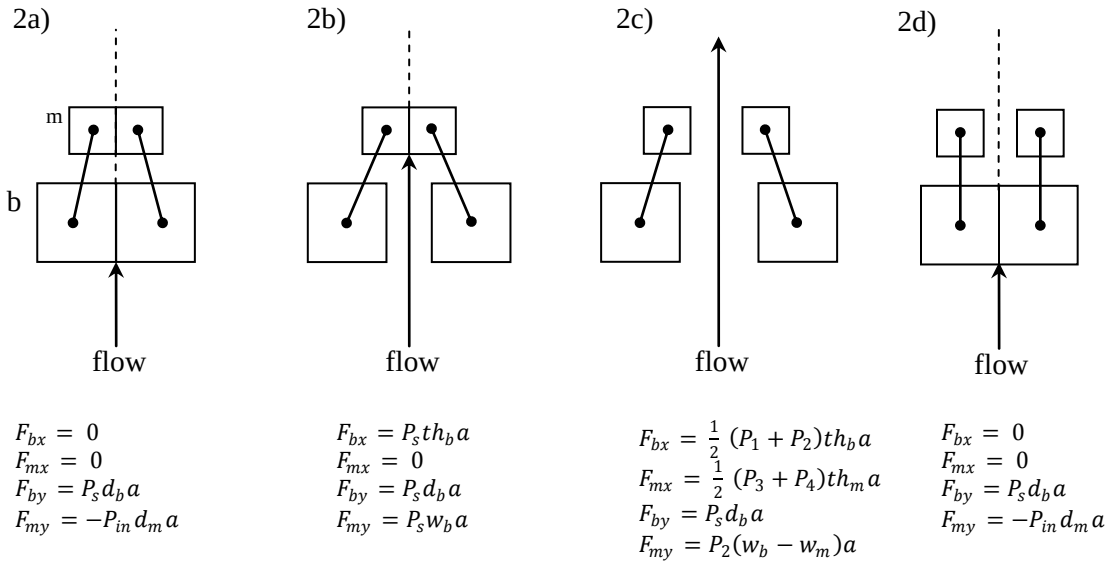


Figure 2.4: Diagrams and corresponding equations of the aerodynamic component of Titze's 16-mass model; all diagrams are adapted from Titze (1973: 153). Diagram (1a) is a schematic showing particle dimensions used in calculating aerodynamic properties such as areas, resistances, inductances, and the forces acting on the particles. Diagram (1b) is an equivalent circuit for volume flow (U) in the glottis. Diagrams in (2) represent the four different force conditions used to determine which set of force equations are applied, which are provided below each diagram. The diagrams are ordered sequentially to illustrate the phases of a single glottal pulse, with vocalis-mucosa phase delay.

Titze's 16-mass model is focused on glottal dynamics and as such does not explore the epilaryngeal vocal tract in detail. This model, however, is essential to the present 3D model as the biomechanical simulation of the vocal folds it implements is based directly upon it. Furthermore, the simulation of the aryepiglottic folds developed for the 3D model is directly inspired by Titze's model. This is largely on account of the virtual absence of a more suitable precedent for modelling the biomechanical aspects of the arypieglottic folds. Perhaps the closest match is that of Sakakibara et al. (2004a), who present a 2-mass model that represents both glottal and ventricular masses in order to synthesize vibratory patterns in ethnic singing styles found in the Altai Mountains that employ ventricular fold vibration. Their model, however, does not explore aryepigloittic fold vibration as observed in aryepiglottic trilling. Titze's (1973) model is regarded as an ideal framework for the 3D Laryngeal Constrictor Model's aryepiglottic trilling biomechanical model as it provides enough degrees of freedom to visually synthesize aryepiglottic fold movement, while remaining a tangible computational prospect, as opposed to the loftier model presented in Titze (2006).

2.4 Summary of the foundation of the model

The goal of this chapter has been to elucidate the constellation of issues that pertain to the laryngeal vocal tract and the attempts made to understand its linguistic function and its anatomical and physiological nature. Table 2.8 provides a summary of the linguistic, anatomical, and physiological scope of the model. Much of the specifics relating to language specific articulation are not explicitly explored in the 3D model; however, the research that has been conducted on these languages forms the backdrop for the development of the model and provides linguistic context. Most important is the articulatory theory that has been developed on a basis of the empirical research into the articulatory properties of these languages. Much of the attention of the early parts of this chapter has been focused on the research and theoretical proposals made by Esling, Edmondson, and their colleagues. The model itself is based upon this research and the emergent observation that the aryepiglottic constrictor is seminal in what is described broadly as laryngeal constriction.

To complete the description of this articulator, the anatomical and physiological aspects of the larynx have been considered. The anatomical domain is by no means free of controversy, and it seems that the structures of most import to the model, i.e. those of the epilaryngeal tube, are indeed considerably varied from individual to individual (Crevier-Buchman, L., personal communication). Nevertheless, there is sufficient documentation on the general form and function of the aryepiglottic folds to design a model that approximates their behaviour. Biomechanical modeling is solely based on the work that has been conducted on the vocal folds, as precedence for aryepiglottic fold models is absent.

The following chapters now introduce the 3D Laryngeal Constrictor Model itself, beginning with the rudimentary aspects of its development (Chapter 3) and working on up, through modeling laryngeal posturing and articulation (Chapter 4) to the more sophisticated functions such as the biomechanical modeling of the vocal and aryepiglottic folds (Chapter 5).

Table 2.8 Summary of the scope of Phase One of the 3D LCM

Domain	Details in Phase One of the 3D LCM
Linguistic	V1, V2, V3; segmental articulations, phonation types
Anatomical	vocal, ventricular, & aryepiglottic folds; laryngeal structures
Physiological	laryngeal movement; vocal & aryepiglottic fold vibration

Chapter Three

Developing the Model and Determining Basic Movement Characteristics

3.1 Orientation to 3D computer model development

This chapter deals primarily with the materials and methodology used in the basic rendering of the 3D Laryngeal Constrictor Model. Section 3.2 discusses the use of histology to create digital representations of laryngeal structures using a process dubbed *vertex tracing*. Section 3.3 covers the implementation of simple mathematical constraints on the laryngeal structure featured in the model, with a focus on vocal, ventricular, and vocal fold posturing.

Representations of cartilages, muscles, ligaments, and other tissues are constructed as 3D polygon *meshes* in the 3D Laryngeal Constrictor Model. In its essence, a mesh is a computer object that describes the geometry of a structure it represents using vertices, which are graphical primitives that contain information about position in 3D space, and additional data such as surface normals (used in lighting calculation), texture, and potentially color information. Vertices are organized in such a way that the computer can connect them together to form polygons (typically triangular), which are the basic surfaces of the represented object. The process of creating the 3D representation involves specifying the positions of the vertices (in what is referred to as *model space*).

There are two subclasses of meshes: *static* and *dynamic*. In the 3D model, static meshes represent those structures that typically do not deform (e.g. stretch, flex, or contract). Dynamic meshes, on the other hand, are used to represent structures where deformation is a physiological necessity. Rendering three-dimensional scenes on computer is relatively straight forward using static meshes. Their position can be manipulated by combinations of simple transformations (translation, rotation, and scaling) applied to their vertex data. Rendering dynamic meshes is a considerably more complex matter. These meshes can be controlled by the use of a *frame hierarchy*, which allows a mesh to be subdivied into sets of vertices that are linked together like bones in a skeleton, which can then be independently manipulated.

Beyond rendering meshes, the program also maintains a physical simulation of the particle systems that are used to represent mechanical vibration in the larynx (i.e. of the vocal folds and the aryepiglottic folds). Details concerning the aryepiglottic fold biomechanical simulation are discussed in Section 5.3.

3.2 Mesh Construction

As discussed in Section 1.3, meshes in models of the vocal tract are typically developed from MRI or CT images. However, for the model discussed here, access to these modalities was unavailable and thus an alternative and somewhat innovative approach was employed: anatomical structures in histological sections of the larynx have been segmented and reconstructed into the 3D meshes. Depending on the circumstances, future versions of the model may employ MRI or CT imaging to develop more fine tuned meshes from living subjects.

3.2.1 Mesh source data

Much of the histological data was obtained from the work of Hirano and Sato published in their *Histological Color Atlas of the Human Larynx* (Hirano & Sato 1993). It is important to note that this data set is incomplete with regards to the needs of the 3D Laryngeal Constrictor Model (see next paragraph). Furthermore, the axial histological sections come from a different individual than those of the coronal sections: the axially sectioned specimen comes from a 57-year old Japanese speaking female, while the coronal sections were obtained from an 81-year old Japanese speaking female. Most of the meshes in the 3D model are based on the structures preserved in the axial sections. However, there are certain structures which are more fully represented by the coronal set.

The histological data set is limited in terms of the extent of the laryngeal architecture included in the sections of both planes: the axial sections extend from 12 mm below the glottis to 13 mm above, while the coronal sections extend from the anterior commissure tendon to 8 mm posterior to the posterior macula flava. In the axial plane, interslice distance is 2 mm, except the section beginning at 3 mm below the glottis and ending at 2 mm above the glottis, where the interslice distance is 1 mm. The anterior most coronal sections are described geographically according to the anatomical landmarks with which they intersect (anterior commisure tendon, anterior macula flava, and quarter sections of the vocal folds. No interslice distance is provided until the posterior macula flava is reached, at which point the interslice distance is 2 mm. Due to the indeterminacy of the position of the anterior coronal histological sections, the primary sections used to obtain traces for the model meshes are the axial ones. While this data set provides excellent images to reconstruct all major laryngeal cartilages, muscles, epithelial tissue,

membranes, ligaments, and even vascular, neural, and adipose tissue, there are some notable structures absent from the data set, and truncations of other structures. The inferior surface of the cricoid cartilage, the superior portion of the epiglottis, cuneiform cartilages, the upper tips of the corniculate cartilage, the upper extent of the aryepiglottic folds, hyoid bone, the middle portions of the superior surface of the thyroid lamina, and the upper portion of the superior horns of the thyroid cartilage are all absent from the data set. Also not evident in the data set are the pharyngeal cavity, the tongue root, and the extrinsic laryngeal musculature.

Due to these problems, it was necessary to base the 3D model on a composite body of data gathered from various sources, including anatomy textbooks, MRI and CT studies of laryngeal anatomy, in addition to the histological data. The goal of representation was to strive for generalized or idealized anatomy. A central assumption affecting the generality of the model meshes is that the structures are bilaterally symmetrical. Furthermore, any structural abnormalities observed on the specimens where ignored and an attempt was made to correct for this by manually interpolating the correct structural details. Thus, in addition to the smoothing of the various irregularities found on the specimens presented in Hirano & Sato (1993), only one half of any given symmetrical structure is represented. Inevitably, the overall process is highly subjective and, therefore, achieving a higher level of accuracy stands as one of the goals for further development of the model, when a more complete data set avails itself[53].

3.2.2 Image data scaling & alignment

Blender 2.45 (win32 dynamic version), an open source 3D modeling suite from Blender Foundation, was used as platform to construct the model meshes. This platform is ideal as it is possible to import the histological data images and scale them to a grid, allowing for the images to be orthogonal to the Cartesian coordinate axes. This was important as it made certain that the complete meshes were exported from the modeling program pre-aligned to the axes, and that all meshes were congruently aligned with one another. The result is that the meshes' model space matches their default orientation when they are rendered into world space[54].

[53] Such as MRI: Selbie, Gewalt, & Ludlow's (2002) cartilage meshes are a fine example of this type of refinement. However, it should be noted that with increasing anatomical accuracy, the degree to which the meshes exhibit individual irregularities is also increased, which is not desirable if the goal is generalized morphology of the larynx.
[54] Model space refers to the coordinate space of the mesh (vertex coordinates) and world space refers to the shared coordinate spaces all meshes are placed in before being subjected to rendering by the computer hardware.

The first step in creation of the model meshes was to obtain traceable data sources (e.g. histological sections or MRI and CT images). For non-digital image sources, the originals were all scanned at a constant size and resolution to ensure that there were no differences in scale or quality from one image to the next in an image sequence. The first of a sequence of images was then imported into Blender and scaled to the grid system by attributing a value of 1 mm^2 to a single grid unit (Figure 3.1). Average measurements obtained from Eckel, Sittel, Zorowka, & Jerke (1994) and Titze (2006) were used to constrain this process. It is acknowledged that these measurements were obtained from entirely different sources than the specimens used for mesh construction. This, however, was assumed to be an irrelevant source of variation as the global size of the model is not as important as the relative proportions of the meshes to one another. Since the histological slices are from female larynges, measurement data from female larynges was used wherever possible, as is the case with Eckel et al. (1994: 35-36).

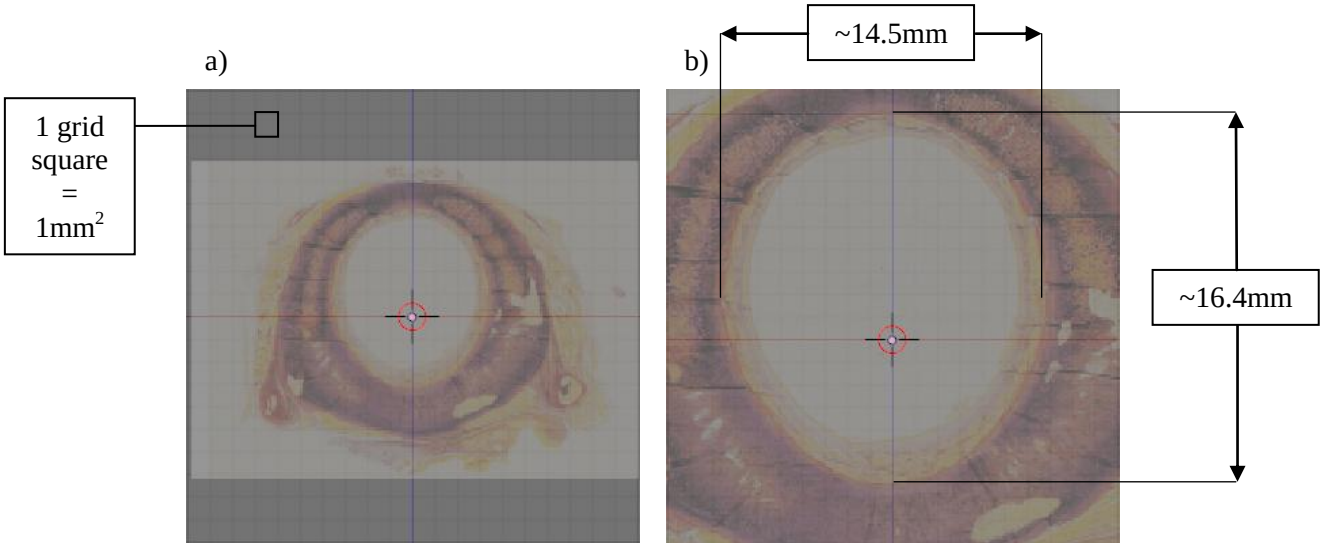


Figure 3.1: An illustration of how scaling was implemented using the Blender rendering grid. In (a), the basal most section (at 12mm below the glottis) from the axial histological data set was loaded into Blender as a transparent image with no scaling applied. In this slice the caudal portion of the cricoid cartilage is visible along with the inferior cornu of the thyroid cartilage. Each square on the grid was attributed an area of $1mm^2$, which forms the basis of scaling. Shown in (b) is the scaled image. Scaling was accomplished by approximating the scaling value so that the measured areas corresponded with the average measures of those distances. Specifically, the measurements used are the internal sagittal distance of the caudal edge of the cricoid cartilage (16.4mm) and the internal transverse distance of the caudal edge of the cricoid cartilage (14.5mm) as reported in Eckel et al. (1994: 36, measurements #56 and #60 respectively).

Once a scaling was established the scaling could be applied to every image in any given data set. These scalings do not, however, transfer between data sets as the size of the images

cannot be guaranteed to be of the same scale. Once again Eckel et al. (1994) measurements were used wherever they could be. In the event that a data set did not contain the structural relationship measured in Eckel et al. (1994), scaling of the mesh was postponed until the rendering process.

3.2.3 Vertex tracing

The next step in mesh construction is vertex tracing. Vertex tracing involves directly tracing a structure of interest in a data image using vertices. The number of vertices impacts the resolution of the model as it determines the number of polygons (triangular) that make up the mesh. A balance between detail and computational expense must be maintained. The structures being represented are organic and continuously vary in their contours. As vertex tracing is essentially a way in which to digitize the structures it necessarily involves a degree of quantization. No explicit limit was set on the number of vertices per unit of distance. To facilitate rapid tracing, the degree of approximation of structural contours, determined by the number of vertices used, was decided subjectively. The only impact that this has on the resulting mesh is the overall smoothness of its appearance. Again, given limitations on processing power, a certain degree of discreetness in the appearance of the mesh is to be expected (see Figure 3.2).

Only one-half of a given structure was traced as midsagittal symmetry is assumed. Where the midline ‘seam’ vertices are concerned, they were left until the entire half mesh was complete, at which point they were joined to their mirror image counterpart. Paired structures were traced in their entirety, but only one of the pair is traced. Figure 3.2 illustrates the vertex tracing process. Once one image in the sequence was traced, the process started over again for the subsequent image in the histological data set. Tracing of axial planes was done in the *x*-*z* plane, while coronal sections were traced in the *x*-*y* plane. The height of the next trace was determined by the interslice distance on the *y*-axis. For coronal traces, the interslice distance was not always provided and thus was assumed to be in 2 mm increments. Once all of the images in a sequence were traced the result was a set of vertices connected to one-another on their respective planes, but not between adjoining planes (Figure 3.3).

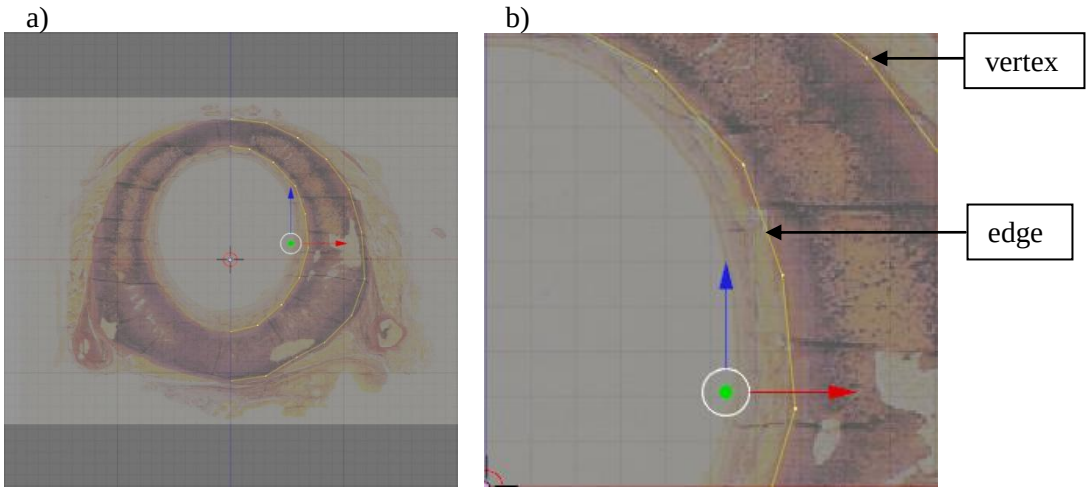


Figure 3.2: Vertex tracing of the cricoid cartilage, illustrated in (a), involved the selection the current render target's (in this case the cricoid cartilage's) hemi-structure in the histological profile and the delineation of its contour with vertices linked by *edges* in the *x*-*z* plane (the red arrow shows the *x*-axis and the blue arrow shows the *z*-axis). The edges constitute one edge of a future triangular polygon formed during the polygon face construction phase. An effort was made to keep the points equidistant while also capturing relevant structural details. The result of this compromise is an approximation to the original structure that expresses the overall form but omits structural peculiarities. In (b) an enlarged view of the vertex trace is presented. Vertices are the yellow points connected by yellow lines, which represent future polygon edges. In this image the degree of deviation from the original structure can be observed.

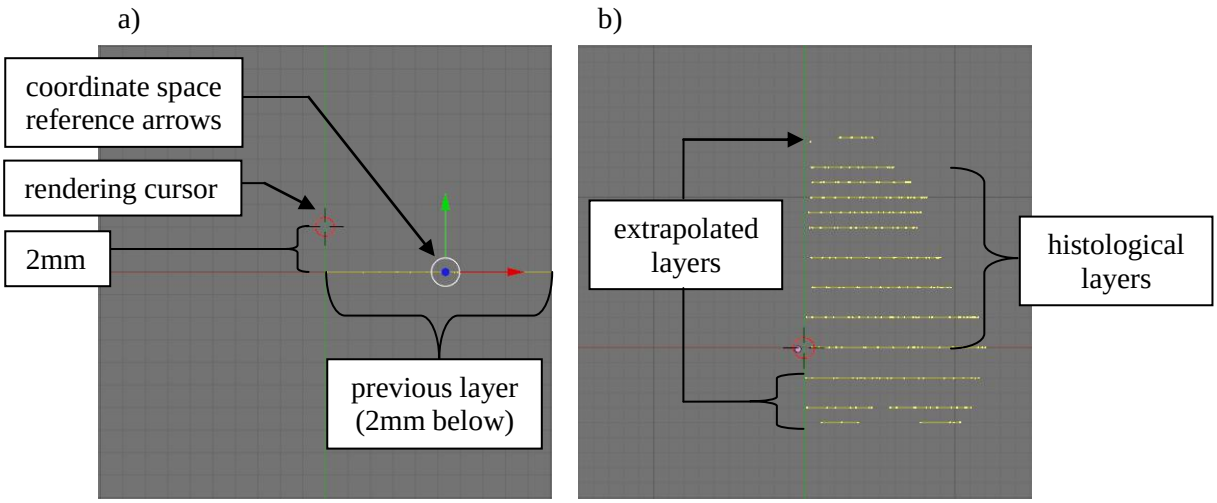


Figure 3.3: Images illustrating how vertex layer placement is carried out. Vertex tracing of the next layer of the cricoid cartilage in the histological sequence, illustrated in (a), begins by positioning the rendering cursor at the height of the next layer, which is at 10mm below the glottis, and 2mm above the previous section. Note that the coordinate space reference arrows now show that the *y*-axis (green arrow) is vertical and the *z*-axis points into the screen (blue arrow). The previous section appears as merely a line now because the *x*-*z* plane is perpendicular to the screen. Also in this view, the background image is no longer displayed; however, it will reappear when the viewing angle is returned to the *x*-*z* plane to begin tracing at the next layer. Image (b) shows the complete collection of layers traced in the histological sections. In the case of the cricoid cartilage these run from 12 mm below the glottis to 2 mm above the glottis. Near the glottal region Hirano and Sato (1993) decrease

the interslice distance of histological sections to 1mm. This is reflected in the increased number of layers near the top of the traced cricoid cartilage (which now appears to be a posterior view of the cartilage). At the bottom and top of the hemi-cartilage there are a total of three layers added that are not part of the histological data set. These are the extrapolated layers; they are used to 'cap-off' the histological sections so that the resulting mesh does not appear completely flattened at the superior and inferior margins of the cartilage.

Before polygon face creation could proceed, it was common for additional layers to be extrapolated to ensure that the final cartilage received a rounded and organic appearance, rather than a flattened appearance at the margins of the structure. This process was required because the histological data set is merely a sampling of a continuously varying structure and many structures of the larynx are incompletely represented, particularly at their extreme margins (most frequently the inferior and superior margins). To begin the process of extrapolation, a layer height was assumed. Typically, the layer was placed at a distance of 1 mm from the existing adjacent layer. For the new layer, vertex placement was most frequently determined by the vertices of the adjacent layer. The objective of this type of extrapolation was to place the extrapolated vertex equidistant in the tracing plane from all nearest neighbour vertices of the adjacent layer. The result was a line of vertices that appears to run down the middle of the adjacent layer's vertex set. Finally, the height of each vertex was adjusted from a profile view of the cartilage to round the appearance of the cartilage in a natural looking manner. For special cases (Figure 3.4), vertex placement in the extrapolated layer was determined by the presence of tubercles or processes that were not present in the histological images. In this case, the dimensions of the protuberance were determined by Eckel et al. (1994), if measurements are available. If not, then the structure was subjectively created, using photographic images and artistic representations of the actual structures (e.g. Zemlin 1997). Unavoidably, this step involved a great degree of subjective judgment to proceed.

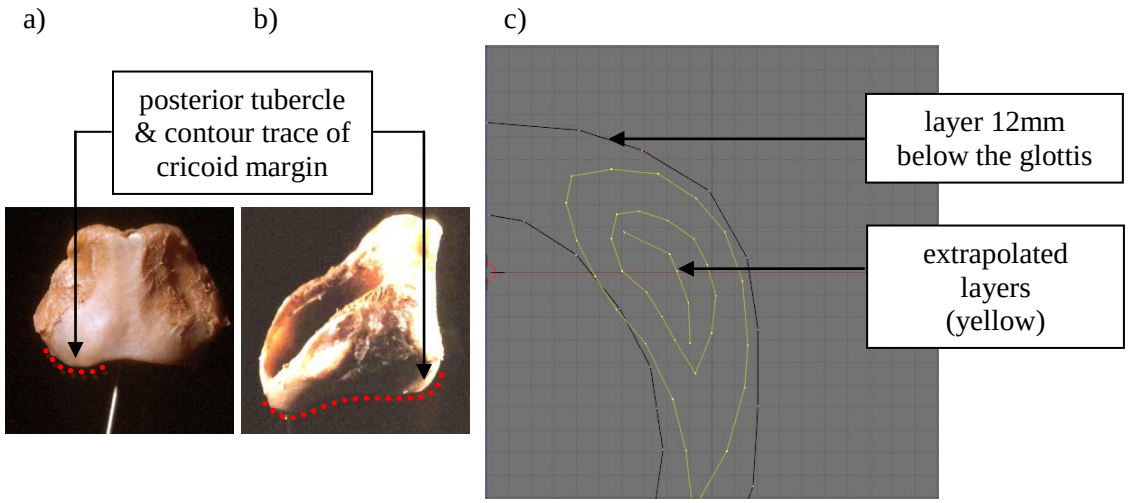


Figure 3.4: Images illustrating layer extrapolation for the posterior and inferior tubercle of the cricoid cartilage. Two images of the cricoid cartilage are visible with the inferior tubercle delineated with a red dotted line in (a: posterior view) and (b: lateral view). These images are used to inform the choices in vertex placement in the extrapolated layers, which are visible as yellow outlines in (c). Three layers are used in this instance to represent the posterior tubercle/foot on the lamina of the cricoid cartilage. Their horizontal placement is at 14 mm, 16 mm, and 17 mm below the glottis. The use of these values is determined by once again approximating measurements from Eckel et al., which specifies that the height of the cricoid from its lower border to the arytenoid articulatory facets is 19.8mm on average (1994: 36, measurement #68). Thus, counting the height of all the layers, the cricoid mesh height is 19mm. This is within the range reported in Eckel et al.'s measurements (between 18mm to 22mm in height).

The next step was to link together nearest vertex neighbours into triangular polygons (faces), which gives shape to the mesh. Finally, once all of the external faces were created, the overall mesh was duplicated, mirrored and corresponding seam vertices were merged together to create the complete mesh of the laryngeal structure (Figure 3.5). The mesh data could then be exported from Blender into a file format that is specific to the DirectX[55] architecture (a *.x* file) for use in the laryngeal model.

[55] See Section 3.3 for more on the software used to develop an engine to render the 3D model.

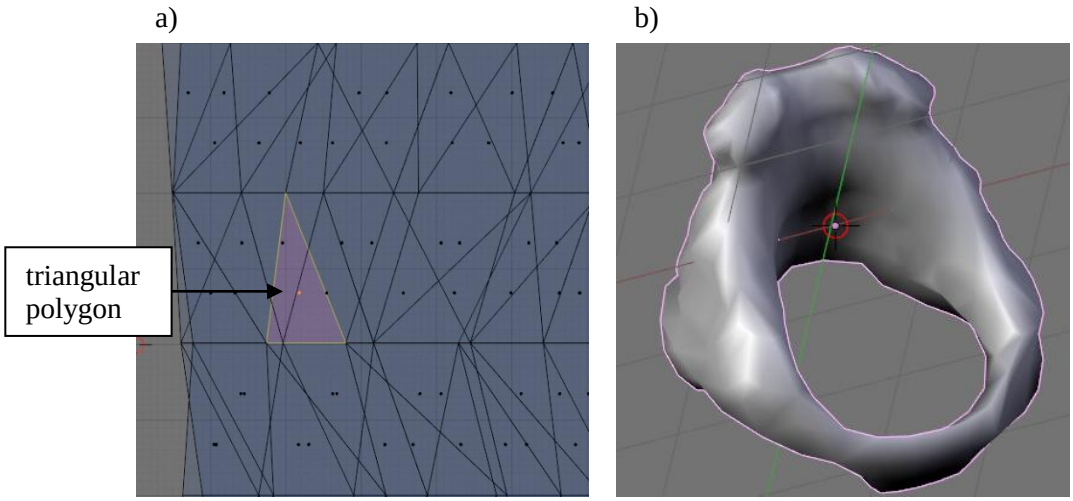


Figure 3.5: Final steps in mesh construction. Image (a) provides a close-up of the mesh structure that was obtained by linking together nearest neighbour vertices from adjacent layers to form a representation of a solid structure. The surface normals of the triangular polygons were set to point outside of the mesh to define an internal and external surface of the structure[56]. In (b) the complete mesh is visible. To arrive at this final state, the hemi-structure (in this case the cricoid cartilage) was duplicated and mirrored. Corresponding vertices lying at the medial edge of each layer were merged together to fuse the structure together.

3.3 Rendering the Model

Developing the model's meshes represents the first step in creating the 3D Laryngeal Constrictor Model. The next phase of development involved the creation of a software engine to describe how the meshes are positioned and orientated in 3D space, as well as move and interact with one another. Visual Basic .Net 2005 (Microsoft Corporation) was used in conjunction with the DirectX 9.0 Software Development Kit (SDK; October 2006, Microsoft Corporation) as the platform for developing the model program. In addition to developing the engine to render the model, a graphical user interface (GUI) was also designed to allow for a high level of interactivity with the model.

Fundamentally, the rendering engine specifies and controls how model components are manipulated, either through user interaction, or via dynamical simulation. All basic model movements are based on implementing observations on actual laryngeal posturing as relatively simple mathematical constraints that are designed to approximate the more complex physiological reality of movements in the larynx and pharynx. The primary motivation behind this approach is to relieve the computer processor from the computationally demanding task of maintaining a complete simulation of laryngeal-pharyngeal muscular biomechanics and contact

[56] In the (a) image Blender has added a square point for each of the triangular polygons to visually locate the Barycentric coordinate of the polygons (essentially, the polygon's center point).

mechanics[57]. Additionally, the simplified mathematical approach is easily implemented using the *DirectX 9.0* libraries without sacrificing the naturalness of how the model moves. Finally, explicit and detailed documentation of the geometrical aspects of movements in the laryngeal vocal tract is rare and incomplete, making an approximation the only option.

Most of the movement constraints concern the relationship between cartilages linked by a joint and moved by shared muscles. The most complicated of all is the motion of the arytenoid cartilages about the cricoarytenoid joints. Another important consideration is the impact of the cricothyroid muscles on the relative positioning of the thyroid and cricoid cartilage to one another. This section reviews the simple mathematical principles and visual illustrations that are used to constrain the model in an attempt to represent real physiological constraints.

3.3.1 Vocal fold posturing

The positioning of the arytenoids is considerably influenced by the shape of the arytenoid articulatory facets on the cricoid cartilage and the corresponding concave articulating surface on the inferior aspect of the arytenoids themselves. The course of this articulatory process runs along a dorso-medio-cranial and ventro-latero-caudal path, which constitutes the rotational axis of the arytenoid cartilage (Sonesson 1959: see Figure 3.6). Sonesson reports that the orientation of this axis is at 24 ± 1° from the sagittal plane and has a declivity of 35 ± 2° from the axial plane (1959: 297). Zemlin (1997: 11) provides figures that are generally wider, holding that the sagittal angle is as large as 30° and the axial plane angle as large as 45°. The rotational axis is illustrated in Figure 3.6. Leden & Moore (1961) provides a brief overview of the mechanics of the joint which attributes arytenoid movement possibilities to the shape of this joint interface. They decompose the motion into three components: rotation around the long axis of the ellipsoidal articulatory facet, translational movement along the long axis of the articulatory facet, and limited rotation about a secondary axis that runs from the posterior cricoarytenoid joint through the vocal process (Leden & Moor 1961: 547). The secondary axis is not widely recognized[58], as is the case with a putative vertical axis through the arytenoid body. Sonesson (1959: 300-302)

[57] This involves the deformation of solid bodies (or semi-solids) when they contact each other at one or more points. In modeling terms, it not only involves incredibly complicated mathematical and physical simulations derived from continuum mechanics, but also maintaining a collision detection schema that is fine grained enough to deal with the complexities of laryngeal structures. Work on this type of modeling is exemplified by Titze (2006).

[58] This observation is based on its utter absence from speech anatomy text books such as Zemlin (1997) and Palmer (1993).

describes the movement effect of individual contraction of the intrinsic laryngeal muscles attached to the arytenoids. Contraction of the lateral arytenoid muscle (LCA) results in infero-medial rotation of the vocal process and antero-lateral translation of the arytenoid body along the articulatory facet. The posterior cricoarytenoid (PCA) acts primarily to abduct the vocal folds along the rotational axis, without the addition of translational movement. Interarytenoid (IA) contraction results in postero-medial translation. Finally, the thyroarytenoids (TA) cause an antero-lateral translation with medial rotation[59]. Maximal translational effect along the cricoarytenoid joint surface amounts to about 2 mm (Sonesson 1959: 302).

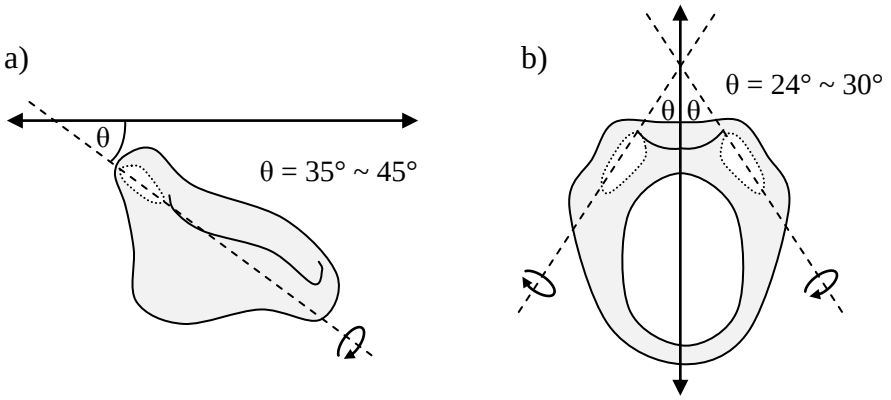


Figure 3.6: Angles designating the rotational axis of the arytenoid cartilage based upon the longitudinal axis of the articulatory facet for the cricoarytenoid joint on the cricoid cartilage. Image (a) illustrates the declivity of the axis, corresponding to about 35° ~ 45°. Image (b) shows the deviation from the sagittal midline of the cricoid cartilage, occuring at 24° ~ 30° (both sides are shown).

While the models presented by Sonesson (1959) and Leden & Moore (1961) are adequate to identify movement patterns for individual contraction of the arytenoid muscle group, they do not describe how arytenoid positioning is impacted by simultaneous contraction of several muscles. Zemlin (1997: 146-148) offers a brief description of some of the compound possibilities. Complete adduction of the vocal folds is possible with compound activity of the lateral cricoarytenoids and the interarytenoids. Independent contraction of either muscle cannot completely close the glottis: acting unopposed, the lateral cricoarytenoids force the vocal processes to be brought together, leaving a posterior gap, characteristic of whisper, while the interarytenoids will adduct the arytenoid cartilage bodies together, but leave the vocal processes

[59] Which Sonesson notes : "im Prinzip die gleichen Rotations- und Translationsbewegungen auf den Arytaenoidknorpel wie der M. cricoarytaenoideus lateralis Muskel." (1959: 300). Presumably, however, the medial force of the vocal processes generated by contraction of the LCA muscles is greater than that of the TAs due to the added torque from the cricoarytenoid joint as the LCA draws the muscular process ventrally.

somewhat abducted. If posterior arytenoid contraction accompanies contraction of the adductory muscles then the net result is that the vocal folds are tightly approximated. It is unknown whether the interarytenoids and lateral cricoarytenoids achieve the same adductory state regardless of the sequence in which they contract. As they are both capable of independent activity, it follows that the muscles could contract in sequence. For the purpose of the model, it is assumed that the sequence of contraction of these two muscles results in the same end state. Zemlin (1997: 130) suggests that compound contraction of the thyroarytenoids with other intrinsic muscles will result in increased vocal fold tension; however, he does not say specifically which muscles accomplish this. In this case it is assumed that, while activity of the cricothyroid is most probably the primary muscle generating the tension during simultaneous thyroarytenoid contraction, activity of the other adductory muscles may also increase vocal fold tension during thyroarytenoid contraction. Interactive behavior between the cricothyroids and the posterior cricoarytenoids is claimed to maximize the abductory posture of the arytenoids (Bartlett 1989: 37), which occurs during deep inhalation.

The 3D model requires a constraint system to control the positioning of the arytenoids that is flexible enough to manage user manipulation of muscle contraction extent. The model's default configuration for the vocal folds is at a point where the activity of the intrinsic laryngeal musclulature was deemed to be minimal. This involves a positioning that is part way between complete abduction and complete adduction, where the muscle activity for the posterior cricoarytenoid muscle and the adductory muscles is conceivably at its lowest (based on EMG studies; e.g. Brancatisano, Dodd & Engel 1991: 977). This state, sometimes described as cadaveric, is best exemplified by the vocal fold positioning observed post-mortem or during paralysis, where the folds are less patent than at any phase of the breathing cycle (Bartlett 1989: 40). During this posture, there is roughly a 5mm separation between the vocal processes (Eckel et al. 1994: 35, #13)[60].

Rotation is handled using spherical linear interpolation[61] and setting boundary conditions along which the interpolation can occur. The identification of boundaries was based on diagrams of arytenoid location relative to the cricoid cartilage in Zemlin (1997: 132, 148). During maximal

[60] Eckel (1994: 35) notes that these measurements cannot be considered completely reliable due to post-mortem changes in the vocal folds elasticity and positioning. However, it is currently the only suitable measurement that corresponds to the resting state of the vocal folds.

[61] This process, commonly referred to as SLERP, interpolates movement between two rotations (expressed in quaternion form) using a parameter to specify the extent of movement (0 to 1, i.e. 0% to 100%).

posterior cricoarytenoid contraction (characterized by forced inhalation), the maximal excursion of the vocal processes from the midline roughly corresponds to the interior border of the lateral leg of the cricoid arch (on either side) as seen from above. Adduction, on the other hand, should result in the vocal processes contacting; this is the case for contraction of the lateral cricoarytenoids with and without concomitant interarytenoid activity. These three primary arytenoid configurations are illustrated in Figure 3.7.

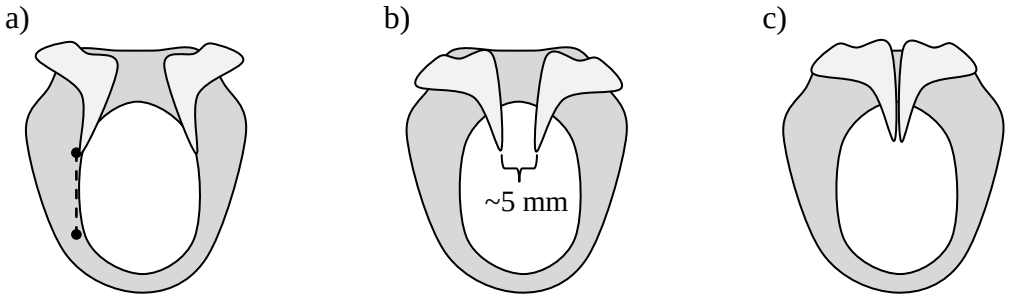


Figure 3.7: Illustration of the three primary arytenoid configurations used to define the lateral (a) and medial (c) extents of rotation along the cricoarytenoid facet for spherical linear interpolation of arytenoid position. The dotted line in (a) indicates the guide used to set the lateral extent of rotation, which involves positioning the arytenoid so that the vocal process roughly lines up with the interior border of the cricoid cartilage from a superior view. At approximately the mid-way point along this interpolated path the arytenoid cartilages are defined as being in cadaveric or resting position (b). The resting state is based loosely on the 5 mm inter-vocal process distance reported in Eckel et al. (1994: 35, #13) for female larynges, which is proposed to roughly correspond to the point at which the muscular activity of the ab- and adductory muscles is at a minimum.

The medial limit on rotation of the arytenoids is determined by when they contact one another. A constraint system is employed based on collision detection between the vocal processes and the arytenoid bodies. Contact between the vocal processes will result in a change of arytenoid rotation that will drive the arytenoid bodies together with continued contraction of the interarytenoid muscles. The effect of continued lateral arytenoid muscle contraction would be to increase the medial compression of the vocal folds until a state of maximal adduction and medial compression is reached, presumably at the furthest extent of the two muscle groups' contraction[62]. To create this constraint the vocal processes and the arytenoid bodies are abstracted as spheres that approximate their shape. The position of the spheres is transformed in the same way as the cartilage mesh itself and the program checks to see if the spheres overlap. Upon overlap of the spheres, the transformations applied change to either allow medial rotation

[62] In the current phase of the model, however, the effect of increasing medial compression in the fully adducted state is not modeled.

and translation of the arytenoid bodies, while the vocal processes are stationary, or prevent further movement from occurring when the arytenoid bodies finally collide. Figure 3.8 illustrates the use of spherical volumes to detect collisions between parts of the arytenoid cartilages.

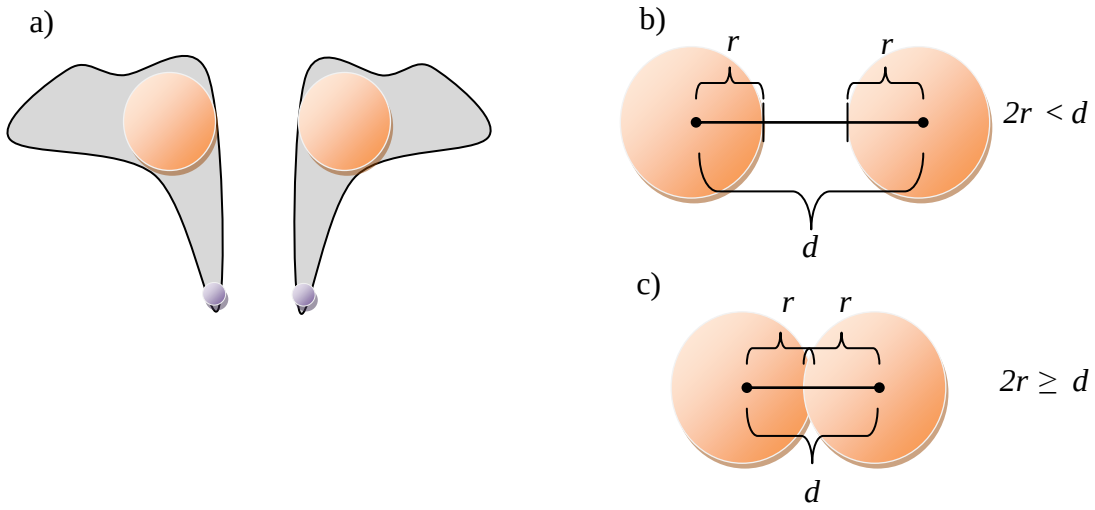


Figure 3.8: Illustration of how spherical bounding volumes are used to detect collisions between the arytenoid cartilages in the model. Image (a) shows the approximate locations of the bounding spheres positioned over the arytenoid cartilage; the orange spheres are used to identify the overall center of mass of the arytenoid cartilage and the medial border of the cartilage subject to collision with the corresponding arytenoid body; the purple spheres represent the smaller bounding volumes used for detecting collisions between the vocal processes of the arytenoid cartilages. No collision occurs when the distance between the centers of the bounding spheres is greater than the sum of the spheres' radii, as in (b). When the sum of the spheres' radii is greater than or equal to the distance a collision is deemed to occur, as in (c).

Finally, the posturing of the vocal folds under isolated maximal contraction for each of the intrinsic muscles controlling the arytenoids is presented in Figure 3.9. The positions are determined based on descriptions and diagrams of the muscles' functions in Zemlin (1997: 126-136), Palmer (1993: 106-111), and synthetic results using biomechanical models developed by Hunter, Titze, & Alipour (2004), and Titze (2006: 133). Configurations presented in Titze (2006) are particularly useful as they depict vocal fold configurations with the corresponding level of muscle activity. The values used in their model are assumed for the 3D LCM.

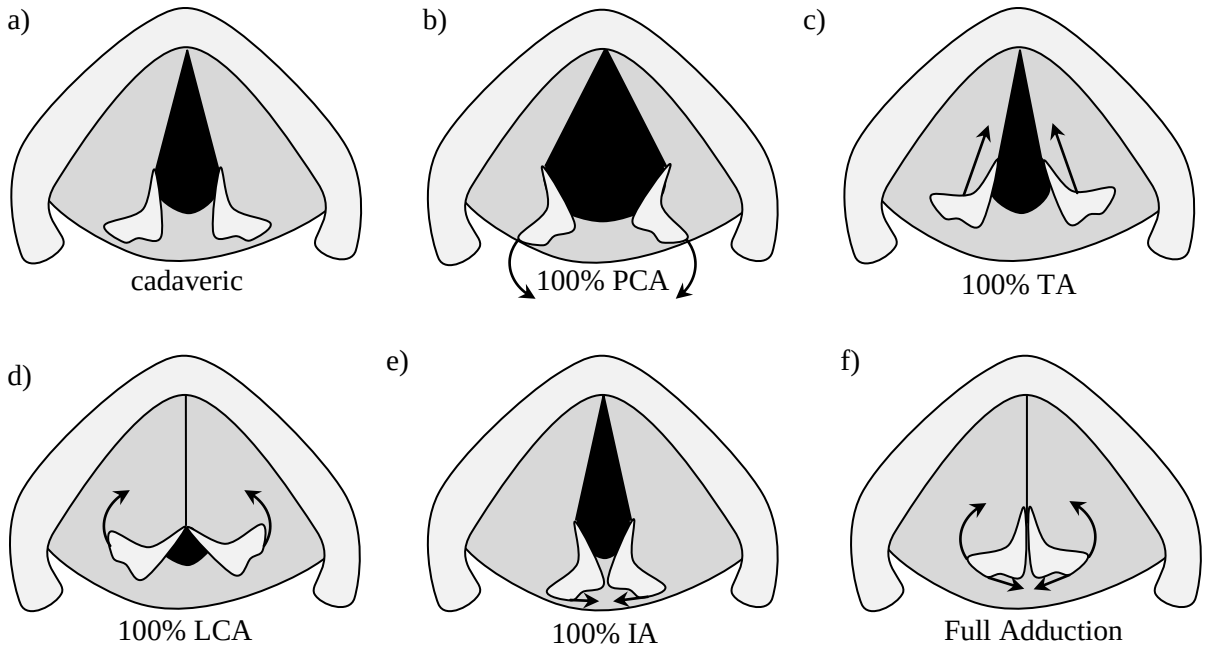


Figure *3.9*: Illustration of vocal fold configurations (top-down view of the larynx without the laryngeal collar) based on isolated, maximal contraction of intrinsic laryngeal muscles, adapted from Titze (2007: 133): (a) cadaveric, (b) 100% posterior cricoarytenoid, (c) 100% thyroarytenoid, (d) 100% lateral cricoarytenoid, (e) 100% interarytenoid, and (f) full adduction (representing 50% LCA and 90% IA, with 20% TA activity).

To achieve increased longitudinal tension of the vocal folds, the cricothyroid muscle must be activated. As discussed in Section 2.2.2, it is assumed that both the thyroid and cricoid cartilages rotate under contraction of the cricothyroid muscle. The angle of movement is defined as the angle between the axial plane running through the cricothyroid joint and the base of the thyroid angle for the thyroid cartilage, and the apex of cricoid arch for the cricoid cartilage. Effectively this angle defines the anterior gap that exists between the cartilages, which closes as the cricothyroid contracts. In Titze's (2006: 127) investigation into cricothyroid physiology, translation values for the thyroid cartilage along the cricothyroid joint were obtained, indicating a range of 2 mm for the ten excised larynges used in their investigation and an angular range of ±9°. In an MRI study, Honda et al. (2004: 1055) provide evidence that the translation effected by the cricothyroid joint ranges up to 1.25 mm (in a live subject) between low and high fundamental frequency. They note that the angular change between the two cartilages during the movement is only 5° in the live subject. Figure 3.10 illustrates the cricoid and thyroid transformations under cricothyroid contraction applied in the model.

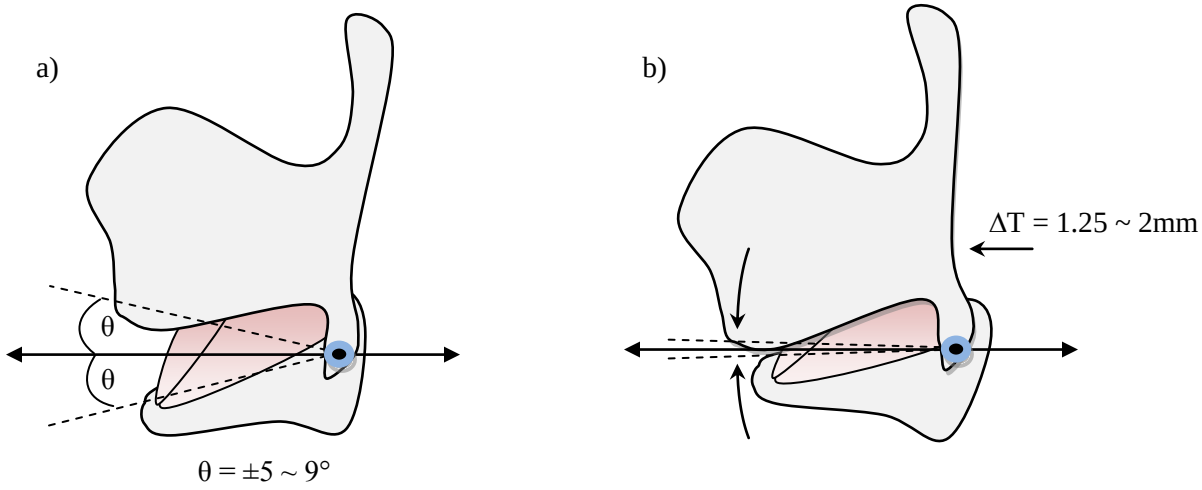


Figure *3.10*: Illustration of the cricothyroid muscle's impact on rotation and translation of the cricoid and thyroid cartilages. Image (a) shows the cricoid and thyroid cartilages in a neutral state. Image (b) illustrates the effect of cricothyroid contraction, which produces rotation towards the axial plane for both cartilages (along the cricothyroid joint rotational axis) and anterior translation of the thyroid cartilage (ΔT) amounting to approximately 2 mm.

3.3.2 Ventricular fold posturing

With regards to the ventricular folds, the primary functional parameter of concern for the 3D model is representing their adduction. It is assumed, based on laryngoscopic videos and evidence from anatomical literature (see Sections 2.1.4 and 2.2.3) that the ventricular folds are capable of complete contact at the sagittal midline. Unfortunately, there is insufficient research in regards to the level of muscle activities required to fully adduct the ventricular folds. There is also a lack of agreement concerning exactly which muscles are responsible for adduction. As discussed in Section 2.2.5, the thyroepiglottic, aryepiglottic, and thyroarytenoid muscles are assumed to generate medial movement of the ventricular folds, shown in Figure 3.11. The ventricularis is assumed to play a role in bunching together the ventricular folds in the sagittal direction and drawing the folds downwards, generating ventricular incursion, a key feature of the valve model of the laryngeal vocal tract (see Section 2.1.6). Explicit values for muscle activities are estimated using normal laryngoscopic videos, which are examined in Section 4.2.

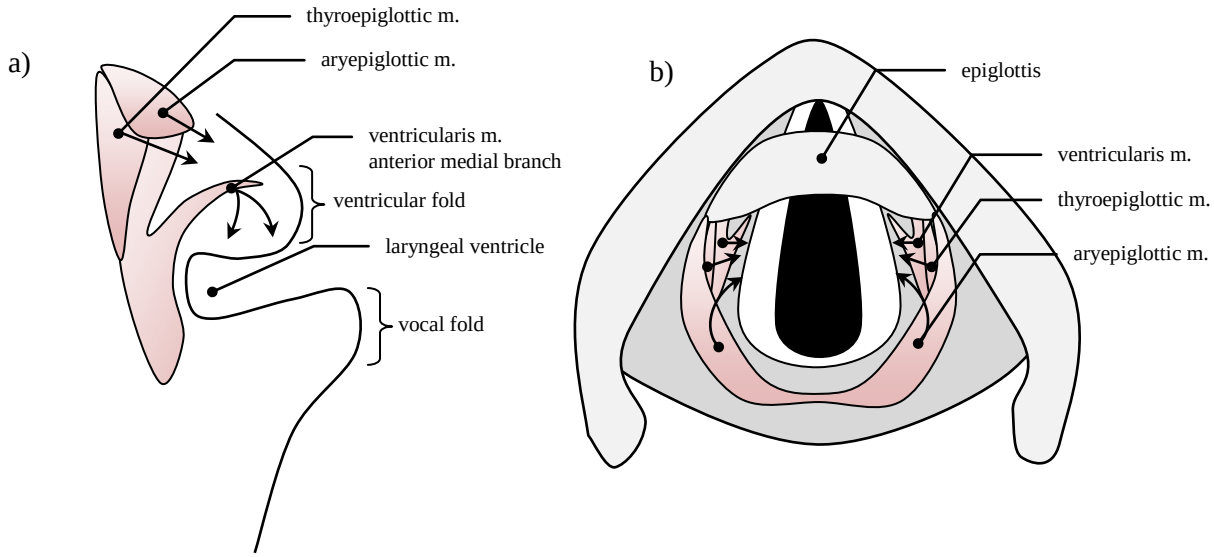


Figure *3.11*: Illustration of the muscular forces assumed to bring about the adduction of the ventricular folds. Image (a) is a coronal section through the left fold (with depth compressed) to show the relationship between the muscles and the force vectors operating on the ventricular folds. In accordance with Reidenbach's (1998) observations and others, the ventricular folds are pushed towards the midline by the action of the ventricularis, thyroepiglottic, aryepiglottic, and thyroarytenoid muscle (not shown). These muscles, in addition to the anterior medial branch of the ventricularis muscle also effect a downwards motion on the ventricular fold. Image (b) is a superior view of the larynx with the muscles "exposed" to allow for visualization of the medially directed force vectors in the axial plane.

3.3.3 Aryepiglottic fold posturing

Under conditions of laryngeal constriction during swallow and pharyngeal articulations, the basic behaviour of the aryepiglottic folds is to approximate the region of the tubercle of the epiglottis, which appears as the base of the epiglottis in laryngoscopic videos. The same three muscles (i.e. thyroepiglottic, aryepiglottic, and thyroarytenoid muscles) responsible for medial movement of the ventricular folds play a role in positioning the aryepiglottic folds, which indicates that the two valves are functionally linked to some extent (see Sections 2.2.3-5). The function of each of these muscles, represented in the model, is assumed to bring about different results in terms of configuration and shape. The thyroepiglottic muscle is interpreted as generating the anterior force to draw the aryepiglottic folds and cuneiform cartilages towards the base of the epiglottis, while the aryepiglottic muscles assist in drawing the epiglottis back, in addition to facilitating the anterior movement of the aryepiglottic folds. The thyroarytenoids are assumed to play a role in translating the arytenoids anteriorally, which facilitates the overall

process of approximating the aryepiglottic folds to the base of the epiglottis. Figure 3.12 provides an illustration of the muscle force vectors involved in aryepiglottic fold positioning.

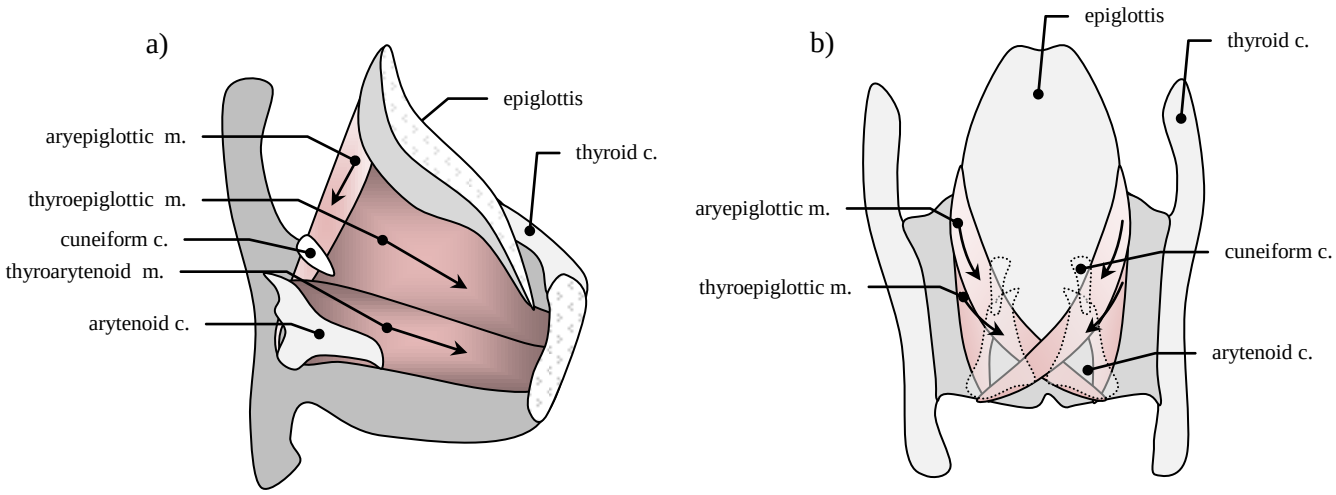


Figure 3.12: Illustration of the muscular forces operating of the aryepiglottic folds to bring about their constriction. Image (a) is a mid-sagittal section through the larynx, showing the muscles thought to achieve aryepiglottic constriction. The force vectors are depicted as arrows associated with the identified respective muscles. Image (b) is a posterior view of the larynx with the thyroarytenoid removed to show the medial direction of the force exerted by the aryepiglottic and thyroepiglottic muscles. The cuneiform and arytenoid cartilages have been made transparent in this view.

Epiglottis rotation also needs to be represented in the model. For normal individuals (i.e. those without dysphagia), the range for the rotation of the epiglottis during swallow is 63° (Paik et al. 2008: 331). This value is implemented on the assumption that epiglottal movement is the same in swallow as in constricted articulations; Figure 3.13 illustrates epiglottal rotation in the model.

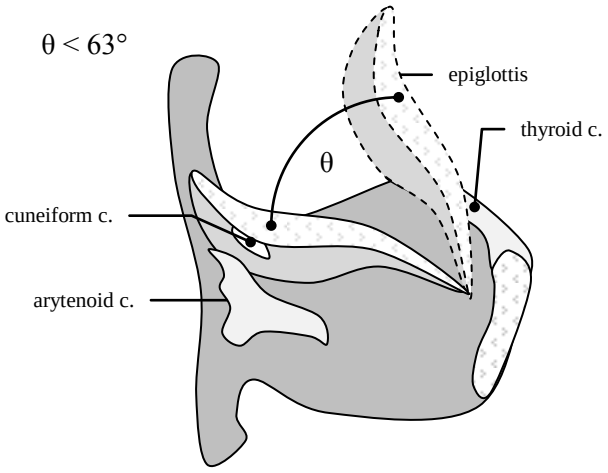


Figure 3.13: The range of rotation of the epiglottis during swallow reaches up to 63° on average. This value is used to constrain the epiglottis movement in the model, such that it the rotation never exceeds 63°.

3.4 Summary of the basic modelling methodology

This chapter has expatiated on the methodology used to construct the 3D model of the larynx and its basic movement properties. The histological images from Hirano and Sato (1993) were identified as a central data source for the construction of 3D meshes, which are used as the basis for the representations of laryngeal structures. Structures preserved in the histological sections were segmented and rendered using a process described as vertex tracing.

Much of the more sophisticated, articulatory movements involve the concerted action of a number of intrinsic laryngeal muscles, particularly with regard to the arytenoids. Thus, the latter portion of this chapter discussed the dynamic structures of the model that are under user control. The movement properties of the primary laryngeal structures (vocal, ventricular, and aryepiglottic folds) discussed in the preceding chapter in Section 2.2 were illustrated. Basic transformations (translation and rotation) are applied to structures with simple movement, such as the epiglottis, cricoid, and thyroid cartilages. The arytenoids, however, are controlled using spherical linear interpolation to approximate their movement along the convex surface of the cricoarytenoid facets.

Chapter 4 focuses on the empirical data collected from laryngoscopy, which is integral to characterizing the more complicated aspects of laryngeal movement. The two main topics concern laryngeal structure movements during articulation and phonation, and aryepiglottic trilling.

Chapter Four

Deriving Synthetic Articulations from Laryngoscopy

4.1 Orientation to Synthesizing Articulation

Model articulations are based upon movements observed in laryngoscopic videos of actual laryngeal articulations. This process involves deducing the degree of muscle contraction and synergies involved in achieving the target states of the articulations. To this end, various types of laryngoscopic recordings made at the University of Victoria are analyzed, including high speed, stroboscopic, flexible nasoendoscopic, and rigid endoscopic laryngoscopy[63].

Articulations from both a trained phonetician and those from native speakers of a wide variety of languages have been considered. The data from the trained phonetician are central to the reconstruction process due to their consistency and the completeness of the articulatory set, which includes both segments and illustration of phonation types/voice qualities. These data are careful phonetic performances and considered to be representative of the "cardinal 'benchmarks'" (Esling, Fraser, Harris 2005: 390) for articulatory categories found in the laryngeal vocal tract. The samples from native speakers are used to assess the variation of articulation patterns that occur, but are not directly modeled given the variation in laryngeal morphology and articulation. Due to the vast difference in quality and frame rate between the normal laryngoscopic footage and the high speed laryngoscopy, different analysis techniques were used to obtain movement data. In both cases, however, the MATLAB (Mathworks, Inc.: R2007a) platform and its *image processing toolbox* were used to conduct the analysis. Section 4.2, provides a discussion of normal speed laryngoscopy, while Section 4.3 presents two different investigations into high-speed laryngoscopic video of voiced aryepiglottic trilling.

4.2 Normal Laryngoscopic Video Analysis

Normal speed laryngoscopy is used to assess laryngeal-pharyngeal states for the set of articulations represented by the model. These data form the foundation for designing movement constraints for the 3D model's structures. The goal is to provide a general characterization of

[63]All laryngoscopic research was conducted by Dr. John Esling of the University of Victoria and his colleagues, with the exception of the high-speed videos, which were recorded at Université Paris III, UMR 7018, and the ORL section of the Hôpital Européen Georges Pompidou with the aid of Dr. Lise Crevier-Buchman.

each articulation that will serve to identify relative positioning and potentially sequencing information as well. Individual variations in the morphological characteristics of laryngeal anatomy render absolute measurements relatively meaningless. Rather, an approach similar to Esling & Clayards (1999) is adopted that involves expressing differences in measurements between articulatory states as ratios. This relativized presentation of the measurements allows for greater flexibility in comparing articulations to one another and between speakers, serving to underpin the model with more idealized, rather than specific, anatomical and articulatory data.

Each articulation studied using normal laryngoscopic videos represents some degree of activity along the continuum of laryngeal constriction. There are a number of variables that intersect with regards to the degree of constriction employed. Notably the aryepiglottic sphincter may be freely engaged to varying extent while the height of the larynx is manipulated (Esling 1999a). Despite this flexibility, however, there are two distinct endpoints on the continuum of constriction. When the larynx is maximally constricted the larynx raises, the aryepiglottic sphincter is fully engaged and the epiglottis and tongue root are fully retracted. Additionally the pharyngeal constrictor muscles may also be engaged to further narrow the pharynx and enhance the constriction. On the other hand, with the larynx lowered and the posterior cricoarytenoid muscles fully contracting, the larynx and pharynx could be considered to be in a state of maximal opening.

4.2.1 Normal laryngoscopic video Analysis: methodology & materials

The basis for characterization of laryngeal and pharyngeal states is by means of measurements of anatomical landmark relationships. The landmarks selected are intended to index the state of the epilaryngeal constrictor and the glottal aperture. The following landmarks are used to characterize the epilaryngeal constrictor: the distance between the apices of the corniculate tubercles, the distance between apices of the cuneiform tubercles and their distance to the epiglottic tubercle, the distance between the interarytenoid notch and the epiglottic tubercle, and the areas circumscribed by the ventricular folds and by the aryepiglottic folds and epiglottic tubercle. While laryngeal height frequently facilitates laryngeal constriction, there has been no attempt to rigorously quantify the vertical positioning of the larynx; rather, changes in laryngeal height are noted descriptively. The glottal aperture is characterized by the distance between the

protuberance formed by the vocal processes, an approximation of glottal area[64], and corniculate tubercle distance. In Figure 4.1, the anatomical landmarks used for measurement are identified. Esling & Clayards (1999) present measurements of the external angle formed between the anterior section of the aryepiglottic fold (which lies between the margin of the epiglottis and the cuneiform tubercle) and the posterior section (which lies between the cuneiform tubercle and the corniculate tubercle). Rather than repeat the measurement of these angles, the data they report is considered for integration into the model. Overall the methodology applied is an expansion of Moisik & Esling's (2007) approach to characterizing articulatory states using landmark difference ratios. New to this approach is the inclusion of area measurements, an explicit specification of the position of measurement for glottal width (i.e. distance between the vocal processes), and the addition of measurements of cuneiform tubercle-epiglottic tubercle measurements. The results obtained here parallel those reported in Moisik & Esling's report, and consequently only the results from the present analysis are considered.

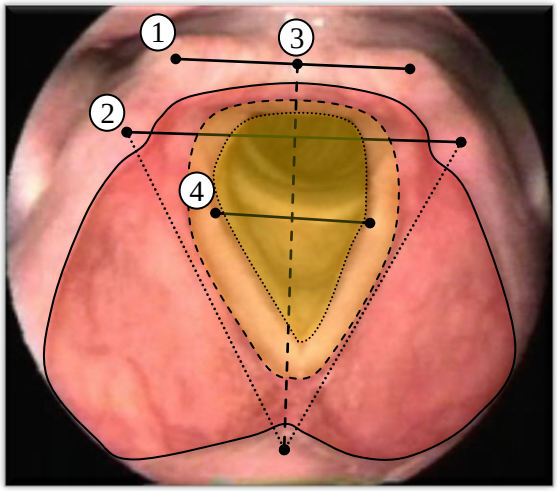


Figure 4.1: A frame from a laryngoscopic video showing the larynx in a maximally open configuration (deep inspiration) with the anatomical landmarks identified. Measurement (1) is the distance between the apices of the corniculate tubercles; measurement (2) is the distance between apices of the cuneiform tubercles (solid line) and their distance to the epiglottic tubercle (fine dotted lines); measurement (3) is the distance between the interarytenoid notch and the epiglottic tubercle (coarse dotted line). The red region (solid outline) denotes the area circumscribed by the aryepiglottic folds and epiglottic tubercle. Note that this area subsumes the other areas identified, the ventricular (orange region with coarse dotted outline) and glottal areas (yellow region with fine dotted outline).

Laryngoscopic videos have been retrieved from the library of footage of laryngeal and pharyngeal articulations collected at the University of Victoria, primarily under the directorship

[64] An approximation is used because vocal fold vibrations are blurred in normal speed laryngoscopic videos.

of Dr. John Esling. Many of these videos actually feature Dr. Esling, a trained phonetician, performing the articulations himself. There also is a vast number, however, of native language speakers from a generous cross-linguistic selection of languages that employ some form of pharyngeal contrast in their phonological systems. The camera positioning is, inevitably, variable from video to video, in addition to individual variations in the morphology of laryngeal and pharyngeal anatomical structures. Thus, the primary analysis is carried out on footage of articulations performed by Dr. Esling. The videos from native language speakers are used for comparison and corroboration that the articulations of the phonetician parallel naturally occurring ones.

While there are variations in the methodology employed from session to session over the years that this research has been conducted, usually the work followed a consistent protocol. The specific setup was typically on a software-based Kay laryngoscopy system that included dual halogen (fixed) and xenon (strobe) light sources, a Panasonic KS152 camera, and a Mitsubishi S-VHS BV-2000 video cassette recorder running at 30 frames/s. Using a GV-D1000 NTSC Sony DV Cassette Recorder (and a 59900U JVC VCR for S-VHS), the videos were captured on computer using Adobe Premier (version 6.50; Adobe Systems Inc.) and Vegas Studio (version 6.0b, build 126; Madison Media Software Inc./Sony) on a system running Windows XP. Still frames from each have been selected for analysis.

4.2.2 Normal laryngoscopic video analysis: results

Measurement of the laryngeal landmarks is carried out using ImageJ, open source image analysis software developed by the National Institutes of Health. These values are provided in Table 4.1. No anatomical scale is identified given the indeterminacy of the placement of the endoscopic camera relative to the larynx. Rather, measurements are provided in terms of ratios based on comparison between landmark measurements in a fully abducted, unconstricted configuration, and those of the target articulatory state. The measurements recorded for inspiration are set to 100% because this laryngeal state is considered to be maximally open/unconstricted[65]. In addition to the ratio values, exact measurement in pixels is provided for

[65] With the exception of falsetto for this speaker: the distance between the interarytenoid-notch and the epiglottic tubercle is higher than the value noted for inspiration. This condition is unsurprising given the increased activity of the cricothyroid muscle. In terms of articulatory state, falsetto is not as open as inspiration for the other metrics and thus is not normalized to 100%; rather, inspiration is used for the basis of comparison.

distance measurements and pixels^2 for area measurements. Finally, for those articulatory states that involve voicing, the fundamental frequency of the voice is also reported. A bar graph of the ratio data from the aryepiglottic area measurements is presented in Figure 4.2. This data is chosen as it is highly reflective of the articulatory contrast between constricted and unconstricted laryngeal states, implicating the aryepiglottic folds as a primary factor in their production.

Table 4.1 Measurement ratios of laryngeal anatomical landmarks

State	F0	CoT	CuT	C-ET		IA-ET	VP	AE area	VentF area	VF area	LH
				R	L						
Inspiration	-	100% (87)	100% (126)	100% (127)	100% (130)	100% (143)	100% (62)	100% (18756)	100% (6609)	100% (3431)	N
Modal Voice	146	9% (8)	48% (61)	78% (100)	88% (115)	79% (113)	16% (10)	60% (11397)	43% (2851)	4% (168)	N
Falsetto	415	13% (12)	56% (71)	89% (114)	99% (129)	107% (154)	3% (2)	93% (17556)	58%[66] (3884)	9% (322)	N
Breath / [h]	-	11% (10)	56% (71)	77% (98)	80% (105)	88% 127	33% (21)	64% (12134)	53% (3540)	24% (834)	N
Breathy Voice	78	12% (11)	43% (55)	54% (69)	55% (72)	63% (91)	29% (18)	33% (6343)	38% (2569)	19% (680)	L
Whisper / [ɦ]	-	25% (22)	85% (108)	39% (50)	60% (79)	51% (74)	12% (8)	17% (3329)	20% (1337)	17% (614)	R
Whispery Voice	115	29% (26)	67% 85	44% (57)	51% (67)	54% (78)	9% (6)	20% (3861)	18% (1255)	8% (302)	R
Creaky Voice	45	6% (6)	56% (71)	29% (37)	37% (49)	32% (46)	12% (8)	7% (1318)	3% (248)	1% (25)	N
Harsh Voice (low)	60	10% (9)	35% (45)	7% (10)	31% (41)	22% (32)	?	1% (241)	0% (30*)	0% (3*)	L
Harsh Voice (mid)	111	18% (16)	59% (75)	23% (30)	42% (55)	31% (45)	?	11% (2096)	2% (133)	0% (2)	R
Harsh Voice (high)	418	13% (12)	39% (50)	39% (50)	46% (61)	49% (71)	1% (1)	27% (5152)	3% (247)	1% (30)	N
Glottal Stop [ʔ]	-	14% (13)	50% (64)	33% (43)	57% (75)	48% (69)	0% (0)	24% (4678)	0% (0*)	0% (0*)	L
Epiglottal Stop [ʡ]	-	19% (17)	50% (63)	25% (33)	34% (45)	19% (28)	0% (0)	1% (296)	0% (0*)	0% (0*)	R

Note: F0 = fundamental frequency (Hz); CoT = distance between the corniculate tubercles; CuT = distance between the cuneiform tubercles; C-ET = cuneiform-epiglottic tubercle distance; IA-ET = interarytenoid notch-epiglottic tubercle distance; VP = distance between the vocal processes of the arytenoid cartilages; AE area = area enclosed by the aryepiglottic folds and base of the epiglottis; VentF area = area enclosed by the ventricular folds; VF area = area of the glottis; LH = larynx height (descriptive: N = normal height; L = lowered; R = raised). Values marked with a <*> have been estimated, while values marked by a <?> were deemed too difficult to assess using this method.

[66] In Agarwal, Scherer & Hollien (2003: 111) the ventricular fold distance was 34% greater for modal than falsetto phonation. This deviates from the relationship presented here, where the area for the ventricular aperture is greater for falsetto than modal phonation. Their measurements, however, were taken from a single coronal laminagraph slice through the membranous vocal folds (of several male and female subjects). The distance between the ventricular folds is evidently smaller anteriorally, but the area is enlarged nevertheless because the larynx is more patent during falsetto due to cricothyroid activity.

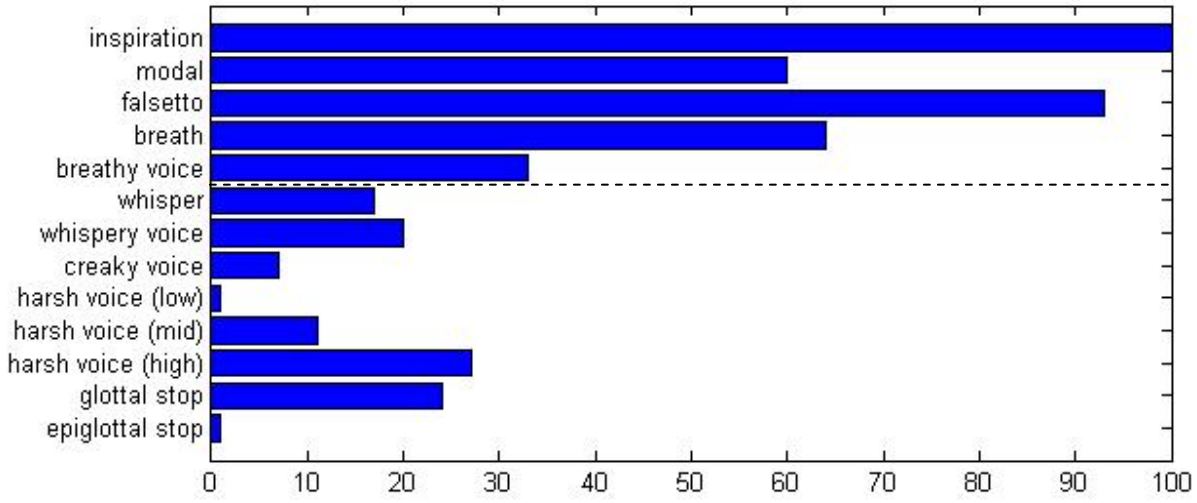


Figure 4.2: Ratio values expressed as percentages for the aryepiglottic area measurement (the AE area column in Figure 4.1)[67]. States above the dashed line are unconstricted, those below are constricted.

Interpretation of the data presented in Table 4.1 must take into account that no correction was made for the effect of laryngeal height on the measurements. Thus, a distance between two landmarks in lowered larynx state will appear shorter than in raised larynx state[68]. Furthermore, the data represents sampling from a single individual, and should be regarded as characteristic rather than precise. Nevertheless, some interesting patterns emerge. The first observation concerns the values obtained for the corniculate tubercle distance; for nearly every articulation examined, apart from (deep) inspiration, the corniculates are tightly approximated. This indicates that there is a high degree of interarytenoid activity throughout the range of unconstricted and constricted articulations. The corniculate distance for whisper and whispery voice (25% and 29%) must be considered in light of the fact that the larynx is raised. The cuneiform tubercle distance indicates laryngeal elevation for sounds that are constricted, particularly for whispery voice. More indicative of aryepiglottic sphinctering are the measurements between the cuneiform and the epiglottic tubercles: the values generally decrease for constricted sounds. As the measurements are taken from the apices a complete seal at this place, as in an (ary)epiglottal stop, does not decrease to zero, especially given that this articulation involves considerable raising of the larynx. For this particular speaker it is evident that the left cuneiform is generally

[67] Note that pharyngeal fricative [ħ] and pharyngeal approximant [ʕ] are equivalent to whisper and harsh voice (low/mid) repsectively.
[68] As the larynx is further away from the camera, meaning the structures appear smaller.

more posterior in positioning than the right[69]. Interarytenoid notch-epiglottic tubercle distance provides a strong indication of aryepiglottic sphinctering as well. The value for falsetto on this measure actually extends beyond the inspiratory value. This antero-posterior stretching induced by the cricothyroid is evident in harsh voice at high pitch. While this is clearly a constricted laryngeal state, harsh voice at high pitch represents an articulatory paradox between the muscles responsible for constriction, likely the thyroarytenoid (Esling et al. 2007), thyroepiglottic, and aryepiglottic muscles and the cricothyroid muscle, which is forcing the laryngeal vestibule into an open configuration. Also evident with this measure is the justification of classifying breathy voice an unconstricted articulation, while whispery voice is constricted. The difference in the position of the interarytenoid notch between these two phonation types might be attributable to the addition of aryepiglottic sphinctering, which would pull forward on the arytenoids, drawing the interarytenoid notch anteriorally. Obtaining measurements for the distance between the vocal processes proved to be considerably challenging for articulation where the supraglottal structures obscured the vocal folds partially or completely from view. Along with the measurement of glottal area, this measure provides an indication of enough differential contraction between the interarytenoid and lateral cricoarytenoid muscles to result in a posterior glottal chink forming, which is central to the noisy phonation of both whispery and breathy voice. Aryepiglottic area strongly indicates the degree of constriction involved in the articulations. Breathy voice and harsh voice at high pitch share similar values here, which is likely a result of the confounding influence of laryngeal height, once again. However, harsh voice at high pitch is exceptional for a constricted state due to cricothyroid influence. Also, this state shares much in common with glottal stop in terms of aryepiglottic area. Visually, the aryepiglottic folds appear taut in both of these articulations. Ventricular area highlights the role of the ventricular folds in forming a glottal stop identified by Edmondson & Esling (2005), an effect that they refer to as *ventricular incursion* (Section 2.1.4).

4.2.3 Normal laryngoscopic video analysis: discussion

The final phase of this analysis involves direct visual assessment of the frames used for measurement of laryngeal landmarks and comparison of these frames with frames taken from

[69] This anatomical fact of the speaker emerges once again when the high-speed laryngoscopic films of aryepiglottic trilling by the same speaker are considered in Section 4.3.

similar articulations occurring in a wide variety of languages. The goal here is to solidify the articulatory characterization of the laryngeal states in terms of a binary classification scheme involving the descriptive opposites: constricted-unconstricted. The languages chosen employ constriction phonologically, either segmentally or by means of superimposing the constriction over tonal or vocalic articulations. Table 4.2 lists a sample of the languages used and details concerning the phonological occurrence of pharyngeal articulations in the language.

Table 4.2 Languages used in the cross-linguistic comparison of laryngeal-pharyngeal (LP) states[70]

Language	Family	Location	L-P Phonology
Cantonese	Sino-Tibetan	Guangdong, China	Glottal reinforcement[71]
Tigrinya	Afro-Asiatic	North Ethiopia	Pharyngeals
Nuu-chah-nulth	Wakashan	Vancouver Island, Canada	Pharyngeals
Nlaka'pamux	Salish	British Columbia, Canada	Pharyngeals
Arabic	Afro-Asiatic	Saudi Arabia	Pharyngeals

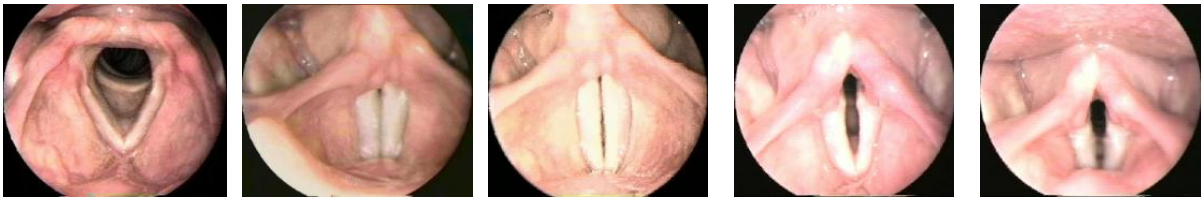


Figure 4.3: A series of laryngoscopic frames showing the unconstricted laryngeal states. From left to right: (deep) inspiration, modal voice, falsetto, breath (exhalation), breathy voice.

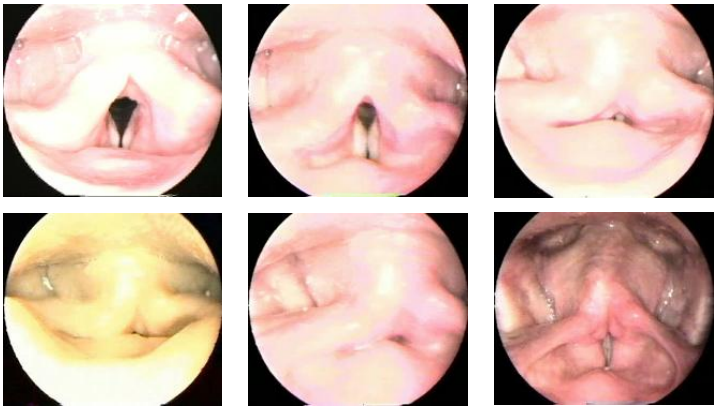


Figure 4.4: A series of laryngoscopic frames showing the constricted laryngeal states. From left to right, top to bottom: whisper, whispery voice, creaky voice, harsh voice (low pitch), harsh voice (mid pitch), harsh voice (high pitch).

[70] Many other languages figured into this cross-linguistic comparison, however, due to space considerations, these images and languages are not discussed here. The reader is encouraged to check the many publications by Esling and his colleagues for excellent documentation of these phenomena across a wide array of languages.

[71] This occurs on coda consonants; e.g. [tit$^{\text{ʔ}}$^{3}] 'order'. See Section 2.1.4 on laryngealization.

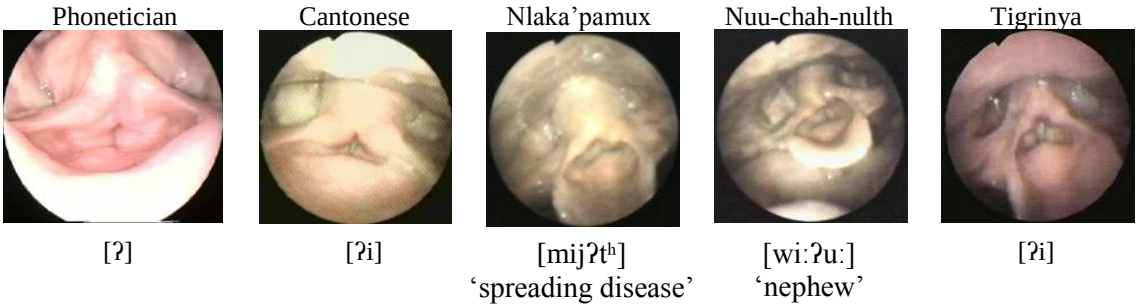


Figure 4.5: A series of laryngoscopic frames showing cross-linguistic instances of glottal stop.

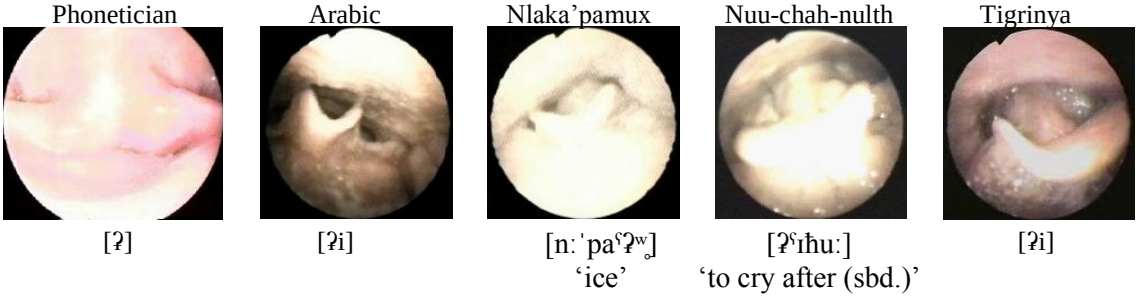


Figure 4.6: A series of laryngoscopic frames showing cross-linguistic instances of epiglottal stop.

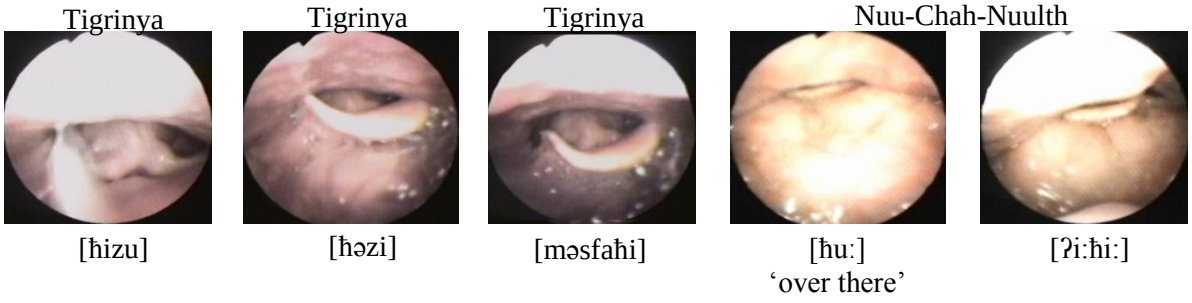


Figure *4.7*: A series of laryngoscopic frames showing instances of voiceless pharyngeal fricative [ħ] during maximal stricture in various contexts. The first frame from Tigrinya indicates that the sound involves tight aryepiglottal stricture. The four frames to the right show that a further epiglotto-pharyngeal constriction is a possible concomitant of the sound.

In Figures 4.3 and 4.4, and the leftmost images of Figures 4.5 and 4.6, the frames used in the measurement analysis are displayed (also see Appendix B)[72]. The unconstricted set (inspiration, modal voice, falsetto, breath, and breathy voice) is clearly characterized by the positioning of the aryepiglottic folds. The internal angle formed between the two folds is persistently about 90° and the aperture formed between the base of the epiglottis and the arytenoids is widely patent. The constricted set (whisper, whispery voice, creaky voice, harsh

[72] These images are obtained from research done by Esling & Harris (2005).

voice (low, mid, and high pitch)), glottal stop, and (ary)epiglottal stop), with a few exceptions, illustrates the effect of closing the aryepiglottic sphincter, to varying degrees. The internal angle of the aryepiglottic folds for this set tends towards 180°[73]. Both harsh voice at high pitch and glottal stop exhibit this tendency to a lesser degree. With harsh voice at high pitch, the reason is likely due to the activity of the cricothyroid, as previously discussed. Glottal stop on the other hand represents a lesser degree of constriction, highly evident when it is compared with epiglottal stop, and hence is characterized by a more angular configuration of the aryepiglottic folds.

The images in Figure 4.5 attest to the occurrence of ventricular incursion in a variety of languages, including those that do not employ pharyngeal contrasts (such as Cantonese). The extent of ventricular incursion is evidently variable, with Nlaka'pamux and Nuu-chah-nulth exhibiting complete contact of the ventricular folds in the images selected, which parallels the glottal stop produced by the phonetician. The Cantonese and Tigrinya examples are cases where there is a slight gap between the ventricular folds at the point of maximal constriction during the glottal stop. The frames of epiglottal stop in Figure 4.6 are less revealing than the phonetician's production, where a complete hermetic seal is plainly visible, formed between the aryepiglottic folds and the base of the epiglottis. In the Arabic, Nlaka'pamux, and Nuu-chah-nulth examples the epiglottis obscures the activity of the aryepiglottic sphincter. In the Tigrinya example it is possible to infer, based on a comparison with the Tigrinya glottal stop in Figure 4.5, that a complete seal is formed. Spectrograms provide evidence that complete closure occurs at the moment of the ostensible epiglottal stop in the laryngoscopic videos. An example from Nlaka'pamux is provided in Figure 4.8.

[73] If one ignores the posterior section of the folds, measuring across the interarytenoid notch.

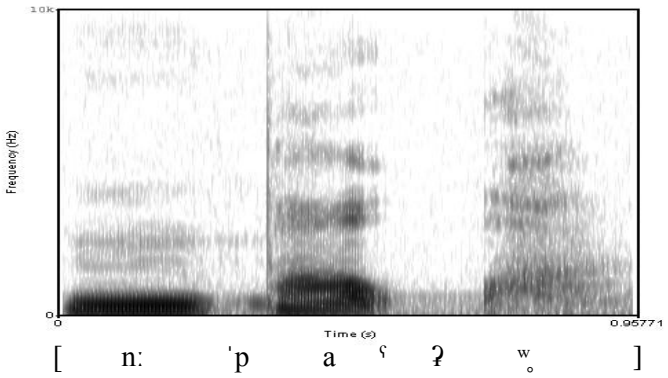


Figure 4.8: Spectrographic evidence of epiglottal stop in an ambiguous frame from laryngoscopic video of the Nlaka'pamux word [nːˈpaˤʔʷ̥] 'ice'. The region of silence corresponding to the position of the film where the frame of epiglottal stop was extracted indicates that the sound is truly a stop.

Perhaps even more revealing is the progressive collapse of the larynx observed during the sequence of frames that show the epiglottal stop articulation in the Tigrinya example, as seen in Figure 4.9. As the aryepiglottic folds approximate the base of the epiglottis the glottal aperture beneath begins to diminish in area, it is not certain whether the ventricular folds are sequenced before aryepiglottic seal formation, if the movement occurs simultaneously with aryepiglottic sphinctering, or if the ventricular folds fully contact at all during the epiglottal stop phase.

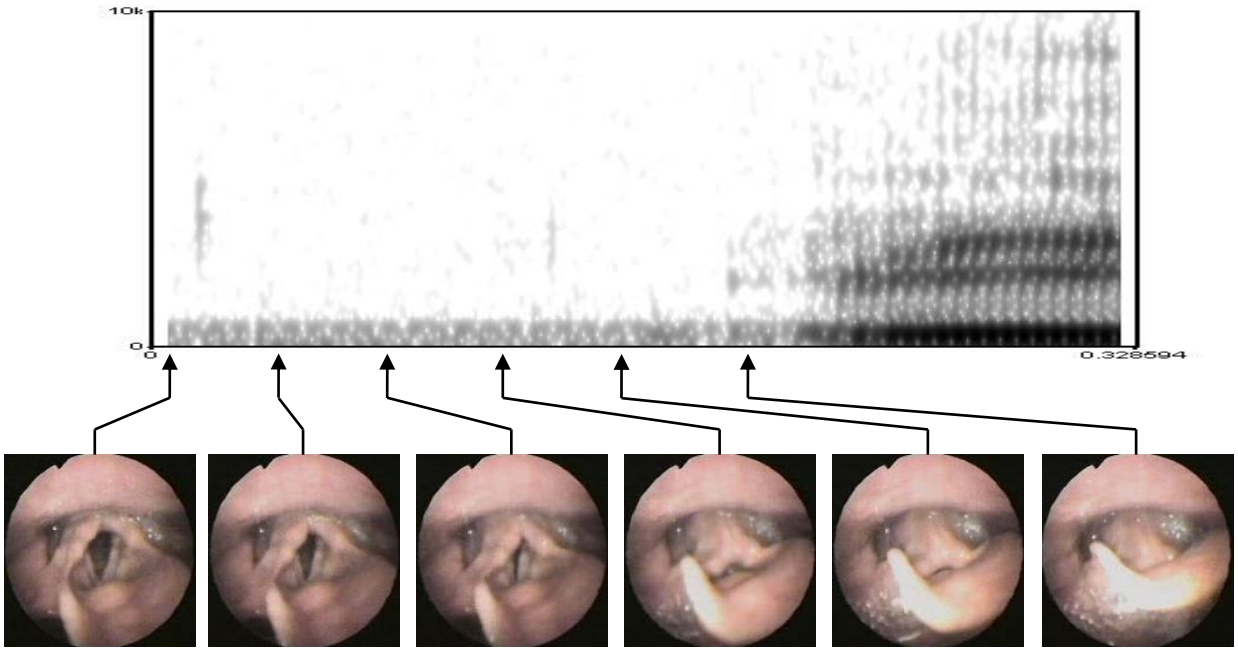


Figure 4.9: Examination of the articulatory sequence of an aryepiglotto-epiglottal stop produced by a Tigrinya speaker. Total duration of the sequence illustrated is 180 ms; the frame rate is of the video is approximately 30 frames/second.

Also worthy of note, but not evaluated in the aperture analysis, is the potential for aryepiglotto-epiglottal sounds, such as the pharyngeal fricative [ħ], to occur with considerable tongue root retraction resulting in epiglotto-pharyngeal constriction as demonstrated in Figure 4.7. This likely enhances the turbulence generated from the extreme postero-anterior aryepiglottal constriction over the glottal airway (see Stevens 1999: 428-429). The extent of contact between the rear pharyngeal wall and the epiglottis is likely to vary across speakers and to be influenced by vocalic context, with retracted vowels facilitating the contact and front vowels opposing it[74]. It is plausible that epiglotto-pharyngeal accentuation of the aryepiglotto-epiglottal constriction is used to enhance the perceptual distinctiveness of the voiceless pharyngeal fricative from the voiceless glottal fricative.

4.2.4 Integrating laryngoscopic data into the model

The final step involved in synthesizing articulations in the laryngeal vocal tract involves developing constraints on model component movements that approximate the behaviour of structures observed moving in the laryngoscopic videos. The most rigorous approach to this problem would involve the study of muscle activation and associated structure movement and deformation using electromyography (EMG). Correlations between these electrical potentials and structural configuration are not often the subject of investigation in EMG studies. As is to be expected, the great majority of EMG research is medical in nature[75], and primarily concerned with fundamental physiological processes such as deglutition and breathing, particularly for pathological or abnormal cases such as vocal fold tumours, paresis, or paralysis (e.g. Heman-ackah & Barr 2007). Linguistically oriented studies, on the other hand, frequently focus on the relationship between the intrinsic muscles of the larynx and frequency, amplitude, and quality of the voice source (e.g. Baker et al. 2001). Therefore, movement can only be inferred through the acoustic data elicited in the session. An example of this is Hong, Kim & Kim (2001) who use

[74] For the Nuu-chah-nuulth speaker, when voiceless pharyngeal fricative is in an intervocalic context with long [i:] vowels, as in the word [ʔi:ħi:] (?), considerable time is taken adjusting from the [i] vowel to the fricative. The duration of the fricative in this word is $\mu = 0.563$ ms ($n = 5$); for comparison, the duration of the fricative during [ʔɪħu:] (?) is $\mu = 0.3275$ ms ($n = 2$), which is also considerably more noisy and at times trilled. This delay may be a result of transitioning between the two articulations, which require considerably different articulatory adjustments to be performed. More research is required to determine the pervasiveness of this coarticulatory effect on pharyngeal production (that is if this is not just a phonological (gemination) phenomenon to begin with).

[75] Furthermore, these studies are frequently carried out on non-human subjects such as cats, dogs, and lambs, due to the invasiveness of the procedure.

EMG to investigate the differential activation levels of the subdivisions of the cricothyroid muscle (pars recta and pars oblique) in controlling preparatory lengthening and tensing of the vocal folds and prosodic effects[76]. Movement is not directly observed in their study; change of cartilage position (in this case anterior rotation and translation of the thyroid cartilage) can only be speculated based on changes in f_0. The activity of the aryepiglottic muscle is rarely measured, due to the considerable range of motion of the aryepiglottic folds, especially during coughing, which makes electrode positioning difficult (see Sakakibara 2004a).

While it initially seems appealing and empirically more robust to establish movement constraints based on EMG research, there simply is not sufficient data to warrant this kind of approach, particularly where it concerns the activity of the aryepiglottic folds. Instead of inferring position based on EMG activity, activation levels (not necessarily associated with EMG potentials) are inferred based on direct visual observation of the larynx, as was documented in the previous section (4.2.3). The ratio data collected in Section 4.2.3 is used as a metric for determining the extent of structural transformations applied in the model. Posturing movements of laryngeal structures (discussed in Section 3.3) offer a continuum upon which articulation operates; limits of movement along the posturing continuum are imposed by determining the parameter values that characterize cardinal laryngeal states, such as maximal abduction. Landmark ratios are then determined using the same methodology used for the laryngoscopic videos. Once the appropriate parameter values are known, they are then hardcoded into articulatory presets, which is the master control that specifies the proportion of activity for all parameters operative for a given articulation. A comparison between laryngoscopic video stills of laryngeal articulations and the corresponding synthesized articulations is discussed in Chapter 6, where the overall model is evaluated.

4.3 High-speed Laryngoscopic Video Analysis

Normal laryngoscopic video has a high enough frame rate at 30 fps to observe gross movements of laryngeal articulations. Vibrating or trilling motion, however, occurs too rapidly for the camera to record anything but a blur of tissue. Trilling of the aryepiglottic folds is one of the central phenomena to be represented in the 3D model. Thus, high-speed laryngoscopy was

[76] It should be noted that the procedure was conducted on patients receiving full or partial lobectomies of the thyroid gland.

obtained to examine this phenomenon further. While laryngoscopic video (of any speed) does not provide depth information[77], it can still be used to characterize how the aryepiglottic folds function. The analysis conducted here is based on the productions of a voiced aryepiglotto-epiglottal trill [ʢ] by a single, trained phonetician. This is a detraction of the present analysis as it limits the confidence with which generalizations can be made. High-speed video, however, is not readily obtained, given the requirement of considerable time, expertise, and special recording equipment. Thus, the observations made from the high-speed video concerning the configuration of the structures of the epilaryngeal tube are generalized to possible configurations found in other individuals. These observations allow for an informed representation of aryepiglottic trilling patterns to be visually synthesized in the 3D model. Inevitably, a more robust understanding of aryepiglottic function can only come from future research with different speakers.

4.3.1 High-speed laryngoscopic video analysis: materials and methods overview

The system used to capture the trilling sequences was a SL Kamera 500 connected to a SpeedCam lite interface (Weinberger), which was set to a frame rate of 500Hz (one frame for every 2 ms). Recording was done in the Paris III Sorbonne-Nouvelle/CNRS-*LPP-*UMR7018 research site at l'Hôpital Européen Georges Pompidou, Paris, with the assistance of Dr. Lise Crevier-Buchman. Two sequences of voiced aryepiglotto-epiglottic trilling were recorded, with a total duration of 1.706 seconds (for a total of 853 frames of trilling motion). Throughout the recording session, an audio signal of the utterances was also obtained using a head mounted AKG C410 microphone, positioned at 4cm from the lips, with an angle of 45° from sagittal plane, sampled at 44100 Hz 16 bit resolution. The audio and video signals were digitized using custom software called Anavox (Vannier-Photelec). A Time-to-Live (TTL) signal was used to ensure synchronization of the audio with the video signal. The video footage was digitized frame by frame in uncompressed format and later stored as uncompressed TIFF[78] files for analysis in MATLAB.

[77] i.e. the camera only presents the structures in linear perspective, which means structures are lost by visual occlusion and depth information can only be inferred by cues from shadow distribution and darkness.

[78] Originally a single multi-image TIFF file was intended to be used. However, due to the style of image directory information storage in TIFF files and the manner in which MATLAB opens image files in general, the use of multi-image TIFF files in MATLAB can result in excessive image loading times. The reason for this is that the image file directory (IFD) of a multi-image TIFF file constitutes a directory chain that must be re-parsed every single time an image is loaded. To alleviate this problem, each frame was stored as an individual image.

To extract meaningful information from the high-speed footage two different methodological approaches are used. The first involves the collection of what is referred to here as aryepiglottic aperture data. The second method involved tracking motion vectors between contiguous frames to identify the directionality of the aryepiglottic fold movement.

Observation of the high-speed videos leads to the immediate conclusion that a single aryepiglottic fold and its associated structures act as a functional unit distinct from that of the other aryepiglottic fold, which taken together with the epiglottis comprises one half of the laryngeal collar. The inference made possible by this basic observation is that there is either loose or no coupling between the paired aryepiglottic folds. This idea receives support from the anatomical fact that the two folds are each connected to their own corniculate and arytenoid cartilage, which are themselves firmly anchored by the intrinsic laryngeal muscles active during a trilling gesture. Thus, for the present analysis, the decision has been made to regard each aryepiglottic fold as a distinct unit. Further observation of the high-speed footage reveals that the overall movement pattern is a tendency towards opening and closing or, in appearance, that of dilation. This observation motivates referring to functional activity of an aryepiglottic unit as that of a change in aperture. Anatomically, the aperture consists of the aryepiglottic fold and its embedded cuneiform cartilage and the corniculate (arytenoid) tubercle posteriorally, and the epiglottic tubercle and margin anteriorally. Given the relatively flaccid nature of the aryepiglottic fold, the primary movement occurs throughout the posterior region, presumably originating at the level of the ventricular fold/ligament, and projecting superiorally throughout the quadrangular membrane and associated structures, until it reaches the fold edge. A slight degree of vibration can also be observed traveling across the inner surface of the epiglottis in the form of a mucosal wave. The use of motion vector analysis is intended to quantify the directionality of the gross movements of the aryepiglottic fold and the mucosal wave transmission.

4.3.2 Voiced aryepiglottic trilling: aperture analysis

To carry out the aperture analysis, *regions of interest* (ROIs) were identified over each aryepiglottic aperture. To select a suitable ROI, the images were loaded into ImageJ, open source image analysis software developed by the National Institutes of Health. ImageJ allows for image sequences to be tracked using polygons defined by vertices that can be freely adjusted by the user. An initial, minimally sized region of interest (in the form of a rectangular polygon) was

established over a single aryepiglottic aperture and then the sequence was advanced to the next frame. If any region of the aperture occurred outside of the ROI, the polygon was adjusted to enclose the area. This technique was carried out until the entire length of the sequence had been viewed. It should be noted that the ROI polygon was only enlarged during this procedure. Once complete, the pixel coordinates of the ROI were recorded and then the image sequence was reset to the beginning to establish a ROI for the opposite aryepiglottic fold. This technique was applied to all trilling sequences analyzed.

With the ROIs established, the next phase of the analysis commenced. For each frame in the sequence, the entire frame (stored as a TIFF file) was first loaded into MATLAB working memory and the image was cropped down to the ROI of the current aperture sequence being analyzed. The cropped region was then inverted[79] and converted to binary format (stored as a logical array) using an optimal luminosity threshold determined through several trial runs. The choice of the value was based on experimentation; the goal was to find a value that converted everything but the aperture area to black. This was difficult to achieve as the range of luminosity in the aperture cavity was considerable, as was the variation in light that reflected off the aryepiglottic folds and epiglottic tubercle regions. Inevitably, some frames contain light bleeding into the aperture region. To minimize the impact that this would have on the area measurements, a binary mask was applied. The binary mask was selected by hand: the goal was to obtain the maximal region of the area occupied by the luminous portions of the aryepiglottic fold and tubercle of the epiglottis. However, no masks were applied to the left aperture analysis, as there were no frames in which the aperture appeared to be truly closed. Thus, there were no frames for the left aperture that could be used as a mask, as this would result in a loss of area information for the aperture itself, clearly an undesirable result. The application of the image mask was then followed by the use of a MATLAB algorithm that fills in any holes in the aperture region. At times, due to increased aperture size, specular light reflection from the laryngoscope light source on the internal surfaces of the aryepiglottic fold aperture meant that there were 'holes' in the image when it was converted to binary format. Thus, to improve the accuracy of the measurements it was desirable to fill in the holes using MATLAB's `fill` routine. Finally, once the image was fully processed, the area of the remaining aperture could be measured using

[79] Inversion was done to turn the dark area of the aperture white, which is what is interpreted as an area according to the `bwarea` function.

MATLAB's built in function, `bwarea`, for measuring the area of binary 'on' regions[80]. Once all of the aperture area data were collected, they were then converted to percentage form, using the largest area value in the sequence as the basis for 100% opening of the aryepiglottic aperture. This process is illustrated in Figure 4.10, with the exact values used in Table 4.3.

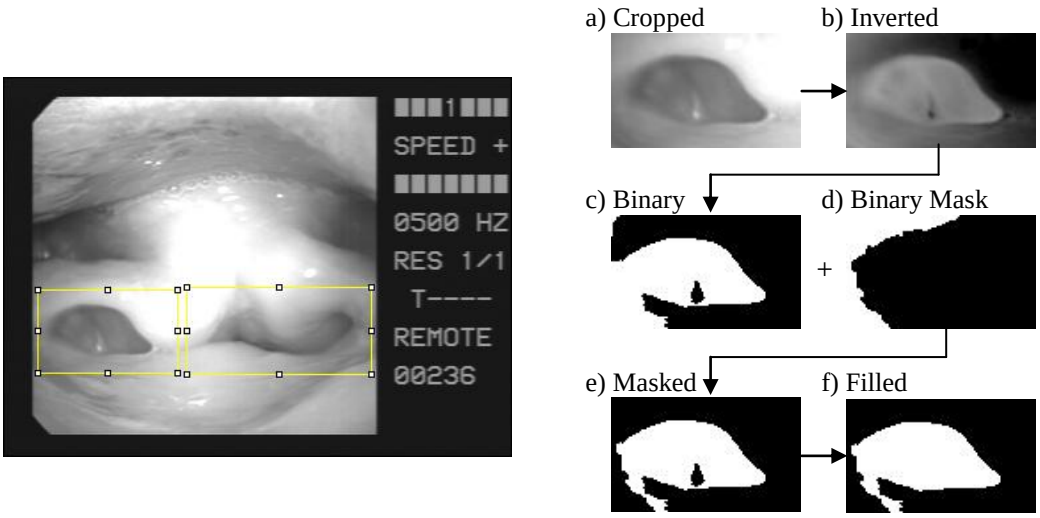


Figure *4.10*: Illustration of the image processing involved in the aperture analysis. The main figure on the left is a frame (#236) extracted from the first sequence of voiced aryepiglottic trilling. The two *regions of interest* (ROIs) for the left and right apertures have been outlined in yellow rectangles/polygons. The series of images on the right illustrate the processing sequence (for the right aryepiglottic aperture). Image (a) is the cropped right aperture used to illustrate the transformation from the greyscale image to the processed binary image, on which the area measurements are taken. The same steps, save for the application of binary masks, were applied to the left aryepiglottic aperture. Image (b) is a 1:1 inversion of the cropped ROI. The image is then converted into binary, seen in image (c), using a greyscale thresholding value (provided in Table 4.3).

Table 4.3 Values used for the aperture analysis & number of frames per sequence

Sequence	#1		#2	
# of Frames	578		275	
Start Frame	200		1060	
End Frame	777		1334	
Duration (ms)	1.156		0.550	
Aperture	Right	Left	Right	Left
Threshold	0.46	0.55	Same as seq.1	
ROI width[81]	133	106	" "	
ROI height	59	54	" "	
Masking frame	200	-	1054	-
Min Area[82]	0	153.1	0	310.7
Max Area	2442.9	2233.6	3083.9	2414.7

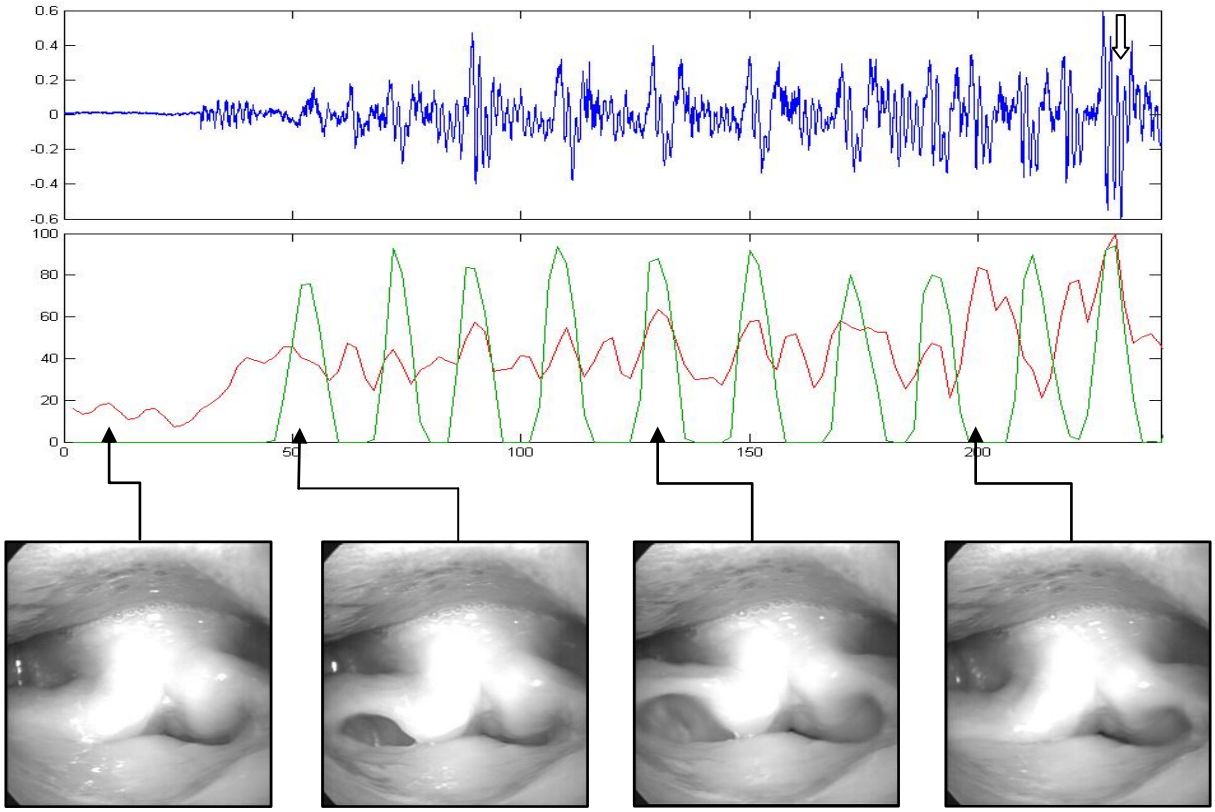
[80] I.e. the areas represented by the digit *1*, which appear as white regions in the binary pictures in Figure 4.10
[81] All height and width dimensions are given in units of pixels.
[82] Areas provided are in pixels^2.

The use of the binary format to conduct the analysis has the obvious drawback of confounding some of the area measurements with information about the luminous areas of the laryngeal collar surfaces. The loss of accuracy comes with the benefit of the speed of an automated analysis. Automated edge detection algorithms developed for other imaging applications (such as ultrasound) fail to properly identify the aperture region due to the tremendous variation in luminosity in the high-speed videos. Additionally, at times the aperture region appears to be blurred due to the tissue velocity, which presents further problems to edge detection algorithms. Furthermore, MATLAB has a host of functions that work exclusively on binary images, particularly the gap filling and area measurement routines. Thus, for the present, binary image analysis is used. This does not preclude the possibility of reanalyzing the images when an improved edge detection algorithm is developed. A visualization of the results of the analysis for the first 250 ms of the first sequence of trilling is presented in Figure 4.11. Frequency and jitter estimates (in the form of standard deviation) for both sequences are presented in Table 4.4.

Table 4.4 Frequency and jitter of trilling sequences compared to modal voice[83]

Sequence	#1		#2		Modal
Aperture	Right	Left	Right	Left	
Frequency (Hz)	54.6	91.7	51.8	99.0	104.2
Jitter (Hz)	4.8	4.5	4.8	3.5	0.00006

[83] Frequency was determined using scripts developed by O'Haver (2006) under the Matlab platform. Jitter is interpreted as the standard deviation of peak-to-peak position differences (i.e. variation in pulse period).



a) Frame 220: 10ms b) Frame 226: 52ms c) Frame 265: 130ms d) Frame 300: 200ms

Figure 4.11: Multiple time series plot of the subject's right (bottom: green) and left (bottom: red) aryepiglottic aperture vs. the audio waveform (top: blue) during the first sequence of voiced aryepiglottic trilling, with selected frames from the high-speed laryngoscopic footage. Aperture ratios are given as percent opening relative to maximal opening in trilling sequence. The abscissa displays time, which commences at the first frame of the trilling analysis sequence (frame 200). Broad arrows in the waveform are used to indicate some potential locations on the waveform where both apertures are open, giving the waveform its distinct double peaked appearance. A thin arrow is used to identify a correspondence between increased aperture areas and markedly increased amplitude in the waveform.

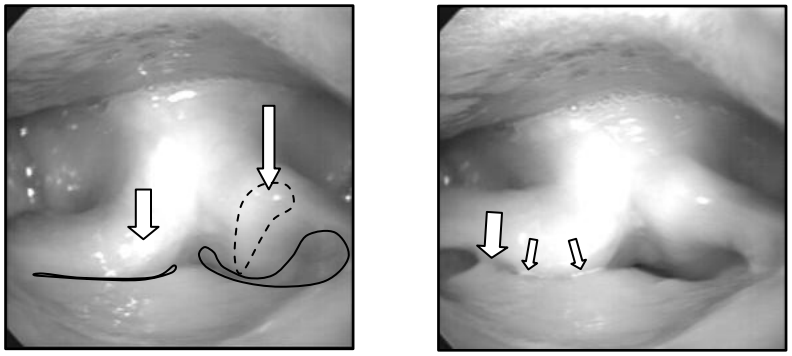
In Figure 4.11, Frame 220 (a) shows the aryepiglottic folds in their pre-trilling configuration; the subject's right aperture is completely closed while the left remains slightly open due to cuneiform positioning (see Figure 4.12). Next, Frame 226 (b) shows the onset of trilling[84]; at this point the right aperture is explosively forced open, while the left aperture gradually increases (from ~20ms to ~55ms) in area as air pressure (expelled from the glottis) works its way up the epilaryngeal tube, setting the left aryepiglottic fold into its pulsing rhythm. From here trilling is fully engaged. There are numerous times when the left and right aryepiglottic apertures synchronize; an image of this phenomena is presented in (c), taken from

[84] The left fold does exhibit movement prior to this, and the waveform indicates glottal pulsing.

frame 265. However, as frame 300 (d) reveals, the trilling pattern of folds is not symmetrical; in this image the left aperture has dramatically opened, while the right remains completely closed. This motion occurs after a synchronized pulse at around 190ms. Generally, the left fold behaves dicrotically, as is evident throughout the aperture analysis, where the initial pulse is of greater amplitude than the following one. As the moment at 200ms (d) illustrates, this is not always the case. Spectral analysis reveals the fundamental frequency of the sequence to be roughly 110 Hz, which is double the frequency of the right aryepiglottic aperture. The left aperture, however, appears to be oscillating at a rate that is closer to frequency of the vocal folds (μ = 95.3 Hz). It is difficult to identify the acoustic contribution of the vocal folds from the waveform as the aryepiglottic sources are superimposed over the vocal source, meaning that there is not a one-to-one correspondence between the physical aperture size and the acoustic signal. As Table 4.4 indicates there is considerable jitter for both apertures in both sequences (μ = 4.4 Hz), which stands in stark contrast to the highly invariant modal voice glottal pulse (e.g. jitter = 0.00006 Hz). It can also be noted that the amplitude of the right aryepiglottic fold pulse tends to be considerably greater than the left pulse. An interpretation of this is that the left aryepiglottic fold vibrates more consistently with the glottal pulse due to its open configuration, meaning that pulse amplitude is reduced in compensation for greater pulse frequency. Negative Bernoulli pressure change across the left aryepiglottic fold surface is likely cutting short its range of movement, by causing the fold to adduct soon after pulse onset.

The occurrence of the asymmetry between the left and right aryepiglottic apertures might receive an explanation from the unique anatomy of the individual performing the trilling articulation. As Figure 4.12 illustrates, when the laryngeal constrictor is engaged in its pre-trilling posture, there is a distinct difference between the right and left cuneiforn and aryepiglottic fold configuration. The right cuneiform fully contacts the epiglottic tubercle and results in a mucosal seal (likely from mucous surface tension) to be formed between the mucosa of the aryepiglottic folds and the epiglottic tubercle's mucosa. The left cuneiform, on the other hand, is seated such that the cartilage axis is pointed obliquely in a postero-lateral direction. As the aryepiglottic fold shape is greatly influenced by the cuneiform cartilages (Painter 1986: 330), the left aryepiglottic fold cannot fully come into contact with its corresponding epiglottic tubercle region. The right aryepiglottic aperture in this case might be referred to as being in a tight configuration, while the left aryepiglottic aperture could be said to be in a loose

configuration. The hypothesis put forth is that it is this anatomical configuration (unique perhaps to this speaker) is what gives rise to the asymmetrical voiced aryepiglottic trilling pattern observed. This issue will be discussed further in Chapter 5.



a) Frame 1: 0ms b) Frame 670: 1s 340ms

Figure 4.12: Images identifying anatomical details of the speaker's epilaryngeal tube that contribute to the pattern of trilling observed in the high-speed laryngoscopic footage. Time in these images refers to the overall footage, as opposed to exclusively the trilling sequence as in Figure 4.11. The slender arrow in figure (a) points to the location of the left cuneiform tubercle; the estimated cartilage shape has been delineated using a dotted line. The broad arrow indicates the location of the right cuneiform tubercle. Solid lines are used to illustrate the shape of the aryepiglottic apertures. Image (b) illustrates the mucosal seal impacting the shape of the right aryepiglottic tubercle during the opening phase of a trilling pulse. The large arrow here points to the coupling of the aryepiglottic and epiglottic mucosa about to be violently separated from the force of the trilling pulse. The smaller arrows point to region along the aryepiglottic aperture where the mucosa is reflecting the camera light, indexing its presence. It is presumed that the mucosal seal must be broken every time the aperture opens, and that the seal is reformed once the right aryepiglottic fold contacts the epiglottic tubercle after the pulse has passed.

4.3.3 Voiced aryepiglottic trilling: motion vector analysis

The motion vector analysis conducted here tracts changes between two contiguous images of the high-speed laryngoscopic footage. Motion vectors are a means to quantify the direction and magnitude of pixel change between two frames. MATLAB does not provide any native support for obtaining motion vector data from a sequence of images. It does, however, contain a useful function, `quiver`[85], that can be used to plot and visualize vector fields. There exist numerous approaches to assessing motion in an image sequence, and to obtain motion vectors, such as systems inspired by biological vision (e.g. Etienne-Cummings 2001), high-level parametric (e.g. Cremers & Soatto 2005) and low-level statistical models (e.g. Polana & Nelson 1997), and techniques used to encode motion in video image sequences for the purpose of digital

[85] The courier font is used here, and throughout, to designate that the entity represents a computer function.

signal compression (e.g. Zheng & Chau 2003). The approach adopted here is to use a simple, block-wise analysis on the binary images generated during the aperture analysis.

The algorithm developed takes the first of a pair of temporally contiguous *regions of interest* from frames in a video sequence. For consistency with the aperture analysis, the same ROIs are used in the motion vector analysis. These images have been previously converted to binary images (during the aperture analysis) and thus the image structure data is numerically simple to track. In the aperture analysis the images were also inverted, thus regions in the original image with luminosity below the specified thresholds are represented numerically by a 1 (termed 'on' pixels). The image is then divided up into square analysis blocks of a desired dimension and each block is scanned to identify the local coordinates of 'on' pixels[86]. The center of each analysis block represents the origin of that block, with coordinates (0, 0); to ensure that the blocks had a single center, odd-numbered dimensions were specified. Once all of the coordinates of the 'on' pixels for a block are collected, the centroids of the *x* and *y* coordinates are obtained independently using formulas provided in Eq. 4.1. The results are then stored in a matrix of the same dimensions as the number of analysis block columns by analysis block rows. Once all of the blocks are assessed, analysis of the second image in the analysis pair can proceed. The same approach is followed to obtain the centroids of the 'on' pixel coordinate for the subsequent image in the frame sequence. With both images assessed, the next step involves calculating the motion vectors for each of the analysis blocks across the frame sequence. This calculation proceeds block-by-block; for each, the difference between the coordinates of first frame and those of the second frame is obtained, which represents the motion vector. There is one special circumstance where a motion vector requires inversion: if the first analysis block is empty and the corresponding block in the second frame has some pixels. The assumption is made that the pixels present in the second frame do not originate within the analysis block, but rather have spilled over from another frame (i.e. an object has moved into the analysis block's region). This assumption exposes the algorithm to making errors where the pixels are in fact originating within the second frame's analysis block. An example of this case would be if the aryepiglottic aperture progresses from a state of full closure to minimal opening, entirely within an analysis block. Depending on the location of the pixels, the motion vector generated will exhibit varying

[86] Under the condition that the analysis block on the edge of the image contained fewer pixels than the rest of the blocks, it was skipped during the coordinate assessment phase. This occurred when the dimension for the analysis block was not evenly divisible into the number of pixels for a given dimension of the analysis image.

degrees of spurious magnitude and directionality. However, it is far more frequent for the pixels to originate outside of the analysis block, and thus inversion of the motion vector allows for the correct trajectory of the motion to be presented (otherwise the motion vector would erroneously point towards the centroid of the second frame's analysis block). An illustration of the overall motion vector analysis process is provided in Figure 4.13.

Eq. 4.1 X & Y Coordinate Centroid Calculations:

$$C_x = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad C_y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

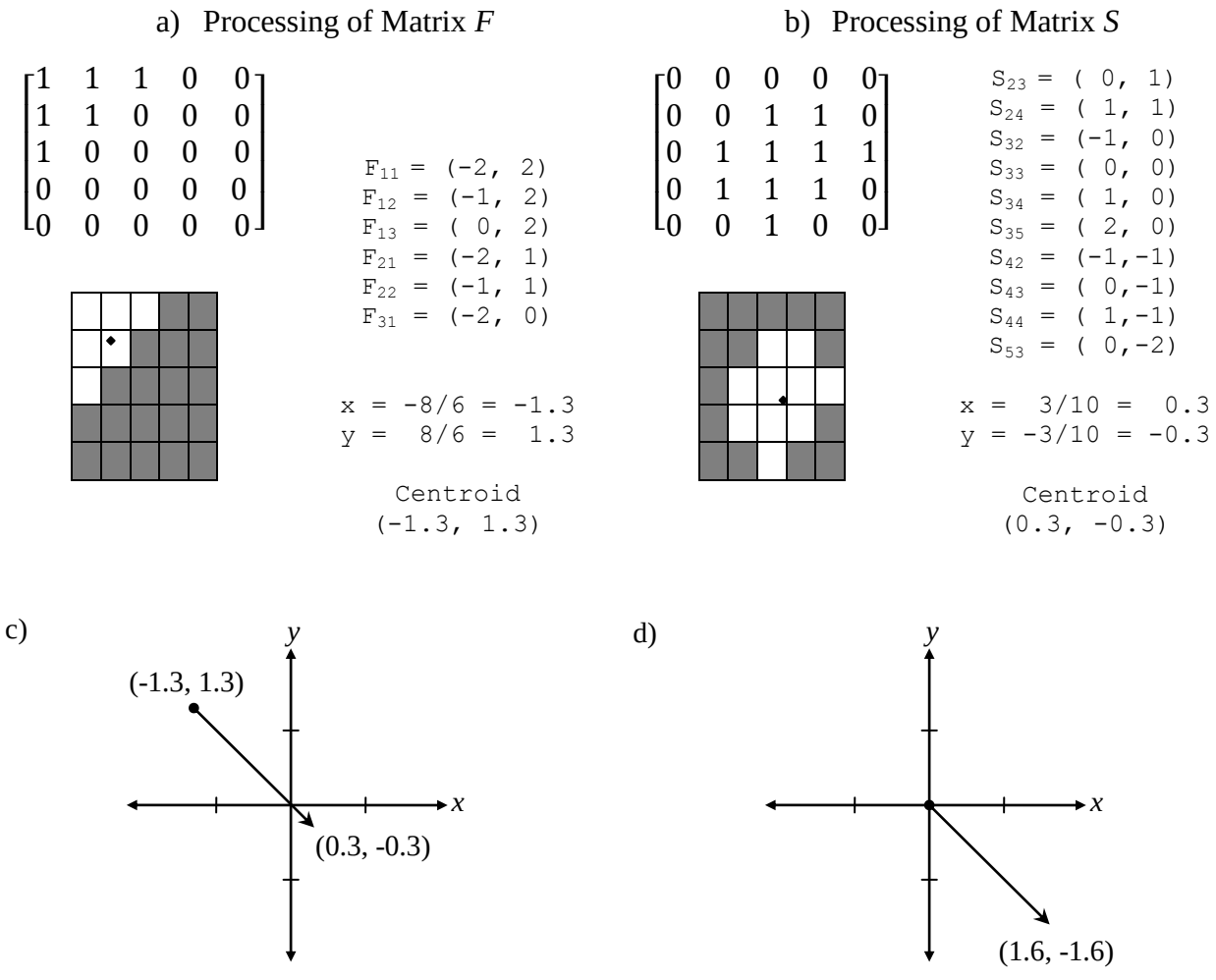


Figure 4.13: An illustration of the motion vector calculation algorithm. The data used in these figures are imaginary, but serve to illustrate, in a simple way, how the motion vector algorithm works. The left hand set of figures in (a) represents an analysis block matrix (top left), an image representation of

the analysis block matrix (bottom left) and the centroid calculation (right) from the first frame in an image pair, while the right hand set of figures in (b) represents the corresponding data from the subsequent (temporally contiguous) frame. The matrices illustrate how binary pixel data are stored and interpreted in the analysis blocks. These particular analysis blocks have a dimension of 5×5 pixels, and are geomentrically interpreted as having the origin of the *x*-*y* planes lay at the center of the matrix (F/S_{33}). Thus, the pixel of the first row and first column of either matrix (F/S_{11}) is interpreted as `(-2, 2)` in geometric space. All 'on' pixels, pixels which are specified as 1 in the analysis block matrix, are enumerated and their coordinates are recorded. Once all of the coordinates have been collected, the centroid calculation applies to both the *x* and *y* coordinates. The diamond in the image representation of the analysis block matrix represents the approximate location of the centroid coordinate.The result of each calculation is the corresponding coordinate of the motion vector origin. The same method is applied to the next image (the second image) in the image analysis pair. The centroid coordinates of this calculation lead to the specification of the motion vector target, which taken together with its origin, allow for the magnitude and a direction of the motion vector to be determined. The coordinate space in (c) represents a plot of this motion vector, corresponding to its actual coordinates as determined from the algorithm. The last step involves translating the motion vector origin to the coordinate space origin, as shown in (d). This allows for the ultimate plot of the vector field to have a more orderly appearance. The actual code used for this analysis is presented in Appendix C.

With the motion vector data aquired, it becomes possible to plot the vectors using MATLAB's `quiver` function. Unless specific origins are supplied, `quiver` will divide the plot up into a grid, which is sized according to the dimensions of the input matrices. To facilitate the interpretation of the motion vector data, the choice has been made to translate all of the vectors to their coordinate space origin (see Figure 4.13d), which normalizes the motion vectors at the cost of removing data concerning the origin of the vectors. The focus of this analysis, however, is the directionality of motion, not the specific location, thus the choice allows for a much more unified vector field to be plotted.

The other dimension of this analysis involves obtaining movement directionality data. The vector field plots serve to visually indicate the general location of motion and the distribution of the motion directionality. However, the motion vectors can also be interpreted into angular data to quantify the overall movement tendencies of the image. There are two primary assumptions made for this analysis to be possible. The first assumption is that motion orthogonal to the *x*- or *y*-axis is interpretable as motion in an antero-posterior or latero-medial direction, respectively. The basis for this assumption is that the placement of the camera within the pharyngeal cavity, during filming, was aligned according to these axes. As is apparent in the high-speed footage, this assumption seems reasonable: the axis formed by the rear pharyngeal wall and epiglottis corresponds well with the *y*-axis in the resulting image. The second assumption is that the camera was stable during filming. Whether this is the case has not been confirmed, but it is dubious that no motion whatsoever occurred during filming as the camera

was not fixed in place within the pharyngeal cavity. Larger movements would likely not influence the analysis on a frame-by-frame basis in any appreciable way, due to the considerably high frame rate. Whether more minute vibrations have an impact on the analysis has not been determined. Thus, the reliability of the results reported here requires further investigation of more video of the same speaker, as well as that taken from other speakers. If the results of future analyses can be shown to reflect the patterns observed in the present results, then technique could be considered more robust in characterizing aryepiglottic movements.

Assessing the directionality of movement is conducted using a simple calculation to obtain the reference angle of the motion vector using the inverse sine of the motion vector's target coordinates. The specific formula for the reference angle calculation is presented in Eq. 4.2. A diagram illustrating the reference angle from the motion vector obtained in the previous example (Figure 4.13) is represented in Figure 4.14. This method of assessing the angularity of the motion vectors conflates opposite motion directions together; thus, purely anterior and purely posterior motions are conflated into the single category antero-posterior. The same is true for latero-medial and non-orthogonal, radial, or oblique motion. The idea is to abstract beyond specific direction and obtain an estimation of the predominant motion types that occur (be they antero-posterior, latero-medial, or radial). Furthermore, rather than assessing raw directionality, the magnitudes of the motion vectors are also considered and act as weights in assessing the proportion of motion in a particular direction category (simplified to 0°, 15°, 30°, 45°, 60°, 75°, and 90°[87]).

Eq. 4.2 Calculation used to obtain the reference angle of motion vectors:

$$\theta = \sin^{-1}\left(|y| \div \sqrt{x^2 + y^2}\right)$$

[87] All radian values are converted to degrees for ease of comprehension. The standard formula, $rad \times \frac{180}{\pi}$, is used for the conversion to degrees.

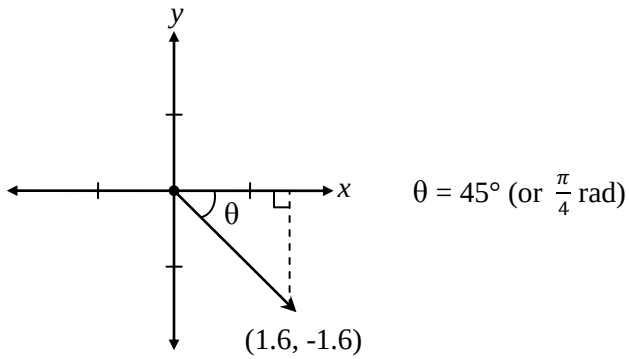
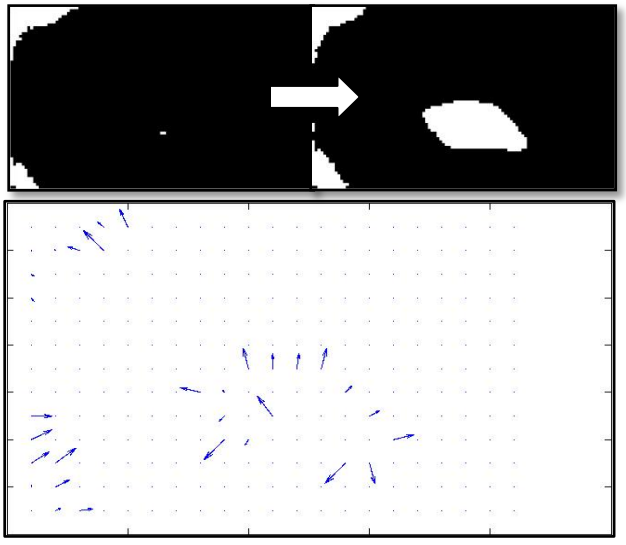


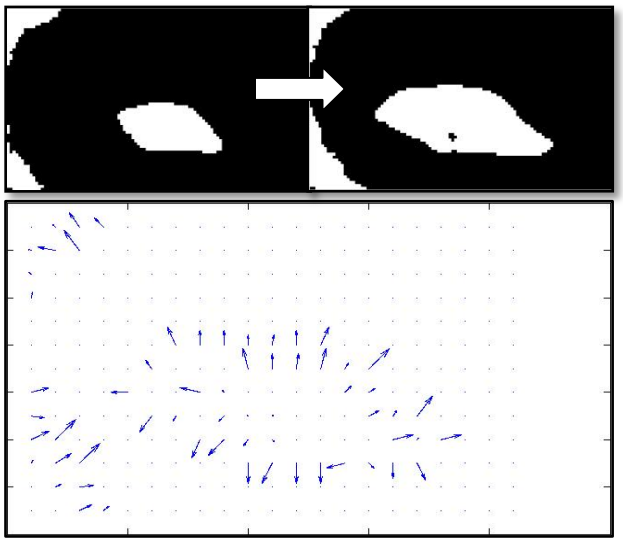
Figure 4.14: Location of the reference angle for the motion vector determined in Figure 4.13. The magnitude of the motion vector is equivalent to the hypotenuse of a right-angled triangle formed with the *x*-axis. This number is divided into the absolute value of the opposite side of the triangle (equivalent to the *y* coordinate). Absolute value is used here to avoid negative angles. The sine is used rather than the tan ratio to avoid undefined answers, where the motion is anteroposterior (where the adjacent side of the triangle is 0).

The results of the motion vector analysis of voiced aryepiglottic trilling during the harsh voice at low pitch sequences are presented below. The next set of figures (Figures 4.15 & 4.16) provides the motion vector plots of the opening phase frames for the left and right aryepiglottic apertures. Directionality bar-graphs for these frame sequences of each corresponding frame sequence are presented below the motion vector plots. The complete, frame-by-frame illustration of one full trilling pulse for each of the aryepiglottic folds is presented in Appendix D.

a) Frame 1 to 2:



b) Frame 2 to 3:



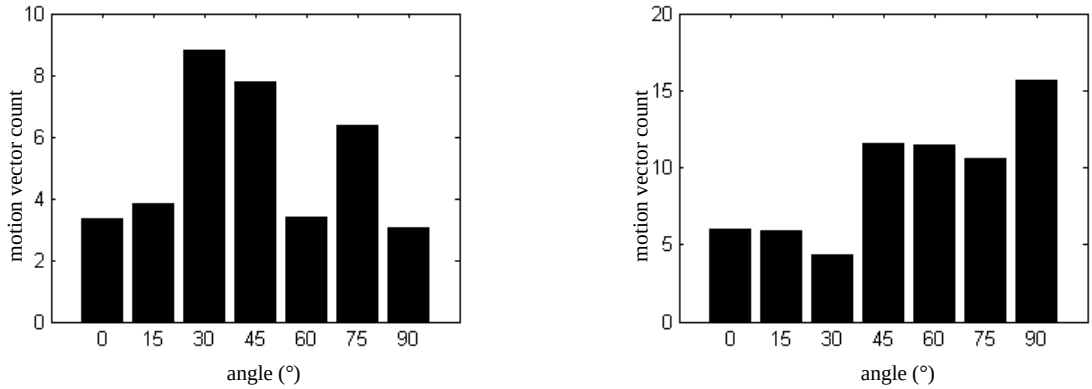


Figure 4.15: Motion vector analysis results: two sequences from the subject's right aryepiglottic aperture. The images on top are the cropped binary images used to conduct the motion vector analysis; the vector plot below shows the direction and magnitude of motion that occurred across the analysis images. In (a) is the transition from frame 1 to frame 2; (b) shows the motion from frame 2 to frame 3. Corresponding bar-graphs of motion directionality are presented below each frame sequence (N.B.: 0° = medio-lateral motion; 90° = antero-posterior motion).

a) Frame 1 to 2 (first pulse):

b) Frame 6 to 3 (second pulse):

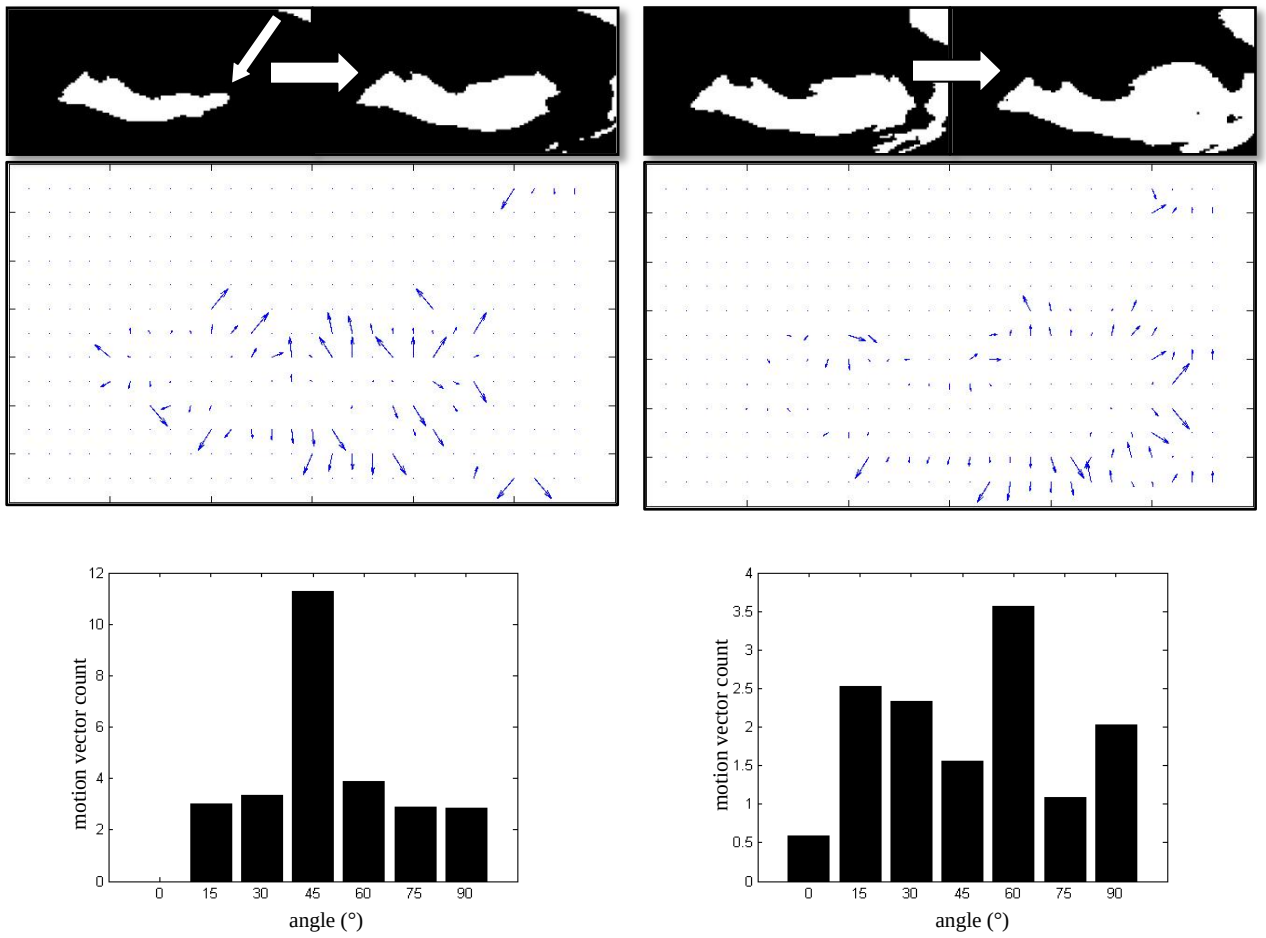


Figure 4.16: Motion vector analysis results: two sequences from the subject's left aryepiglottic aperture. The overall layout of this figure is the same as Figure 4.15. Note that (a) shows the transition between frame 1 and 2. Image (b) shows the transition between frame 6 and 7, which is a secondary pulse or sub-pulse for the cycle of trilling. The thin arrow in the first sequence marks the inflection

point of the left aryepiglottic fold, where there is a considerable bend, as the fold courses up into the lateral margins of the epiglottis.

Our primary concern is with the motion of the aryepiglottic folds during the burst phase of the trilling cycle (i.e. when the aperture begins to enlarge). From the quantitative analysis of motion vector directionality in the first frame sequence (figure 4.15: frames 1 to 2), the subject's right aryepiglottic fold shows strong tendency for fold motion to be angled between 30° and 45°, with weaker medio-lateral and antero-posterior motion (see corresponding bar graph). The subsequent sequence then shows a dramatic increase in antero-posterior movement. For the first sequence (frames 1 to 2), the aperture transitions from a nearly fully closed configuration to a roughly ellipsoidal one as the fold is pulled away from the epiglottis; there is still considerable contact between the right aryepiglottic fold and the epiglottis at this point. The next frame sequence indicates that, once the aperture is formed, motion becomes more antero-posterior, as force from the glottal pulse propels the fold towards the posterior pharyngeal wall. There is a secondary trend towards radial movement as well in this sequence, as the corners of the aperture expand.

The subject's left aperture behaves somewhat differently. For the first sub-pulse, the aperture is moderately patent from the start, and never really closes throughout the entire sequence (as seen in the aperture analysis Section 4.3.2). There is a salient tendency for 45° motion, as seen for the first sequence of the right fold (see corresponding bar graph in Figure 4.16a); the aperture enclosed by the point of inflection in the fold (marked by the thin arrow in Figure 4.16: a) is enlarged along its three sides, while the more medial part of the aryepiglottic fold, which contains the cuneiform cartilage, shows a moderate movement tendency at a 45° angle. The next frame sequence skips ahead to frames 6 and 7, where the next sub-pulse in the trilling cycle occurs. The magnitude of the motion has diminished considerably, compared with that of the first pulse, although the directionality is consistent with the first sub-pulse. Again, the medial portion of the fold appears to be stiffer than the lateral portion, which experiences a radiating pulse, resulting in directionality in the 30° to 60° range.

The motion vector analysis also reveals another interesting observation about movement of the right aryepiglottic fold. The movement direction of the fold itself becomes non-uniform half-way through the aryepiglottic pulse. The medial portion begins its recoil, while the lateral portion is still being expelled outwards by the glottal pulse. Figure 4.17 provides an illustration

of this phenomenon. It could be that, because the lateral portion of the aryepiglottic fold is farther from the glottis, the response of the lateral part of the fold is slightly delayed and out of phase with the medial portion.

a) Frame 4 to 5: b) Frame 5 to 6:

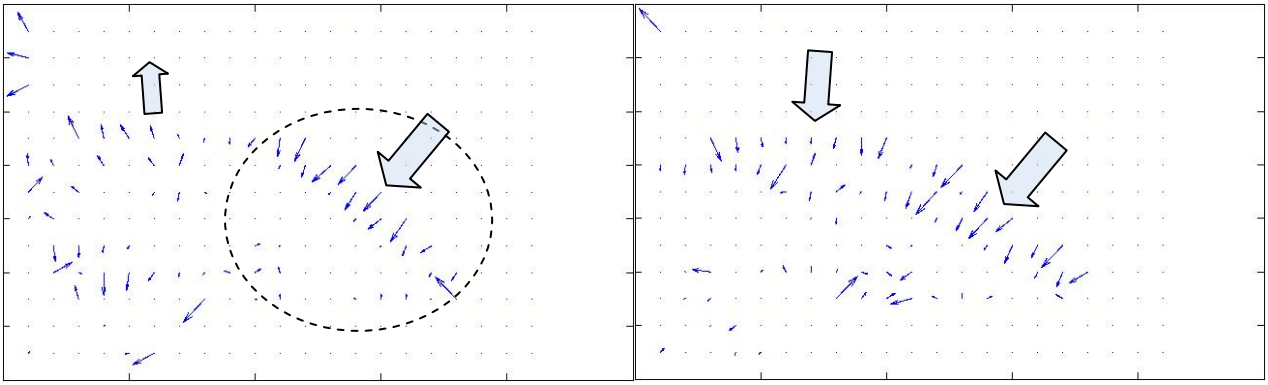


Figure 4.17: Lateral-medial difference in (the subject's right) aryepiglottic fold movement. The left frame sequence (a) occurs immediately after the full aperture has formed and the fold is beginning to recoil. The motion vectors of the medial portion (circled) indicate that the medial portion starts its recoil earlier than the lateral portion, which begins in the next frame sequence. The large arrows are used to indicate the overall estimated motion direction of the motion vectors within proximity of that arrow.

Unfortunately, the directionality assessment of the motion vector analysis results must be considered in the light of a possible confounding factor: as is evident in the binary images, non-aryepiglottic aperture regions are subject to analysis. These areas are the sunken regions around the external margins of the aryepiglottic fold, particularly in the direction of the pyriform fossae. The number of motion vectors counted for these areas is in the neighbourhood of 10-18. Fortunately, however, as motion of the fold is consistent both internally and externally, the values counted often reflect the trends observed for the internal margin of the aryepiglottic fold.

The motion vector analysis is important in determining the force vectors operating on aryepiglottic fold particles in the biomechanical simulation of aryepiglottic trilling discussed in Section 5.3. The goal is to apply forces that reflect the directionality of movement for the trilling pulses observed in Figures 4.15 and 4.16. It is acknowledged that these force contours are a crude approximation of the actual flow patterns that occur in real larynges under conditions of trilling. Simulation of this more complex aerodynamic phenomenon is reserved for future research into the fluid dynamics of air particles during trilling.

4.4 Summary of laryngoscopic contributions to the 3D Model

Three larygoscopic investigations have been discussed in this chapter. The first involved the measurement of specific anatomical landmarks of the larynx which serve to characterize laryngeal constriction; this was done using normal speed laryngoscopic video. The next two investigations examined high-speed laryngoscopic video of voiced aryepiglottic trilling using image analysis techniques. The first of these involved the identification and measurement of aryepiglottic aperture regions to ascertain the frequency and amplitude of the trilling pulse. The second technique assessed motion vectors in contiguous frames of the videos, with particular emphasis on the burst phase of the aryepiglottic pulse.

Data and observations derived from these examinations are incorporated into the model in several ways. Normal speed laryngoscopy serves as a basis for the construction of cardinal postures of the larynx during laryngeal articulations and phonation types. Observations from the high-speed video form the foundation upon which the biomechanical model of the aryepiglottic folds is constructed, both conceptually and with regards to the parameters that control the trilling.

The next chapter focuses on the biomechanical simulations of the model and draws on much of the discussion of Chapter 4, with regards to the process of aryepiglottic trilling. Further details concerning the application of measurement data from the normal speed laryngoscopy is discussed in Chapter 6.

Chapter Five

A Biomechanical Model of Aryepiglottic Trilling

5.1 Introduction

The goal of this chapter is to provide an outline of the assumptions behind the biomechanical model of the aryepiglottic folds and the technical details of its implementation in the 3D model. The chapter begins (Section 5.1.1 & 5.1.2) with a discussion of aryepiglottic trilling patterns, based on observations of the high-speed video of aryepiglottic trilling (Sections 4.2 & 4.3). During this section, a brief excursus into a minor electroglottographic investigation conducted to understand the relationship between vocal fold and aryepiglottic function is provided. The next section (5.2) discusses the 3D LCM's implementation of Titze's (1973, 1974) model of the vocal folds, which act as the input to the model of aryepiglottic trilling. With these preliminary matters discussed, the final section (5.3) describes the myo-elastic (Section 5.3.1) and aerodynamic (Section 5.3.2) aspects of the biomechanical model of aryepiglottic trilling as it is implemented into the 3D model.

5.1.1 Patterns of Aryepiglottic Trilling

Two patterns of aryepiglottic trilling are distinguishable from the high-speed laryngoscopic video (Section 4.3). It appears that asymmetry in the positioning of the cuneiform cartilages for our subject gives rise to bilateral asymmetry of aryepiglottic fold positioning, which is especially important when the epilarynx is largely or fully constricted as in epiglottal stop. These two cuneiform configurations are termed *loose* and *tight*, which are intended to be descriptive of whether or not the cuneiform tubercle contacts the tubercle of the epiglottis. In the loose configuration (the subject's left cuneiform), the long axis of the cuneiform points supero-postero-laterally, which precludes any contact between the aryepiglottic fold and the epiglottis (see Figure 4.12a). Presumably, given the variability in cuneiform development across individuals (Krmpotic 1957), the degree of looseness can be considerable. I assume, however, that for any particular individual, the cuneiform setting observed in a constricted laryngeal posture is largely immutable. The other possibility, the *tight* configuration, entails that the cuneiform cartilage is positioned so that it presses upon the posterior side of the epiglottis, in the region of the tubercle. Consequently, in this configuration the aryepiglottic fold makes complete

contact with the epithelium of the epiglottis. It is suspected that, as a result of this contact, a mucosal seal is created, which adds resistance to aerodynamic forces acting to separate the aryepiglottic fold from the epiglottis during trilling (see Figure 4.12b).

Unlike bilabial, lingual, and uvular trills, aryepiglottic trilling is unique in that there are potentially two structures vibrating instead of one. This gives rise to a particularly complex acoustic output as each of the aryepiglottic folds must be considered a source, in addition to a possible glottal source. Furthermore, the adjustments to the articulators required to generate an aryepiglottic trill, also are suspected to restrict and interfere with the movement of the vocal folds. The consequence of this is that vocal fold vibration tends to be somewhat irregular both in frequency and amplitude, giving rise to harsh phonation (Laver 1980; see Section 2.1.3). By and large, however, the aryepiglottic trill shares much in common with the other trills. Aerodynamically, trilling in general is optimized when the tissue undergoing the trill is well supported by adjacent structures and forms a "funnel-like constriction, with the articulator opening along the direction of the airstream" (Barry 1997: 36). Fundamentally, this configuration takes advantage of the Bernoulli Effect when a sufficient airstream is generated (Catford 2001: 173). Thus, bilabial trills with the lips drawn inwards, lingual retroflex trills, and uvular trills with the anterior surface of the uvula contacting the tongue dorsum are all unattested, in addition to ingressive trilling of any of the articulators[88]. The shape assumed by the aryepiglottic folds in their constricted posture conform to these requirements, but the base frequency of the aryepiglottic trill, at approximately $f_0/2$[89] (see also Sakakibara 2004b), does not correspond well with the frequencies observed for all of the oral trills, which range from 14 – 50 Hz (Barry 1997: 36), with values commonly appearing around 25 Hz (e.g. Lindau's (1985) mean rate obtained for apical trills).

Figure 5.1 presents paired waveforms and electroglottograms (EGG)[90] showing the effect of aryepiglottic trilling on the acoustic signal and concurrent glottal vibration. Basic

[88] With the possible exception of ingressive velo-pharyngeal port trilling as in snoring, which is not attested as a speech sound.

[89] Which equates to approximately 50 Hz for the male larynx. As of yet, aryepiglottic trilling in the female voice has not been studied. If Sakakibara et al. (2004a, b) is correct, that trilling is a function of glottal pulse frequency, then the predicted base frequency for female aryepiglottic trilling would be 100 Hz (assuming a female f_0 of 200 Hz).

[90] A portable EGG device developed by Laryngograph Ltd. was used to obtain EGG signals. The audio signal was simultaneously recorded using an Audio-Technica AT3035 Cardioid Condenser Microphone. The two signals were digitized through the left and right channels of an Edirol UA-25 USB audio capture device at 44100 Hz, 16 bit and captured using Cool Edit Pro 2 (Syntrillium Software Co.) . Finally, images of the signals were retrieved using PRAAT (version 4.6.22; developed by Paul Boersma and David Weenink).

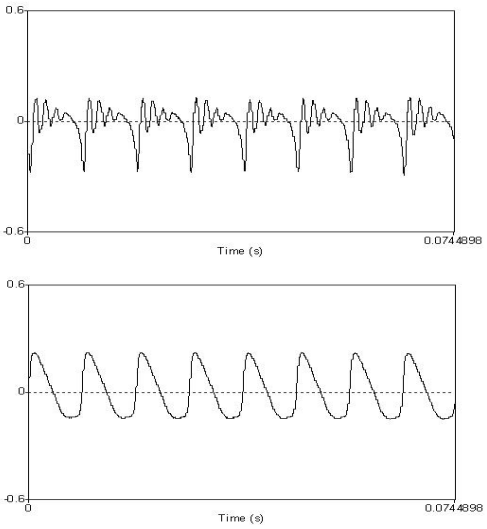
aryepiglottic trilling (*Fig*. 5.1b) without a glottal source (as evidenced by the EGG signal) is highly noisy and shows a salient spike in amplitude (broad arrows) generated at approximately one half of the speaker's f_0 for modal voice[91] with neutral pitch (*Fig*. 5.1a). The fact that trilling occurs at the apparent $f_0/2$ frequency without voicing, indicates that the frequency of pulse is not necessarily causally related to the glottal pulse, but may rather reflect the inherent biomechanical properties of the aryepiglottic folds themselves. Voiced aryepiglottic trill (*Fig*. 5.1c) shows evidence that the primary aryepiglottic pulse is maintained at the putative $f_0/2$ (broad arrows), with an overall appearance of modal voice, save for the presence of noise immediately before the next pulse (thin arrows) occurs. Comparing the EGG signals of these sounds reveals that the vocal folds open more rapidly[92] after closure during aryepiglottic trilling. The acoustic signal shows that vocal fold closure in modal voice corresponds with a valley in the acoustic waveform, while the same is true with aryepiglottic trilling, the valley is followed immediately by a substantial peak, for both the primary pulse and the non-primary (or off-beat) one. Whispery voiced aryepiglottic trilling (*Fig*. 5.1d) was also examined because it more closely matches the phonation type used by the speaker in the high-speed laryngoscopic video (Section 4.3). Again the primary pulse occurs regularly at $f_0/2$ (broad arrows). The acoustic signal is considerably noisier, as expected from whispery phonation, while the EGG signal has a 'double peaked' or peak-plateau waveform, with corresponding negative-positive spikes in the acoustic waveform. This EGG signal corresponds with Esling's (1978: 245-247) ventricular voice, which is considered an extreme variety of harsh voice. He infers from the EGG waveform that it involves rapid glottal opening with a prolonged closure phase, distinguishing it from harsh voice, which has rapid opening and closure phases. The term ventricular is avoided in the present analysis on auditory and acoustic grounds as the phonatory quality and waveform suggest high airflow, indexical of whispery voice. Two other voiced aryepiglottic trilling sequences are analyzed, where fundamental frequency has been manipulated. Figure 5.1e shows that the trilling pulse becomes somewhat

[91] In terms of EGG waveform shape, modal voice for this speaker is somewhat atypical in light of Painter's (1988) normative study on EGG waveforms. Normal or modal voice for Painter's subject (who was a professional singer) has a distinct closed phase, and two opening phases, one gradual, and the other more steep. The EGG signal for Figure 5.1a shows that duration of the closed phase is nearly instantaneous, and the opening phase is monotonic.

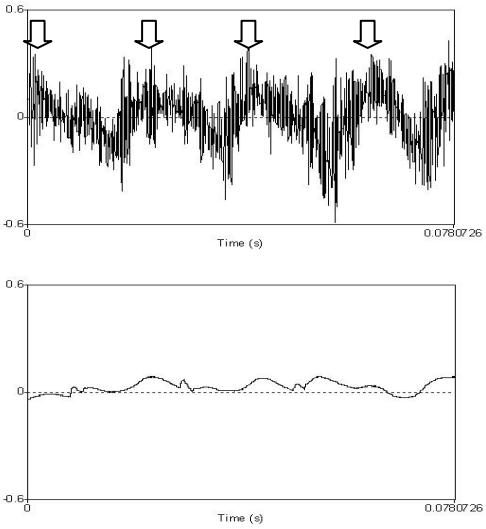
[92] As EGG measures electrical impedance across the glottis (ideally), an increase in glottal area will be indicated by an increase in impedance on the EGG (thus a peak indicates glottal opening), while the reverse is true for glottal closure.

unstable; its timing alternates between two and three glottal pulses[93], although a 3:1 glottal-to-aryepiglottal ratio has been observed as well. An even higher fundamental frequency was explored in Figure 5.1f, which shows the emergence of a new pattern as indicated by the acoustic waveform: the only index of aryepiglottic trilling visible is the appearance of a beating phenomenon, with a recurrent component at 68 Hz identified here as the primary aryepiglottic pulse.

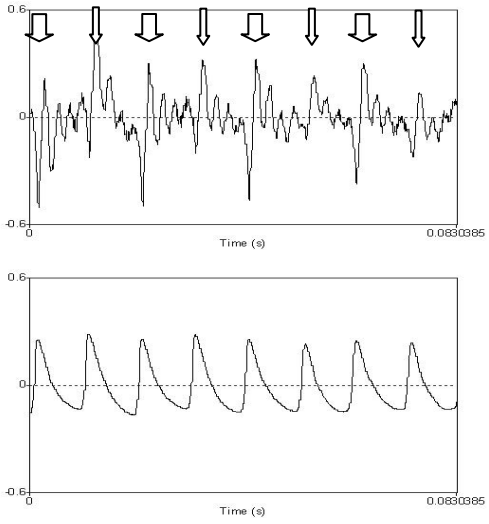
a) modal voice
(F0 = 107.52 Hz)



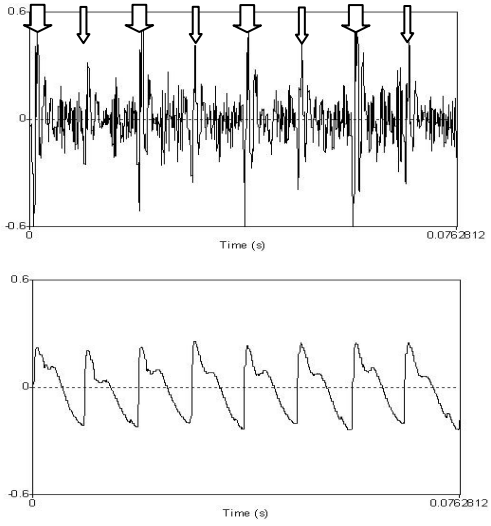
b) voiceless aryepiglottic trilling
(primary pulse = 54.79 Hz)



c) voiced aryepiglottic trilling
(F0 = 96.77 Hz; primary pulse = 48.46 Hz)



d) whispery voiced aryepiglottic trilling
(F0 = 105.63 Hz; primary pulse = 52.91 Hz)



[93] This could be described as a 5/4 rhythm.

e) voiced aryepiglottic trilling (aperiodic)
(F0 = 232.01 Hz; primary pulse = ? Hz)

f) voiced aryepiglottic trilling (high pitch)
(F0 = 347.22 Hz; primary pulse = 68.64 Hz)

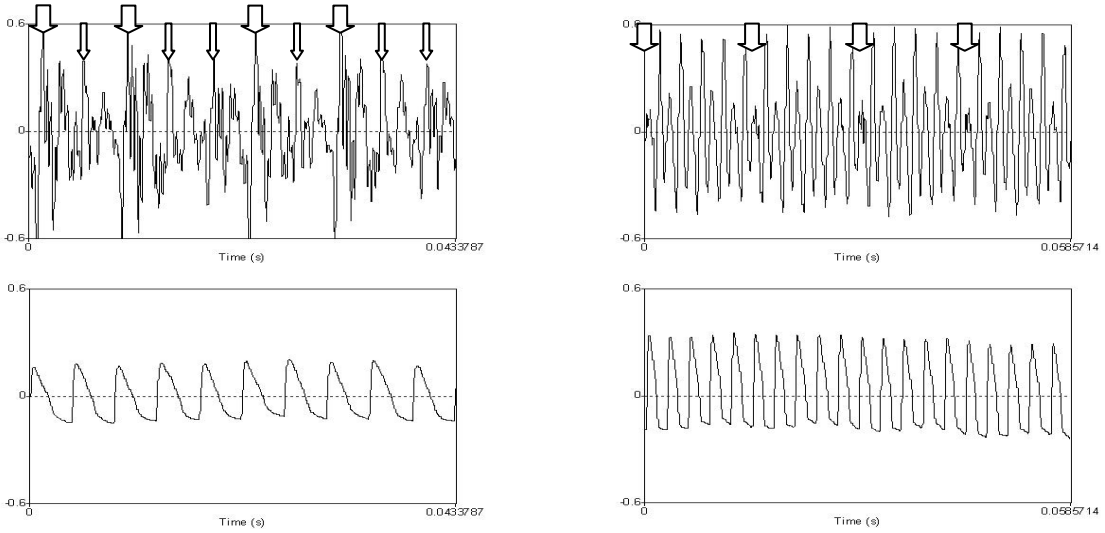


Figure 5.1: Six acoustic (always on top) and EGG (always on bottom) waveform sequences comparing aryepiglottic trills[94]. Broad arrows are used to identify the approximate locations of what is suspected to be the primary aryepiglottic pulse. The thin arrows identify a secondary pulse that could be glottal or a superposition of glottal and aryepiglottal pulses.

The figures discussed above are provocative in that they reveal not only variation in the rhythm of the trilling patterns, but also the considerable stability of the glottal pulse, despite aryepiglottic activity, which is predicted to induce compression of the epilaryngeal tube and likely force the ventricular folds upon the vocal folds, impeding their function. By and large, the tendency for aryepiglottic trilling at $f_0/2$ was observed. This mini-study is only intended to be probative, however, and provides the impetus for a more thorough investigation.

5.1.2 Loose & Tight Cuneiform Configurations[95]

The two cuneiform settings described above are assumed to play a major role in determining the interaction between the glottal pulse and the periodicity of the aryepiglottic trill. As seen in the high-speed video (see Section 4.3), in the loose configuration the glottal pulse is suspected to be transmitted directly to the aryepiglottic fold, and for the most part it oscillates in phase with the glottal pulse. A mucosal wave can be seen ascending through the vestibule of the

[94] All sequences are from a trained phonetician carefully performing each sound.
[95] This observation is originally discussed in Moisik & Esling (2008).

larynx and up into the aryepiglottic fold. Overall there is a tendency towards semi-periodic oscillation of the loose aryepiglottic fold, with varying (or alternating) amplitudes. In this case, the frequency is typically matched with f_0, but at times the fold appears to vibrate at possibly even double the fundamental frequency. Aperiodic oscillation is also observed. Irregularity of the trill in a loose configuration may be a direct result of the constricted setting inhibiting evenly periodic glottal phonation, in conjunction with the elastic properties of the aryepiglottic fold.

The tight cuneiform configuration, on the other hand, is suspected to give rise to what I am calling the primary pulse pattern or base frequency of the aryepiglottic folds. As has been discussed, the tight configuration may be responsible for the creation of a mucosal seal between the epiglottis and the aryepiglottic fold, which must be broken at the beginning of each pulse. It is assumed that the depth of this seal runs roughly from the upper portion of the ventricular folds to the edge of the aryepiglottic folds. Thus, much of the work done by the vocal pulse is to break open the mucosal seal, beginning with the inferior extent, which receives the glottal pulse, and progressively working upwards. If the seal is strong enough, a single pulse is predicted to be insufficient to fully force the aryepiglottic fold into an open position. As the first pulse is being absorbed, elastic recoil of the aryepiglottic fold likely begins. Bernoulli pressures are predicted to be absent as there is not yet a channel for the air to escape through and the energy of the glottal pulse is spent on opening the lower portion of the aryepiglottic seal. The second pulse is evidently strong enough to break the seal and throw the aryepiglottic fold postero-laterally, but also assumed to generate substantial negative pressure as its velocity increases through the channel formed by the aryepiglottic aperture. With this negative pressure and the fold's elastic restoring forces, the fold immediately begins its recoil, which is apparently quick enough to return the fold to its sealed configuration before the next glottal pulse can escape, at which point the sequence repeats. Voiceless aryepiglottic trilling with a tight cuneiform configuration is suspected to behave in a similar manner, with the exception that the airstream is relatively constant. In this case the aerodynamic pressure required to complete one cycle of the trill in a tight cuneiform configuration is likely to be roughly proportional to its voiced counterpart. Figure 5.2 outlines this process through one cycle of aryepiglottic fold pulsing given a tight cuneiform configuration.

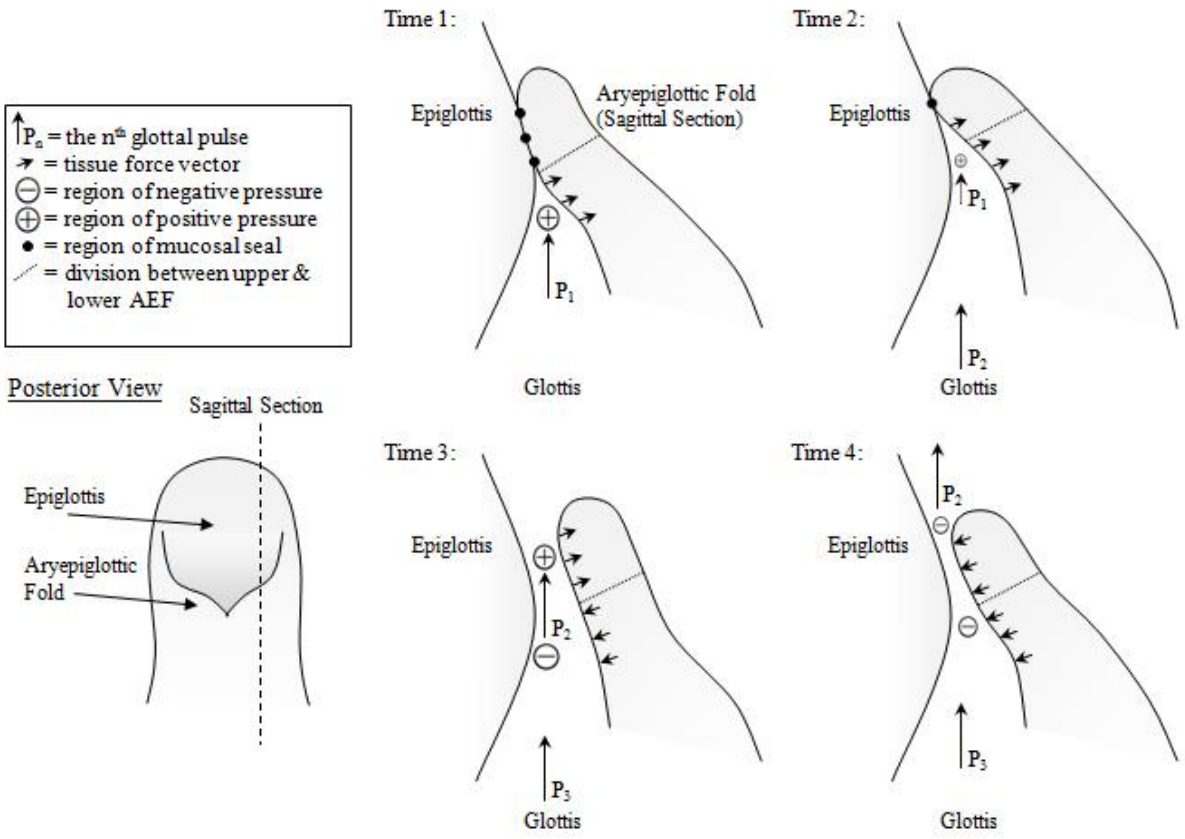


Figure 5.2: Illustration of one cycle of aryepiglottic trilling given a tight cuneiform configuration. At Time 1, the aryepiglottic fold is sealed to the tubercle of the epiglottis by mucous on the surface of the epithelia. The first glottal pulse is trapped within the epilaryngeal lumen and consequently generates expulsive force on the aryepiglottic fold, which causes the lower portion of the aryepiglottic fold to move in a postero-lateral direction. At Time 2, the first pulse has largely been exhausted and its energy converted into work to open up the lower section of the laryngeal seal, which is now held together only superficially by the upper most portion of the aryepiglottic fold (it is possible that it is even beginning to open very slightly at this point). The second pulse begins its assent rapidly up towards the opening. At Time 3, the second pulse forces the aryepiglottic seal to burst open and the aryepiglottic pulse to be generated. Simultaneously the pulse also generates negative pressure as it passes through the newly formed aperture, which, combined with elastic restoring forces, causes the aryepiglottic fold to quickly return to its initial position, with a newly formed mucosal seal.

5.2 Glottal Source

The behaviour of the aryepiglottic folds is inextricably linked to vocal fold function, as has already been indicated in the preceding subsections (5.1.1 & 5.1.2). Not only does vocal fold activity give rise to the broad classes of voiced and voiceless aryepiglottic trills, but it seems that the vocal folds may engage in whispery voicing, which evidently increases the airflow and aperiodicity of the waveform and aryepiglottic pulsing (see Figure 5.1). Breathy-voiced aryepiglottic trilling is ruled out by default as the constriction of the aryepiglottic folds necessary for trilling is antipodal to the unconstricted configuration required for breathy voice. Laver

(1980) and Catford (1977a) identified the articulatory conflict between breathy and harsh voice, but their conception of harsh voice does not necessarily imply aryepiglottic trilling (as discussed in Section 2.1.3). These new aryepiglottic possibilities motivate revising the phonation type taxonomy.

Aryepiglottic stricture is predicted to have a profound impact on vocal fold vibration. This is because of the addition of mass onto the vocal folds from the ventricular folds, which have been pushed over the vocal folds by the action of the aryepiglottic sphincter. Vocal fold vibration in this state is expected to be auditorily harsh, which conforms to Laver's (1980: 118) "extremely harsh voice"[96]. Van den Berg (1955) observes that if the ventricular folds press down upon the vocal folds, the resulting phonation is harsh or "metallic". The effect of irregularity of vocal fold oscillation from cycle-to-cycle, described as *jitter*, is perceptually identified as hoarseness (Protopapas & Lieberman 1997: 2270), which corresponds well with the harsh classification. In turn, this irregularity in glottal pulsing is likely to have a strong effect on the periodicity of the aryepiglottic pulse during trill.

Modeling of aryepiglottic fold biomechanics is, thus, dependent on a model of vocal fold oscillation. It is possible to simulate glottal flow without an entire biomechanical model of the vocal folds, by using a truncated sinusoidal function for glottal area (see Titze 2006: 260). A complete model of vocal fold function is necessary, however, for visual synthesis and is pursued here.

5.2.1 Implementing Titze's 1973-1974 biomechanical model

Titze's (1973-1974) biomechanical model of the vocal folds was introduced in Section 2.3.2. In the implementation of this model, I have attempted to strictly abide by the mathematical formulations Titze used to construct his model. There are, however, a number of deviations from his model that need to be motivated and discussed. Furthermore, the paper discussing his model does not contain all of the details necessary for a full implementation and thus some elements of the current model are not based upon his original formulation.

Perhaps the most important observation to be made about Titze's original paper is that, while the differential equations of motion for particle bodies and volume velocity are provided,

[96] Laver provides the synonym "Ventricular Voice", but maintains that ventricular phonation is not fully understood to merit the creation of a new phonation type for it; rather the label is intended to be more physiologically explicit.

no explicit means of arriving at a solution for these equations is indicated. The differential equations cannot simply be calculated using the independent variables or parameters specified in the paper, they require a solution that either generates the equation for the derivative, known as the analytical solution, or one that approximates the value of the derivative at a specified time based on initial values of the simulation, known as a numerical solution. Perhaps the most widely employed type of numerical solution for first-order ordinary differential equations is the fourth order Runge-Kutta method, which iteratively approximates the derivative with considerable accuracy[97]. Of course, the differential equations of particle motion are of the second-order as they contain second-order derivatives (i.e. acceleration terms), which means that they need to be converted into systems of coupled first-order differential equations. This is done by interpreting the acceleration as the first derivative of velocity. The equation of motion can then be rearranged to solve for acceleration, which enables the derivative of velocity to be assessed at each time step. Equations 5.1 and 5.2 illustrate this process.

Eq. 5.1 Conversion of the second derivative of position (acceleration)

Velocity: First derivative of position[98]

$$\frac{dx}{dt} = v \cong \frac{\Delta x}{\Delta t}$$

Acceleration: Second derivative of position to first derivative of velocity

$$\frac{d^2x}{dt} = a = \frac{dv}{dt}$$

Eq. 5.2 Equation of motion transposed for acceleration

$$F = ma \;\; \rightarrow \;\; a = {}^{F}/_{m}$$

Using this method, the particle system responds with exponential decay to perturbation, as shown in Figure 5.3. Titze (1973: 144) draws on Flanagan's (1971) claim that a suitable damping factor for the abducted vocal fold is $\zeta = 0.16$, but later notes that the value is likely to be

[97] The error term or deviation from the exact solution is described by the local truncation error of the Taylor polynomial, which the Runge-Kutta method is based upon (Burden & Faires 2001: 272-282). For example, the fourth-order Runge-Kutta generates approximately the same error using a time-step of 0.01 ms as Euler's method using a time-step of 0.00000001 ms.

[98] The right most term represents the average velocity for time interval Δt.

at or beyond the critical damping level (i.e. $\zeta = 1$; see Shabana 1991: 79) when the folds are adducted, although no explicit value is provided.

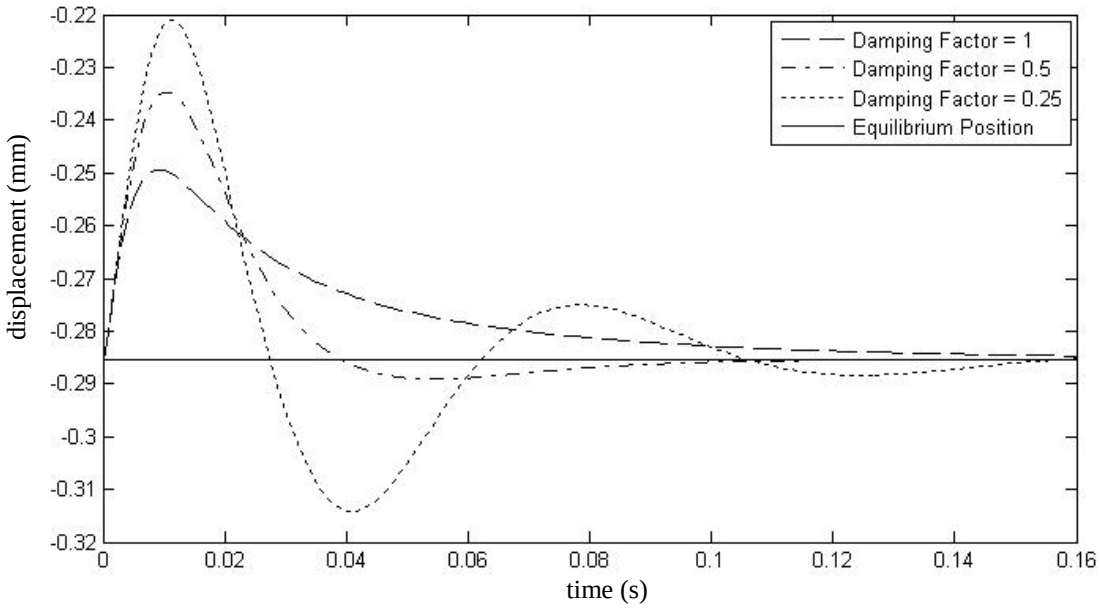


Figure 5.3: Damped vibration of a single particle in the vocal fold simulation using different damping coefficients.

Titze does not provide an explicit definition of the boundaries of the particle system. The effect that boundary location has on the particle system is considerable as the strain gradient is proportional to the equilibrium state of the particles relative to the location of the boundaries. The further the boundary is away from the equilibrium state, the less restoring force will be generated for particle displacement at a given spring constant. A more advanced model, such as one developed using the finite element method (as in Hunter, Titze & Alipour 2004), can model the strain profile through the entire tissue in abducted and adducted states, giving a more realistic characterization of the effect of the boundary on tissue behaviour. The boundary used in the 3D Laryngeal Constrictor Model, however, was determined through trial and error to roughly approximate the shape of the thyroid alae; boundary locations are calculated using the formula in Equation 5.3, which is illustrated in Figure 5.4. The equation determines the *x* position of the boundary relative to the particle in question, while the *y* & *z* coordinates are identical to those of the particle.

Eq. 5.3 Formula for particle boundary *x* coordinates

$$x = \sqrt{\left|\frac{z - 2.4}{-1.4965}\right|}$$

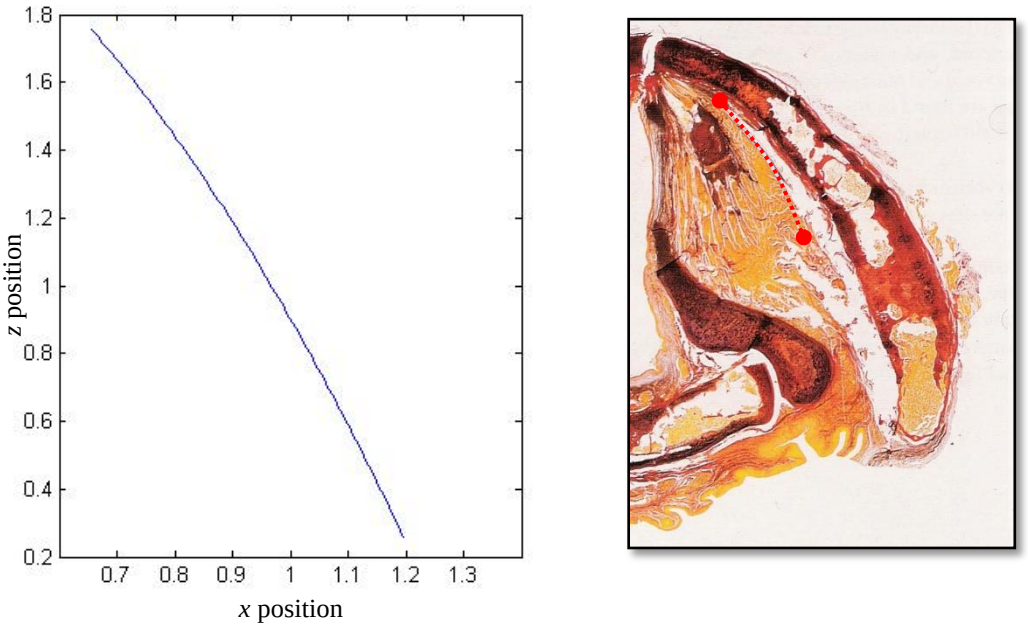


Figure 5.4: Basic approximation of the thyroid boundary condition. The graph shows the contour of the boundary in the *x*-*z* plane. Actual *x*-coordinates are the specific values of the function for the particle's *z* value. The histological section used to create the model meshes is used for comparison; the dotted red line indicates the approximate location of the boundary contour.

Due to the finite time step used in biomechanical simulations, it is necessary to employ a constraint system to prevent particle movement from behaving in anomalous ways. This is particularly evident when stiff springs are used in a simulation (e.g. see Millington 2007: 93). The forces generated from particle displacement can be considerably large for stiff springs, potentially approaching singularities in strain values (Titze 1973: 149). In a discrete temporal setting, these forces are erroneously applied across an entire time step (known as a step discontinuity), producing an unrealistic jump in position. In the real world situation, these forces are instantaneous and dramatically diminish as the spring moves continuously from one position to the next, while opposing restoring forces prevent the system from being unstable.

Particle constraints prevent any such jump in particle movement and ensure that each particle remains within an appropriate distance of its neighbouring particles, neither too near or too far. Titze (1973: 150) employs a constraint system that removes one degree of freedom from

particle movement by converting the coordinates of its velocity, external force, and acceleration into polar form. This occurs when a particle reaches either the minimum or maximum allowed radius relative to its neighbouring particle, or the boundary.

The decision was made to use a different particle constraint system for the present 3D model. Instead of performing the calculations necessary to convert rectangular coordinates into polar form, a simple system based on vector distance is used. DirectX includes a predefined function to calculate the length of a vector, making it more efficient than doing so in program-level code.

The constraint system works as follows. The maximum interparticle distance, or tolerance factor, is determined by dividing the length of the entire particle chain (from its origin to its insertion) by the product of the number of particles and the length squared. Using this formulation allows for the tolerance factor to decrease as the length increases, limiting the amount of movement possible as the tissue stretches. The minimum distance is a constant determined by halving the value that represents the particle's thickness. The actual particle constraint is expressed in the form of the parametric equation of a line, using the tolerance factor divided by the length of the vector representing the difference between the particle and its neighbour. This calculation is performed twice: once for both of its neighbours in the particle chain. The results of the parametric line equation calculations are subtracted from the particle position vector, added together and averaged. Finally, this vector can then be added to the original particle position to obtain its new position. Figure 5.5 demonstrates this process and the equations can be seen in Eq. 5.4.

Eq. 5.4 Equations used to apply particle chain constraints[99]

$$\boldsymbol{i}_{n\pm1} = \boldsymbol{p}_n - \boldsymbol{p}_{n\pm1}$$

$$t_{n\pm1} = r_{max} / \left|\boldsymbol{i}_{n\pm1}\right|$$

$$\boldsymbol{n}_{n\pm1} = \boldsymbol{i}_{n\pm1} - (\boldsymbol{i}_{n\pm1} \times t_{n\pm1})$$

$$\boldsymbol{p}_n = \boldsymbol{p}_n + \frac{1}{2}[(\boldsymbol{p}_n - \boldsymbol{n}_{n-1}) + (\boldsymbol{p}_n - \boldsymbol{n}_{n+1})]$$

subscripts:
max = maximum distance
n = the particle in question
n±1 = the previous & next neighbours

terms:
p = particle position
i = difference between particles
n = constrained position
r = radius
t = parameter for line equation

[99] Terms appearing in bold are vectors, every other term is a scalar.

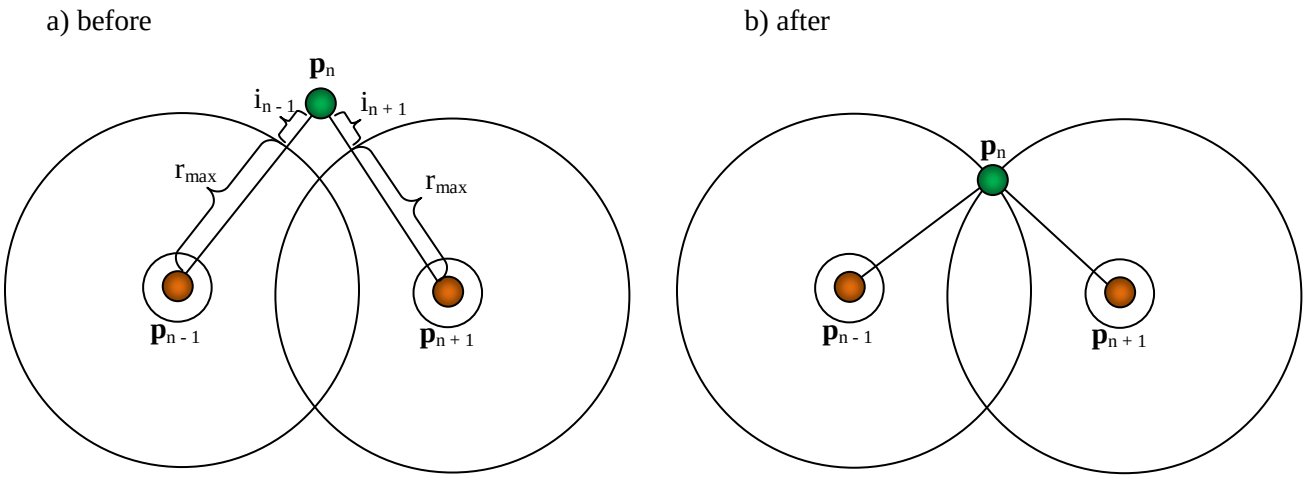


Figure 5.5: Visualization of particle constraint model. Image (a) shows the particle **p** (green) having moved outside of its constraint boundaries defined by the radii of its neighbouring particles (orange). Image (b) shows the particle repositioned after the application of the particle constraints described in *Eq.* 5.4.

One of the other novel features of the 3D LCM is that phonation can be simulated using different vocal fold postures, which is not possible in Titze's 16-mass model (1973), where only the biomechanical properties such as tension and particle mass can be manipulated. Due to this freedom, particle equilibrium position (stored in an array of the initial nearest neighbour values for a given particle) is updated continuously as the model is manipulated by the user. This prevents spurious forces from being generated solely due to control parameter based transformations of particle position.

Finally, there are a number of constants used in the simulation, for which Titze does not provide a value. Additionally, some numbers were manipulated through trial and error as deemed necessary upon repeated examinations of the simulation. A comprehensive list of all simulation constants used in the simulation is presented in Table 5.1. Titze himself notes that many of the constants are not fully substantiated and require adjustment through experimentation (1973: 148-149).

Table 5.1 Values for constants and parameters in the vocal fold simulation

	Parameters		Modal	Falsetto[100]
Biomechanical	Tension	Mucosa	100 g/mm^2	40000 g/mm^2
		Vocalis	40000 g/mm^2	100 g/mm^2
	Spring Coefficient (k)	Mucosa	1000 N/m	1000 N/m
		Vocalis	1000 N/m	1000 N/m
	Mass	Mucosa	0.002 kg	0.002 kg
		Vocalis	0.024 kg	0.003 kg
	Damping Factor (ζ)	Mucosa	0.2	0.5
		Vocalis	0.5	1
Dimensions	Depth	Mucosa	0.06 cm	
		Vocalis	0.1 cm	
	Thickness	Mucosa	0.1 cm	
		Vocalis	0.4 cm	
	Tracheal Area		19.63 cm^2	
Aerodynamics	Air Density (ρ)		0.00114 g/cm^3	
	Fluid Viscocity (μ)		0.000186 g/cm - s	
	Lung Pressure		0 – 1 kPa[101]	
Other	Time Step (Δt)		0.001s	

5.3 The biomechanical model of the aryepiglottic folds

This section is broken into two subsections to describe the layout and properties of the physical model of the aryepiglottic folds (Section 5.3.1) and the aerodynamic conditions that drive it (Section 5.3.2). The biomechanical model of aryepiglottic fold trilling function is likely the first of its kind. Sakakibara (2004a) describe a 2×2 mass self-oscillating model of the ventricular folds (in conjunction with the vocal folds) to synthesize their vibratory pattern and consequent laryngeal flows during ethnic singing styles employing ventricular vibration. Their model, however, does not deal in any part with the behaviour of the aryepiglottic folds, which is central to this discussion. Due to lack of precedence for this model, many of the parameters and

[100] Modal and falsetto registers really represent values along a continuum from minimal to maximal longitudinal tension, primarily proportional to the degree of activity of the cricothyroid and thyroarytenoid muscles.
[101] This number is based on Titze (2006: 261); corresponds to 0-10000 dynes/cm^2.

values chosen in the simulation represent the best guess based on available information. Much of the simulation is modeled after Titze's (1973-1974) Sixteen-mass model (discussed in Sections 2.3.2 and 5.2.1). An example of the indeterminacy of physiological data for the model is measurements of aryepiglottic fold mass. Since no data on the mass of the aryepiglottic folds could be found, the effective mass of the aryepiglottic folds was assumed to be somewhere in the range between the mass of the vocalis muscle/vocal ligament and that of the vocal mucosa, per particle. The same is true for tension, spring coefficients; aerodynamic factors are also assumed to be similar to those of the glottal simulation.

5.3.1 The myoelastic component

Overall, the aryepiglottic folds are interpreted as paired sheets of elastic tissue, which increase in mass from their superior edge to their most inferior regions, particularly in the area of the arytenoid-corniculate cartilages. There is no clear distinction between the aryepiglottic folds proper and the tissue that covers the arytenoids, and, as attested in laryngoscopic observation (e.g. in Iraqi Arabic's allophonic voiceless aryepiglottic trill; Edmondson et al. 2007) the cuneiform-arytenoid complex itself may be subject to vibration during aryepiglottic trilling. The vibrating region of the aryepiglottic folds is assumed to begin, posteriorally, at the region of the colliculus of the arytenoids, where the cuneiform cartilages connect to the arytenoids. The region between the colliculus and the cuneiform cartilages, albeit considerably small, exhibits minor vibrations associated with cuneiform movement under aerodynamic force. The aryepiglottic fold proper starts at the cuneiform cartilage (a junction in the vibrating region of the AE fold) and courses up to the lateral margin of the epiglottis; it is subject to considerably more movement than the tissue associated with the arytenoid complex. The tissue covering the tubercle of the epiglottis also exhibits vibration, in the form of a mucosal wave, but this is not modeled in the simulation. The body of the aryepiglottic folds, largely comprised of the quadrangular membrane and fibres from the thyroepiglottic and aryepiglottic muscles, is also important in the movement of the aryepiglottic folds, as it acts as a sail – picking up the glottal pulse and transmitting it to the aryepiglottic fold edge. In a more extensive simulation, the entire fold surface may be conceived of as a lattice of parallel spring-mass elements, but this is computationally demanding to implement graphically. To simplify the simulation, only the edge of the aryepiglottic fold is modeled visually, while the body of the aryepiglottic fold is represented only abstractly in the

particle system. While a mucosal seal is suspected to form between the aryepiglottic fold and the base of the epiglottis, the increased resistance that this might offer (in the form of coupling between the two tissues) is not prescribed in the simulation. Rather, it is the use of either glottal or aryepiglottal pressure, based on whether the aperture is opened, which determines the frequency of the aryepiglottal pulse. This abstraction of the mucosal seal is embodied in context specific aerodynamic force equations discussed in Section 5.3.2.

The particle system is structured as follows. Each fold is comprised of a series of particles coupled by springs; there are eight particles constituting the edge of the aryepiglottic fold proper, and two between the cuneiform tubercle and the arytenoid colliculus. The body of the fold is comprised of two sets of image particles (duplicates), which are used to add resistance to the oncoming glottal pulse in the event that a mucosal seal is formed, otherwise they behave normally, but are not explicitly rendered in the model (to conserve computer resources). There is an anterior boundary condition defined by the positioning of the epiglottis and a function that computes the approximate location of the surface of the tubercle of the epiglottis (i.e. Eq. 5.4). A schematic of the biomechanical model is presented in Figure 5.6.

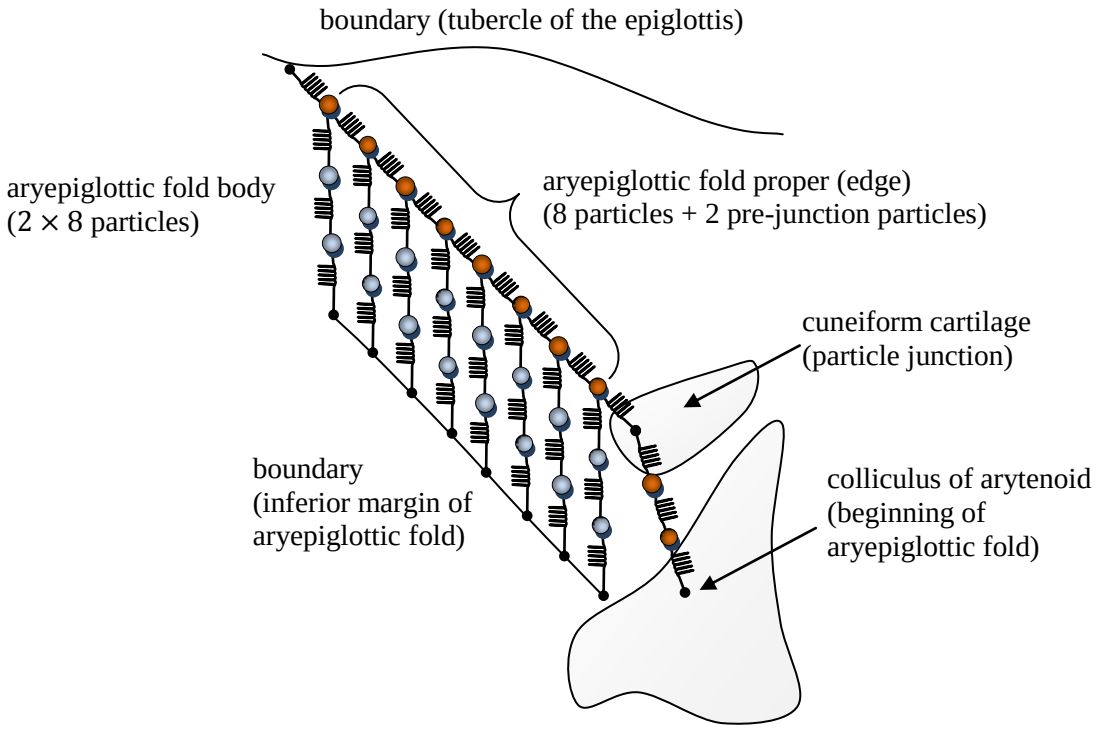


Figure 5.6: Schematic of the biomechanical model of the aryepiglottic fold (left). The orange spheres represent particles that are rendered in the 3D model, while the blue spheres are particles used to simulate the behaviour of the aryepiglottic fold body, particularly under conditions of mucosal seal

formation, but not rendered to conserve computer resources. Location of the body particles is determined by the equilibrium state locations of the edge particles, except for a small decrease in the *y*-coordinate). The diagram is intended for expository purposes and does not reflect the actual dimensions or proportions of the 3D model.

The boundary for the aryepiglottic particles is more complex than the one used in the vocal fold simulation. This is mainly due to the change in direction of the particles at the cuneiform cartilage. Pre-cuneiform junction particles are simply prohibited from crossing $x = 0$ in world space[102] (the model is perfectly centered in world space; the right and left folds fall into their own half of the space). The particles of the aryepiglottic fold proper, on the other hand, have boundaries that roughly trace the epithelial lining of the tubercle of the epiglottis. Like the boundary for the vocal folds, the equation for this boundary was determined through trial and error. The formula (*Eq.* 5.5) was adjusted until the boundary particle locations[103] corresponded well, visually, with the mesh representing the epiglottal epithelium and with images of the epithelium in laryngoscopic pictures. Fundamentally, it is defined as the parametric equation of a line, with modification of the *z*-coordinate to give the boundary a curve in the *x*-*z* plane (Figure 5.7). The boundary allows for movement constraints to be applied, as well as to determine when mucosal seal occurs, i.e. when the aryepiglottic fold contacts the epiglottal tubercle.

Eq. 5.5 Calculating the boundary particle locations of the aryepiglottic fold proper

boudary particle distance:

$$\delta_x = m_x - i_x/n$$
$$\delta_y = m_y - i_y/n$$
$$\delta_z = m_z - i_z/n$$

modification to the z-coordinate:

$$i_z = i_z - 0.02((j \pm 4)^2 + 1)$$

parametric line equation:

$$\boldsymbol{b} = (\boldsymbol{\delta} * j) + \boldsymbol{i}$$

Symbols: [104]
δ = distance between boundary particles
m = mid point of the epiglottis at the tubercle
i = insertion point of the aryepiglottic fold into the epiglottis
b = boundary particle
j = particle index

[102] World space refers to the three-dimensional space and coordinate system where 3D objects can be placed and transformed to construct a 3D scene for rendering by the video hardware.
[103] Visually represented using temporary spheres as markers.
[104] The emboldened symbols are vectors.

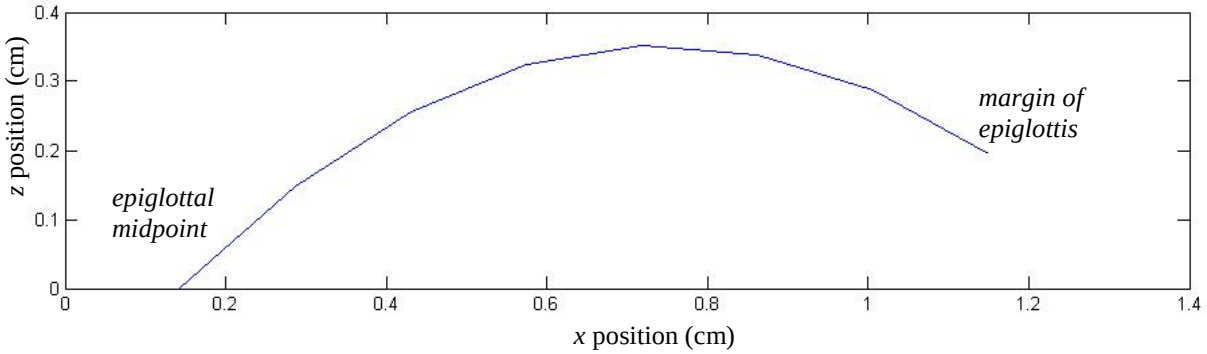


Figure 5.7: Illustration of the aryepiglottic boundary shape in the *x*-*z* plane (right side shown).

5.3.2 The aerodynamic component

The aerodynamic component of the aryepiglottic simulation is an extension of the model of glottal flow, but is reduced in complexity, by comparison, until more refined models avail themselves. The assumption of incompressible flow from Titze's (1973) model is carried forward. The aryepiglottic sphincter is conceptualized as a constriction in the air stream, which modifies volume velocity, and resultant Bernoulli pressure changes accordingly. The distribution of pressure is assumed to be uniform across the entire aryepiglottic fold, including the region from the arytenoid to the cuneiform cartilage. The driving forces of the biomechanical model are specified as force vectors, where the direction is determined by analyzing the motion vector patterns observed in the high-speed laryngoscopic analysis, Section 4.3.3. This technique is used because movement does not simply occur normal to the direction of flow, as is the case in most glottal simulations. The use of force vectors, whose directions are determined from the motion vector analysis, allows for forced particle movement direction to be influenced in an empirically justified way.

Glottal output pressure from the vocal fold simulation[105] is directly applied to the aryepiglottic model as the source. This value is obtained by aggregating the output pressure values for each particle pair, which is then interpreted as the source in the calculation of aryepiglottic volume velocity and aerodynamic driving forces. As there is virtually no research on this subject, the aryepiglottic model is not as refined as the one developed by Titze (1973). The equivalence circuit used to calculate aryepiglottic flow is based on those developed by Flanagan & Landgraf (1968: 59) as well as Titze's (1973: 153). A diagram of the circuit is

[105] Titze (1973) calls this "input pressure" to the vocal tract.

shown in Figure 5.8. The calculation of inductance and resistance, both viscous and turbulent follows directly from the equations for the corresponding terms for the vocalis particles in Titze's model (1973: 153). The overall validity of this model is an area that needs further research, but the model does maintain consistency with the glottal simulation and minimally provides a utilizable volume velocity.

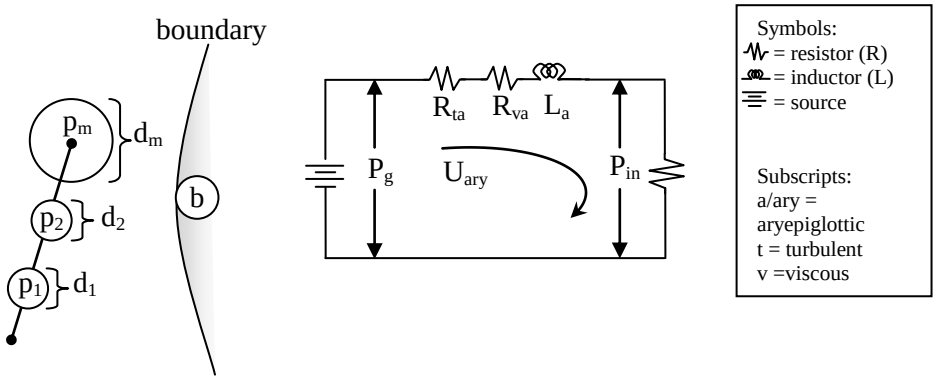


Figure 5.8: Equivalent circuit used to calculate air flow through the aryepiglottic sphincter. The circuit is developed on the assumption that the aryepiglottic sphincter exhibits the same resistance to flow and inductive properties as the vocal folds. The corresponding equations for resistance and inductance are therefore based directly on the equations for the vocal folds. A more complex model is undoubtedly required, but this stands as an issue for future research. The diagram on the left represents the basic dimensions of each particle in the aryepiglottic fold: p_1 is the first particle (at the basal end of the fold), p_2 is the second particle, and p_m is the main particle, whose position is actually rendered. Note that only diameters (d) are used as the specific dimensions for the aryepiglottic folds could not be obtained[106].

The equations based on the equivalent circuit are presented in *Eq*. 5.6. These equations are direct adaptations of those used for the vocal fold simulation and future research will be required to determine a more appropriate form for the equations and constants. Areas are calculated using vectors to simplify the overall process of obtaining the local dimension of the aryepiglottic particle apertures. Aryepiglottic subaperture area is interpreted as the dot product of the unit normal vector and the cross product of the diagonals of the quadrilateral (see Van Gelder 1995) defined by the space enclosed by a particle ($\mathbf{p}_i$), its neighbour ($\mathbf{p}_{i+1}$), its boundary particle ($\mathbf{b}_i$), and its neighbour's particle boundary ($\mathbf{b}_{i+1}$).

[106] This is also the reason why spheres, as opposed to squares in the vocal fold simulation (Section 2.3.2), are used in the representation of aryepiglottic particles.

Eq. 5.6 Equations of aryepiglottic aerodynamics

Area:[107]

$\boldsymbol{d}_1 = \boldsymbol{p}_i - \boldsymbol{b}_{i+1}$

$\boldsymbol{d}_2 = \boldsymbol{p}_{i+1} - \boldsymbol{b}_i$

$A_{ary} = \frac{1}{2}\hat{\boldsymbol{n}} \cdot |\boldsymbol{d}_1 \times \boldsymbol{d}_2|$

Subscripts:
1 = first diagonal
2 = second diagonal
ary = aryepiglottic
i = index
(particle-boundary)
T = total

Resistance:

$R_{v-ary} = 12\mu a^2 \frac{\mathrm{th_{ary}}}{\mathrm{A_{ary}^3}}$

$R_{t-ary} = 0.875 \frac{\mathrm{P_{b-ary}}}{|\mathrm{U}|}$

$R_T = R_{v-ary} + R_{t-ary}$

Dimensions:
d = diagonal
p = particle location
b = boundary
$\hat{\mathbf{n}}$ = unit normal vector of area plane
th = thickness

Inductance:

$L_{ary} = \frac{\mathrm{th_{ary}}}{\mathrm{A_{ary}}}$

Bernoulli Pressures:

$P_{b-ary} = \frac{1}{2}\rho \frac{\mathrm{U}^2}{\mathrm{A_{ary}^2}}$

Quantities:
μ = coefficient of viscosity (0.000186 g/cm - s)
ρ = air density (0.00114 g/cm^3)
U = volume velocity
U̇ = volume acceleration

Pressures:

P_g = determined in glottal sim.

$P_{in} = -(\mathrm{R_T U} + \mathrm{L\dot{U}} - \mathrm{P_{b-ary}} - \mathrm{P_g})$

$P_{ary} = \mathrm{P_g} - \mathrm{P_{in}} - 1.37\mathrm{P_{b-ary}}$

Pressures:
P_g = glottal pressure
P_{ary} = pressure
P_{in} = pressure input to VT (exit pressure)

The equations of aerodynamic driving forces are, once again, modeled after the ones used in Titze (1973). However, because there is no body-cover distinction made, the need to average pressure to determine force is unnecessary. The lack of a body-cover distinction also makes the number of force phases smaller in the loose cuneiform configuration, as the aperture never fully closes in this state. An increase in complexity is made to account for mucosal seal formation in the tight cuneiform configuration. The application of these context dependent forces is illustrated in Figure 5.9 (loose) and Figure 5.10 (tight), with supporting equations in *Eq.* 5.7 and *Eq.* 5.8, respectively.

Eq. 5.7 Magnitude of aerodynamic force: loose condition

$$F = P_{ary}(d_1 + d_2 + d_m)a$$

Subscripts:
1 = first particle of AE body
2 = second particle of AE body
m = main particle of AE fold (rendered)
g = glottal (pressure)

Dimensions:
a = interparticle distance
d = diameter

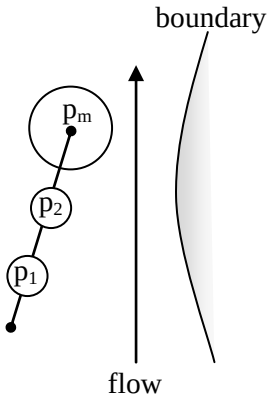


Figure 5.9: Aerodynamic driving forces in the loose cuneiform configuration. As *Eq*. 5.7 describes, the force is determined by summing up the depths of the particles of the aryepiglottic fold body and margin, multiplying by interparticle distance to obtain an area, which can then be multiplied with glottal pressure to obtain the force. In this case it is a magnitude, which is then applied a unit direction vector to determine the movement of the particle (discussed later in this section). To simulate the reduction in exposed area in this state, due to convolutions within the fold, diameters are reduced to radii (done mathematically by multiplying the entire equation by 0.5.

[107] To prevent a non-planar or "bent" polygon from forming, boundary heights are automatically assumed to be those of the particles themselves.

Eq. 5.8 Magnitude of aerodynamic force: tight condition

$F = P_g d_1 a$ $F = P_g (d_1 + d_2) a$ $F = P_g (d_1 + d_2 + d_m) a$ $F = P_{ary} (d_1 + d_2 + d_m) a$

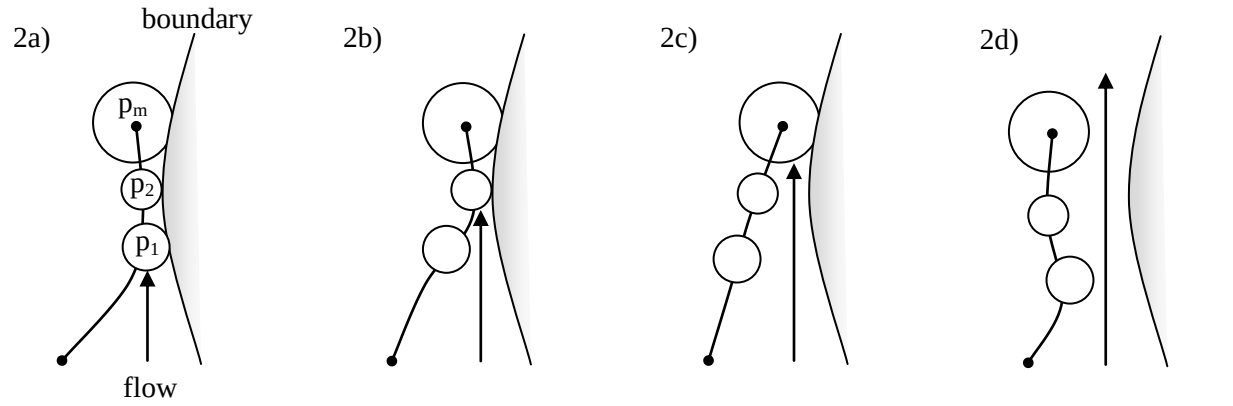


Figure 5.10: Aerodynamic driving forces in the tight cuneiform configuration. Forces are determined by the degree to which the aperture is open; the greater the opening, the more aryepiglottic fold surface area is exposed to the glottal pulse. The diagrams are placed immediately below the corresponding aerodynamic force equations in Eq. 5.8. Full diameter is used in the closed state calculations (2a – 2c), while the open state (2d) behaves identically to the loose cuneiform configuration.

The final aspect to be considered, now that the force magnitudes are known, is in what direction they are applied. Observations from the motion vector analysis are directly incorporated into the specification of the force vector direction (see Section 4.3.3). It is assumed that the direction of expulsive force is the same on either side of the aryepiglottic sphincter (i.e. for either fold). The motion vector analysis reveals that motion of the aryepiglottic fold, in both the loose and tight conditions, occurs in a roughly ellipsoidal pattern (see Section 4.3.3), which indicates posterior movement of the median portion of the fold, with strong oblique-to-lateral components near the lateral most portions. Modeling of this movement pattern is confined to the *x*-*z* plane, as no information about height of fold movement could be ascertained. It is assumed, however, that the fold does get propelled slightly upwards upon receiving the glottal pulse. All vectors are thus given a slight vertical bias. In the motion vector analysis, movement was identified as being ellipsoidal during the burst phase. An elliptical directionality is assigned to the motion vector field that describes the movement of particles in the *x*-*z* plane during glottal pulse transmission. The parameters specifying the semimajor (*a*) and semiminor (*b*) axes are determined by measuring the dimensions of the area of the burst region in the motion vector

analysis frame sequences (shown in Figure 5.11). The averages[108] of the dimensions are reduced as a ratio, 0.94:0.51, which is used to supply the values for the ellipse axis parameters. The major axis runs lateral-medially (a = 0.94), and the minor axis runs antero-posteriorally (b = 0.51). The parametric equation for the ellipsoidal vector field is provided in *Eq*. 5.9. The domain of the equation was restricted to $t = \left[0, \frac{\pi}{2}\right]$ with the x-values for the fold in the negative region of the x-axis negated.

Eq. 5.9 Parametric equation of the ellipse used to determine force vector directions

$$x = 0.94\ \cos t \qquad t = \left[0, \frac{\pi}{2}\right]$$
$$z = 0.51\ \sin t$$

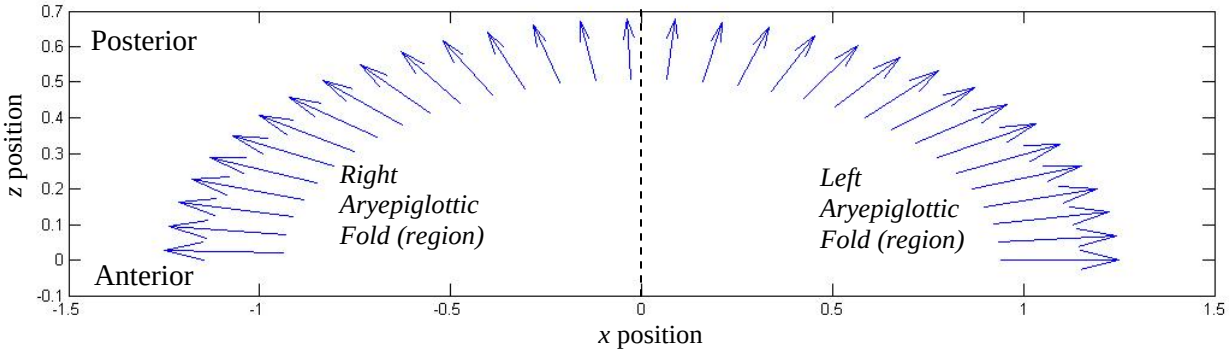


Figure 5.11: Directionality plot of unit force vectors used to determine the direction of aerodynamic force vectors in the aryepiglottic simulation. Once the magnitude of the force is calculated, it is multiplied by the force vector to obtain the actual forcing function for each direction. This plot illustrates both the left and right fold region (using an expanded domain, [0, π]). The midline of the aryepiglottic sphincter falls at exactly 0.

5.4 Summary of the Biomechanical Model of Aryepiglottic Trilling

The purpose of this chapter has been to delineate the biomechanical model used to control the visual synthesis of aryepiglottic trilling. The central objective of this model is to represent the left-right asymmetry that was observed in the high-speed videos in Section 4.3. Based on the observations of the high-speed videos, the hypothesis that trilling asymmetry arises from anatomical asymmetries was formed. This hypothesis specifies that it is cuneiform configuration that gives rise to one of two basic patterns of aryepiglottic trilling. The loose cuneiform configuration, on the other hand, freely oscillates in synchrony with the glottal pulse, which is

[108] Three samples were taken from both sides for a total of six samples of the burst region. Unreduced average dimensions: length = 376 pixels, width = 204 pixels.

relatively unimpeded. Tight cuneiform positioning results in a delayed response of the aryepiglottic fold, as the glottal pulse is assumed to work its way past a mucosal seal between the aryepiglottic fold and the base of the epiglottis. This causes vibration of the aryepiglottic fold to occur at a frequency roughly half of an aryepiglottic fold in the loose configuration.

The biomechanical model represents these possibilities in the form of context sensitive forcing functions that are based on the degree of contact the aryepiglottic fold has with the epiglottis. Three layers of particles are used to represent the body of the aryepiglottic fold, although the upper most layer is the only one that is actually rendered in the 3D model. These particles allow for a graduated contact between the aryepiglottic fold and the epiglottis. In the tight configuration, all particles are in contact and must be sequentially expelled from the epiglottis, causing a time delay in the pulse cycle. Simulation of the loose configuration involves positioning the fold (based on cuneiform location) such that none of the aryepiglottic fold particles make contact with the epiglottis.

The actual performance of the biomechanical model of the aryepiglottic folds will be discussed in Chapter 6, which also evaluates and exhibits the 3D model's graphical user interface, mesh representations of laryngeal structures, basic posturing, synthesis of articulations and phonation types.

Chapter Six

SHOWCASING THE MODEL

6.1 Overview to evaluating the model

As discussed in Chapter 1, the primary focus of the model is the *visual* synthesis of laryngeal and pharyngeal articulations. This chapter represents a culmination of the research and design that has gone into creating the 3D Laryngeal Constrictor Model. All major aspects of the model are considered. As a preliminary, the model's graphical user interface (GUI) is discussed. Section 6.2 provides images of the model's components to illustrate its appearance and establishes basic dimensions of the model. Section 6.3 illustrates basic vocal fold manipulation possibilities, first from a physiological perspective, and then from a linguistic perspective, by examining how the model represents cardinal laryngeal & pharyngeal articulations and phonation types. Finally, section 6.4 discusses the biomechanical simulations implemented in the model by examining dependent variable behaviour.

6.1.1 The model's graphical user interface (GUI)

The graphical user interface of the model is designed to provide the user with a suite of controls to manipulate how the model is viewed and parameters for simulating articulation and laryngeal movements. A screen shot of the GUI is provided in Figure 6.1. The model is presented within a *render space* in the center of the GUI. The camera can be rotated by dragging with the middle mouse button, scaled using the mouse wheel, and translated by holding the SHIFT key and dragging with the left mouse button. By default the lighting cast on the model is programmed to follow the location of the camera; its position can be changed by dragging with the right mouse button. If standard viewing angles are desired, the user can select one from a drop down menu to the left of the viewing space.

Information about the state of the model is gained from the labels in the upper left corner of the render space. The current viewing angle is displayed at the top, followed by percent of available zoom, frames per second (FPS), elapsed simulation frames, and finally selected structures. The FPS count is intended to give the user an indication of the current computer performance; values above 30 FPS will allow the model to be viewed as smoothly as possible.

The elapsed simulation frame count provides a count of the number of times the biomechanical simulations of the vocal folds and/or the aryepiglottic folds have been updated.

Model meshes can be toggled on and off through the use of the *structures* menu-strip item. For paired structures, the user can choose to remove one of the structures. The meshes can be selected by clicking on them with the left mouse button; the program will identify the structure selected by changing it to a green color, and displaying the name of the structure in the upper left hand corner of the screen. To facilitate orientation of model structures in coordinate space, a set of coordinate arrows can be toggled using the A key. The model also features a ruler that can be toggled with the M key and translated in the measuring plane with the arrow keys, and finally will draw the axis of rotation along the cricoarytenoid joint when the S key is pressed.

Flanking the render space are the two parameter control sets. The left set specifies the level of intrinsic laryngeal muscle activity, while the right set controls the parameters modeled after the valves of the throat described in Edmondson & Esling (2005; see Section 2.1.6). To the bottom left of the model is a panel that contains the biomechanical simulation controls. The user can increase subglottal pressure (from 0 to 10000 dynes/cm^2), change the speed of the simulation (placing the slider at its minimum value will pause the simulation), and finally advance the simulation a single frame at a time. Below the valve controls on the right is the articulatory preset panel, which allows the user to select a predefined set of parameter actions that constitute a laryngeal articulation or phonation type, and control the degree of movement towards the target state through the use of a slider.

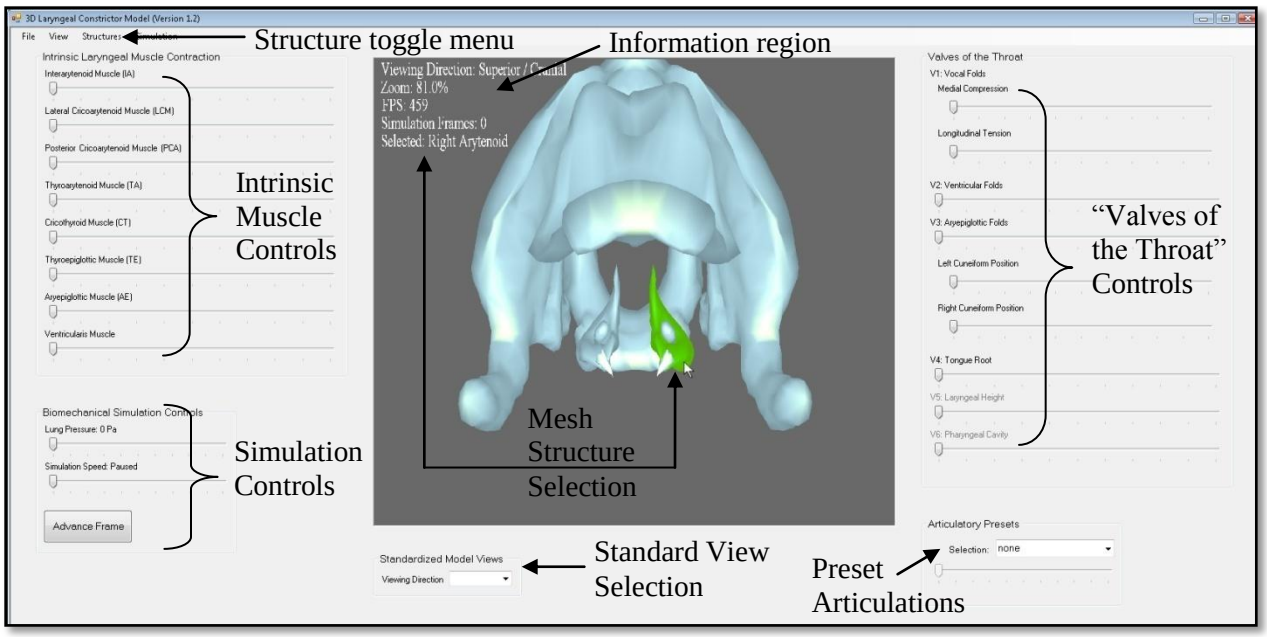


Figure 6.1: The model's graphical user interface (GUI). Areas of interest have been identified.

The purpose of having two model control parameter sets is to underscore the observation that movements of vocal tract structures can be interpreted in terms of physiology, and in terms of articulatory phonetics. It is important to note that, despite the fact that the same fundamental set of structures is manipulated, these parameter sets do not overlap in their control of movement. The physiological parameters are more directly related to individual structures, while the valve controls govern small constellations of muscles along the degree of their parameter activation. In the case of the aryepiglottic folds (V3), the sub-parameters that control cuneiform position, used to simulate tight and loose cuneiform configurations (see Section 5.1.2), do not impact the degree of muscle contraction at all. Figure 6.2 provides a schematic of how these two parameter sets interrelate in the first phase of the model.

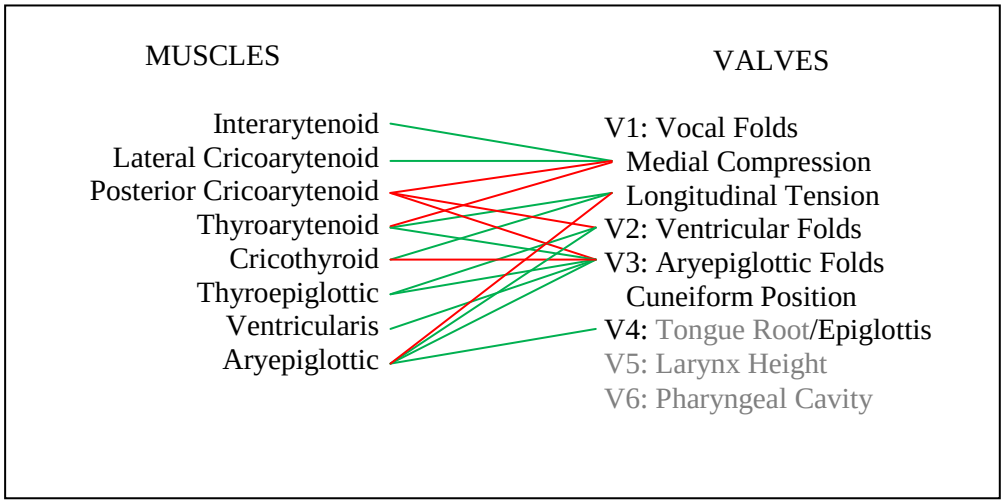


Figure 6.2: Schematic of muscle-valve control relationships. Green lines indicate mutual activation between the parameters, while the red lines indicate that the parameters counteract each other. The tongue root component of V4, V5 (larynx height), and V6 (pharyngeal tube constriction) have been greyed out to indicate that they are not featured in the control sets of the current phase of the model.

6.2 Appearance and measurements

This section provides images of the individual components of the model along with their measurements. The images of the cartilages show five key aspects of each separable model component (i.e. meshes): posterior, anterior, superior, inferior, and lateral. Other tissues are displayed from a handful of angles that provide optimal viewing.

The components are measured using the virtual ruler (briefly mentioned in Section 6.1.1). As described in Section 3.2.2, because the histological images were scaled to conform to measurements projected onto the rendering grid in the 3D mesh development software (Blender), it was possible to impose a metric on the dimensions of the model components. The mapping between the coordinate space used in Blender, and that used in the model rendering software (DirectX 9.0) is 1:1, thus the dimensions are preserved and meaningful. A 3D ruler was created to gauge approximate dimensions of each of the structures for a selection of major anatomical landmarks, particularly those documented in previous research (e.g. Eckel et al. 1993, Titze 2006: Ch. 1).

6.2.1 Cartilages & ligaments

The laryngeal cartilages, and the ligaments that link them together, form the structural and functional foundation of the larynx. The cartilages represented in the model

are the cricoid (Figure 6.3), thyroid (Figure 6.4), epiglottal (Figure 6.5), arytenoid (Figure 6.6), corniculate (Figure 6.7), and cuneiform (Figure 6.8). Measurements taken in this section are summarized in Table 6.1 and compared to averages of corresponding laryngeal dimensions reported in Eckel et al. (1993) and Titze (2006). Of particular interest is the relationship between the cuneiform and arytenoid cartilages: in the model the cuneiforms are seen suspended directly anterior to the arytenoid, with the head of the cuneiform rising slightly above the apex of the arytenoid (Figure 6.9).

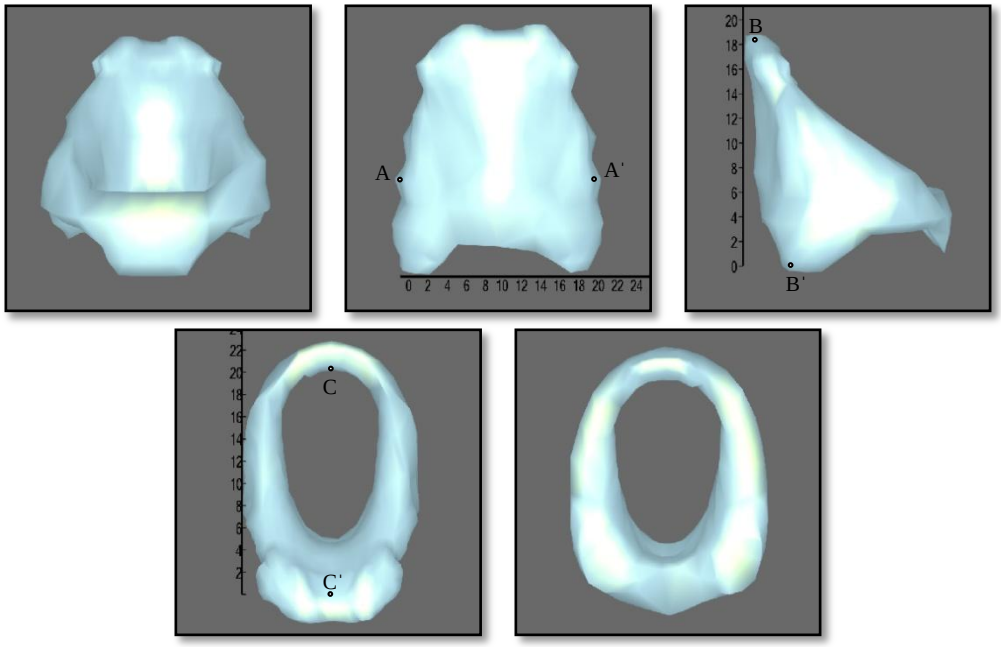


Figure 6.3: The cricoid cartilage as it appears in the model. Views from left to right, top to bottom: anterior, posterior, lateral (right), superior, and inferior. Posterior wall width (A-A′): ~20mm; posterior wall height (B-B′): ~19mm; inside diameter, anterior-posterior, at cranial edge (C-C′): ~21mm.

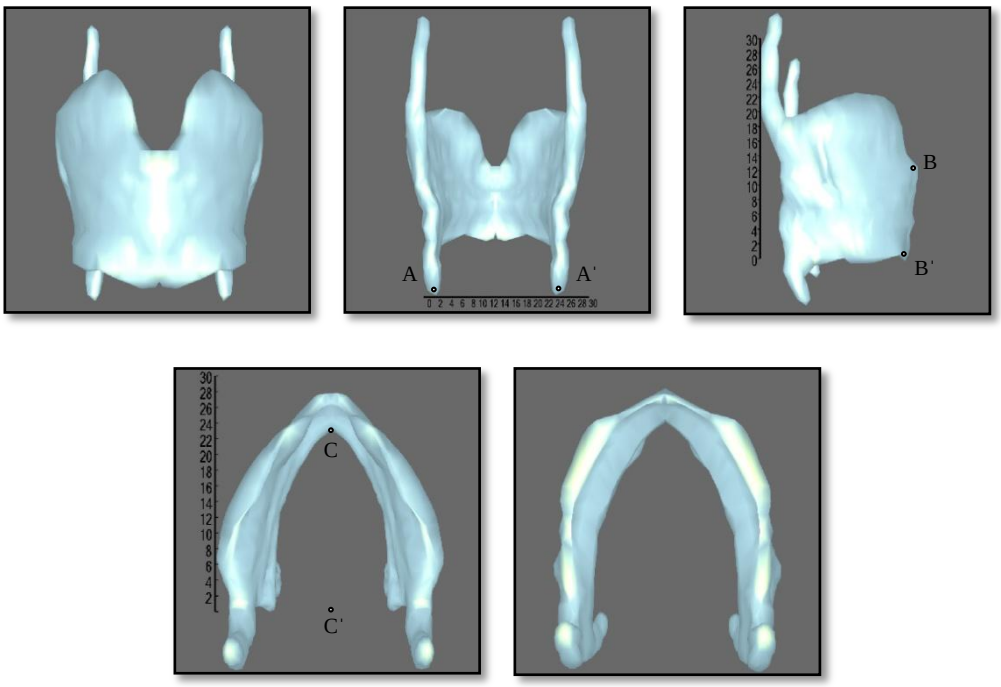


Figure 6.4: The thyroid cartilage as it appears in the model. Views from left to right, top to bottom: anterior, posterior, lateral (right), superior, and inferior. Inter-inferior cornu distance (A-A'): ~23mm; anterior thyroid notch to inferior border (B-B'): ~13mm; bottom inside, anterior-posterior (C-C'): ~24mm.

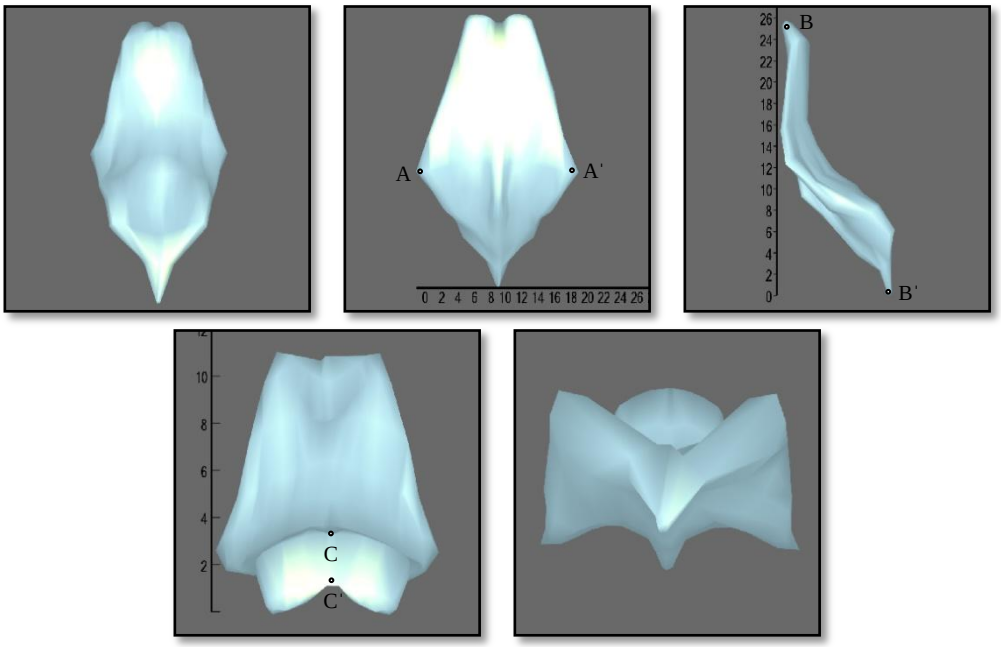


Figure 6.5: The epiglottal cartilage as it appears in the model. Views from left to right, top to bottom: anterior, posterior, lateral (right), superior, and inferior. Widest portion (A-A'):~19mm; petiolus to apex (B-B'): ~26mm; thickness at apex (C-C'): ~2mm.

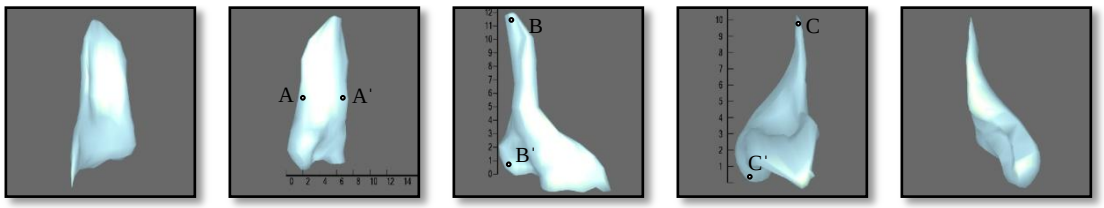


Figure 6.6: The arytenoid cartilage as it appears in the model. Views from left to right: anterior, posterior, lateral (right), superior, and inferior. Posterior girth (A-A'): ~6mm; apex to muscular process (B-B'): ~12mm; vocal process to muscular process (C-C'): ~12mm.

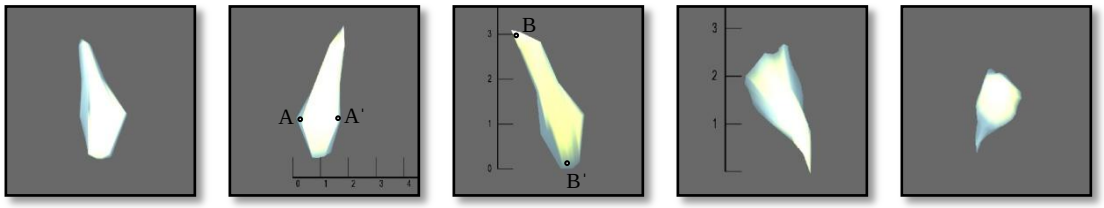


Figure 6.7: The corniculate cartilage as it appears in the model. Views from left to right: anterior, posterior, lateral (right), superior, and inferior. Width (A-A'): ~2mm; height (B-B'): ~3.5mm.

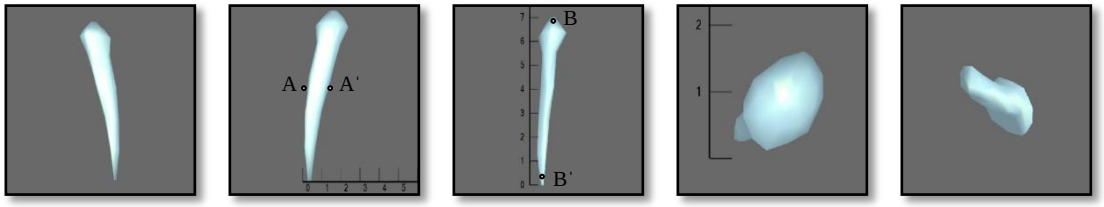


Figure 6.8: The cuneiform cartilage as it appears in the model. Views from left to right: anterior, posterior, lateral (right), superior, and inferior. Medial girth (A-A'): ~2mm; height (B-B'): ~7mm.

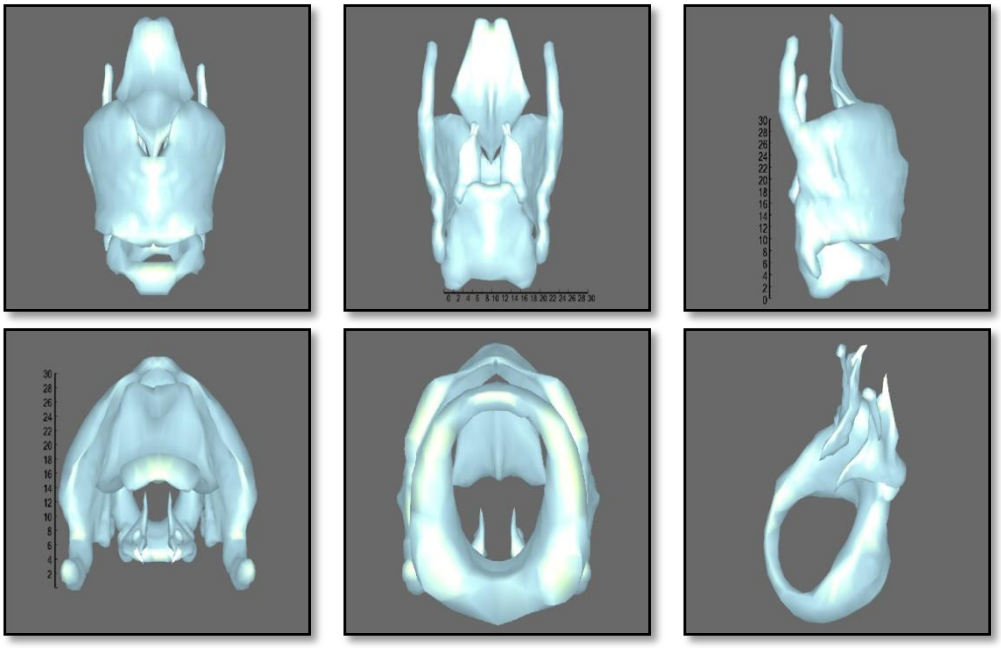


Figure 6.9: The laryngeal cartilages as they appear in the model. Views from left to right, top to bottom: anterior, posterior, lateral (right), superior, inferior, and oblique. Measurements are in mm. The oblique view (antero-superio-lateral, left side) shows the seating of the cuneiform, corniculate, and arytenoid cartilages relative to the cricoid cartilage.

Table 6.1 Measurements of laryngeal cartilage dimensions in the model

Measurement		Model	Titze 2006	Other
C	posterior wall width (A-A′)	~20	21.76 (7; 1.82)	24.40 (2.45)*
	posterior wall height (B-B′)	~19	19.63 (3; 1.65)	19.80 (1.20)
	inside diameter (C-C′)	~21	23.66 (3; 0.23)	17.40 (1.92)*
T	Inter-inferior cornu distance (A-A′)	~23	25.44 (2; 2.46)	29.90 (3.54)*
	anterior thyroid notch to inferior border (B-B′)	~13	13.59 (2; 0.13)	15.80 (1.07)*
	bottom inside, anterior-posterior (C-C′)	~24	22.90 (6; 2.98)	-
E	widest portion (A-A′)	~19	-	19.9 (2.35)*
	petiolus to apex (B-B′)	~26	-	28.2 (2.62)*
	thickness at apex (C-C′)	~2	-	-
A	posterior girth (A-A′)	~6	-	-
	apex to muscular process (B-B′)	~12	14.05 (3; 1.51)	13.95 (2.47)*
	vocal process to muscular process (C-C′)	~12	11.05 (3; 0.23)	12.5 (2.05)* [109]
κ	width (A-A′)	~2	-	-
	height (B-B′)	~3.5	-	-
Δ	medial girth (A-A′)	~1	-	0.5 – 1**
	height (B-B′)	~7	-	5 – 9**

N.B.: C = cricoid, T = thyroid, E = epiglottis, A = arytenoid, κ = corniculate, Δ = cuneiform. All measurements are in mm. For Titze (2006) measurements, sample size and standard deviations appear in parentheses, in that order. For Eckel et al. (2004) measurements, only standard deviation is shown; sample size is always $n = 25$. All values are for female larynges.

*Eckel et al. (1994)

**Krmpotic (1957)

The laryngeal ligaments that have been rendered in the model include the ventricular, vocal, ceratocricoid, (posterior) cricoarytenoid, and thyroepiglottic ligaments. The ceratocricoid and the cricoarytenoid ligaments form the joint capsule for the cricothyroid and cricoarytenoid joints, respectively. The ceratocricoid has posterior, anterior, and lateral components, while the cricoarytenoid ligament is most massive on the posterior side of the joint. The ventricular ligament is rendered as connecting roughly to the colliculus of the arytenoid and running alongside the cuneiform cartilage. It is possible that the ventricular ligament exhibits more substantial attachment to the cuneiform in some

[109] Eckle et al. (1994) present separate left-right measurements for paired structures; this value represents an average of the two measurements.

individuals, but the degree of morphological variation is high from individual to individual (Crevier-Buchman, personal communication). Images in Figure 6.10 illustrate the five laryngeal ligaments that appear in the model. The median cricothyroid ligament is illustrated in Section 6.2.3, which discusses the membranes represented in the model.

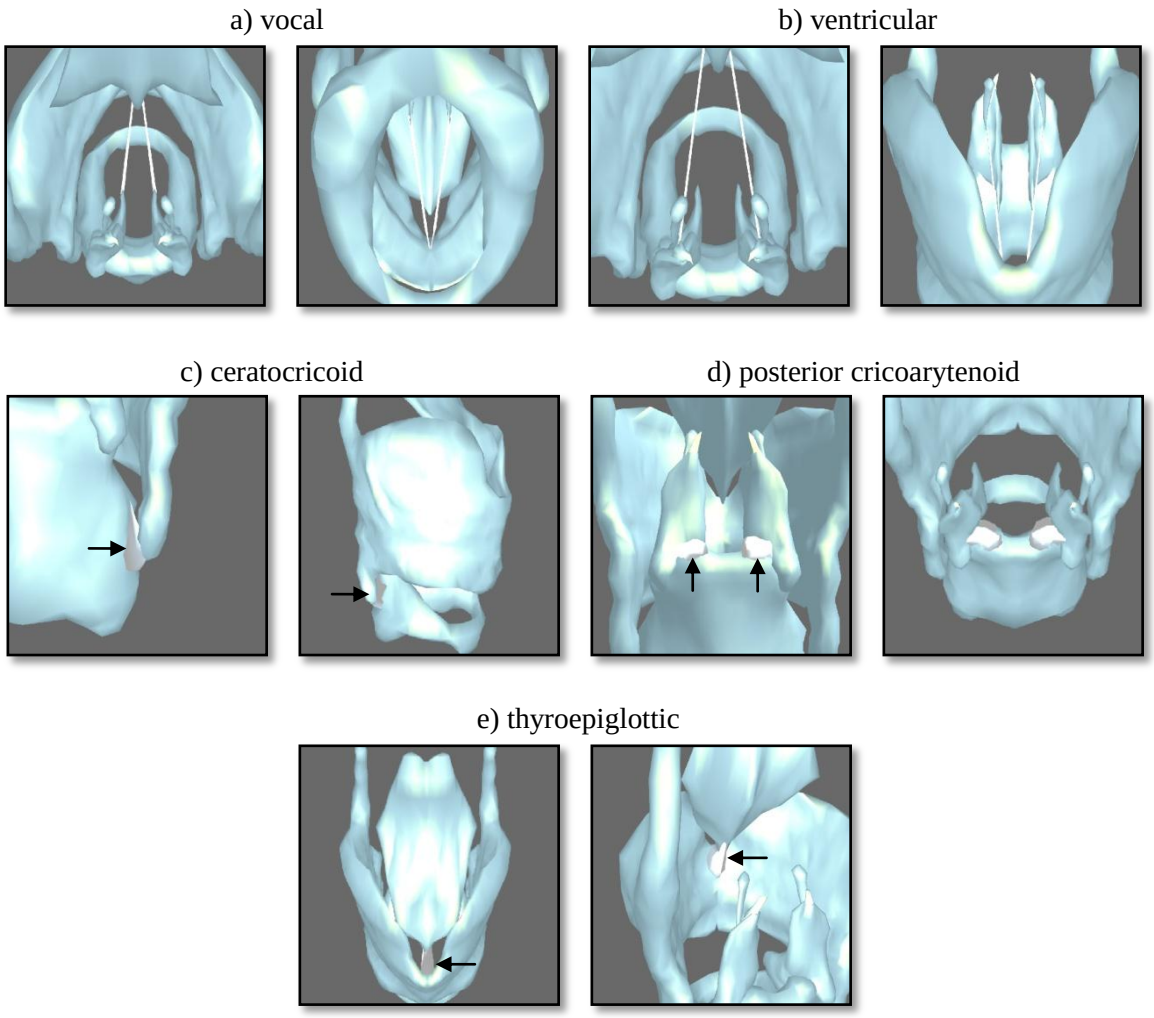


Figure 6.10: The laryngeal ligaments as they appear in the model. Note that the thyroid cartilage and epiglottis are removed in some screen shots to maximize the visibility of the structure, while maintaining the overall cartilaginous framework for context. Ligaments appear in grey/white. The small arrows in (c-d) are used to highlight the ligament being illustrated. The right image in (d) shows the posterior cricoarytenoid ligaments in their abducted position, where they can be viewed more clearly. From left to right, top to bottom, the viewing aspects are as follows: (a) superior, inferior, (b) superior, anterior, (c) posterior, antero-lateral, (d) posterior, supero-posterior, (e) supero-anterior, postero-lateral.

6.2.2 Muscles

Phase One of the 3D Laryngeal Constrictor Model development exclusively represents the intrinsic muscles of the larynx. The cricothyroid, posterior cricoarytenoid, and lateral cricoarytenoid muscles are rendered as individual meshes, while the interarytenoid is rendered as connected to the aryepiglottic muscle (via the oblique arytenoid projection), and the thyroarytenoid is rendered as continuous with the thyroepiglottic and aryepiglottic muscles. The primary motivation behind these fused muscle meshes is that the structures were not amenable to individual segmentation as they appeared in the histological data. Despite this practical difficulty, the fused representations demonstrate the considerable interdigitation that these muscles exhibit, which may be interpreted as a benefit to the model. Separable meshes are desirable to allow the user to isolate muscle sections to clearly visually distinguish the muscle, but this will require more thorough data to represent. Concerning the aryepiglottic muscle, in Section 2.2.4 the research of Van Daele, Perlman & Cassell (1995) was briefly discussed, which indicated that the aryepiglottic muscle is not ubiquitous in the human population. This study is evidently troubling for the model as it throws into doubt whether the muscle should be represented, if the goal is to have idealized anatomy. Numerous other researchers, however, attest that they exist, although no other study could be found that provides an indication of their frequency of occurrence (see Section 2.2.4). As the model is concerned with the aryepiglottic folds, it seems reasonable to include the muscle to demonstrate its appearance when it does occur. Ultimately, the issue is a matter of cosmetics as the muscle is evidently not well developed and not the sole muscle responsible for retracting the epiglottis. Future versions of the model may allow the user to optionally display the muscle in addition to the thyroarytenoid and thyroepiglottic muscles.

Figure 6.11 provides illustrations of the muscles from optimal viewing angles, and for every muscle except the thyroarytenoid, demonstrates their relaxed and contracted configurations. The thyroarytenoid along with the aryepiglottic and thyroepiglottic muscles are more thoroughly illustrated in Figure 6.14, which depicts the vocal fold posture settings (as discussed in Section 3.3.1). Finally, as discussed in Section 2.2.5, the ventricularis

muscle is not explicitly rendered in the model due to its diminutive size, and its apparent absence from the histological data set (Hirano & Sato 1993).

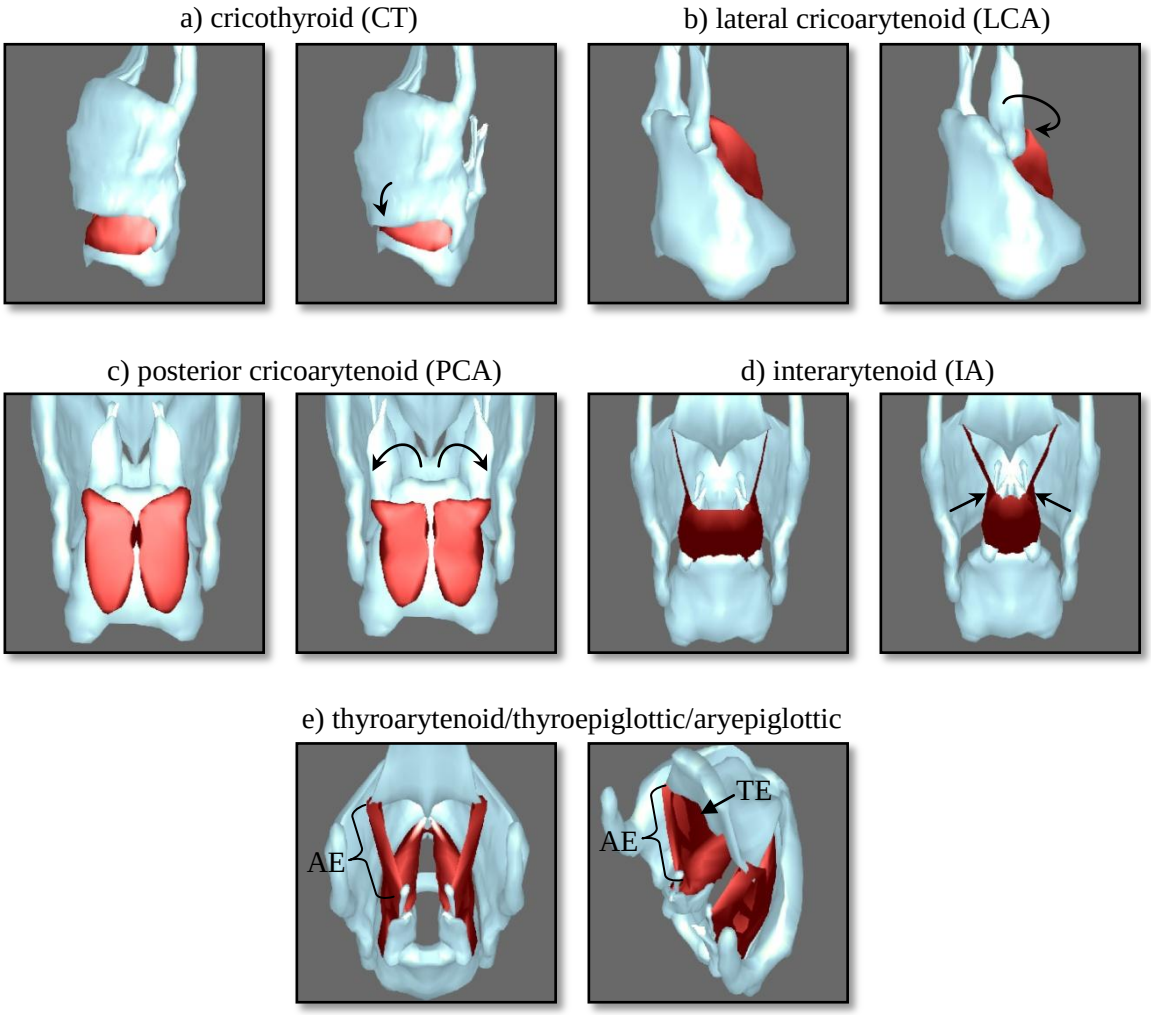


Figure 6.11: The intrinsic laryngeal muscles as they appear in the model. Viewing angles from left to right, top to bottom:(a) lateral (left), (b) postero-lateral (right), (c) posterior, (d) posterior, (e) postero-superior, postero-latero-superior (right). In (a-d) muscle action is indicated by an arrow head. In (e) no movement is illustrated; the aryepiglottic muscles are denoted by AE, and the thyroepiglottic muscles are denoted by TE.

6.2.3 Membranes

Currently, the model only features two membranes, both of which are important to the laryngeal constrictor mechanism. The conus elasticus forms the internal wall of the laryngeal tube from the level of the cricoid to the vocal folds themselves, where it terminates as the thickened portion known as the vocal ligaments (not rendered as part of the conus elasticus). The regions where the conus elasticus blends into the vocal processes

of the arytenoids posteriorally, and the internal angle of the thyroid notch anteriorally, are known as the maculae flavae (yellow spots). In the model, these regions appear as thickened bulbs on the superior margin of the conus elasticus. The cricothyroid ligament is rendered as continuous with the conus elasticus, but is easily distinguished by the sharp change in direction where the ligament meets the conus elasticus proper. The model also features the quadrangular membrane, which gives shape to the aryepiglottic folds and the epilaryngeal tube/laryngeal vestibule. Worthy of note is that the inferior margin of the quadrangular membrane is the ventricular ligament, which means that the aryepiglottic sphincter is inherently connected to the ventricular one[110]. Figure 6.12 illustrates these two membranes as represented in the model.

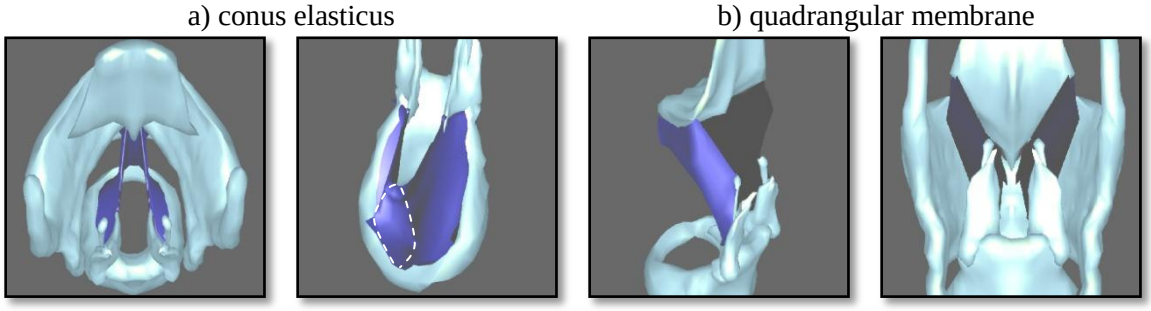


Figure 6.12: The laryngeal membranes as they appear in the model. The membranes appear in dark blue. The white dashed line in the right image of (a) indicates the location of the cricothyroid ligament, which is rendered as continuous with the conus elasticus. Views from left to right: (a) superior, antero-lateral, (b) postero-lateral, posterior.

6.2.4 Epithelial tissues

The present phase of the model pays specific attention to the epilaryngeal tube. As discussed in Section 2.2.5, The surfaces covering the pharynx proper and the tongue root were not attended to in this phase, but are intended to be included as the appropriate data avails itself. There are four meshes used to represent the epithelial tissues of the larynx: vocal, ventricular, aryepiglottic, and epiglottal. The vocal mucosa extends from the internal margin of the cricoid cartilage up to the level of the ventricle. It does not cover the arytenoids. The ventircular epithelial tissue forms the ventricular fold and provides an indication of the location of the

[110] This anatomical fact is an essential component of the overall mechanism of constriction, which involves a "folding and buckling" of the epilaryngeal tube and associated structures (Esling, personal communication). The consequence of constriction at the aryepiglottic level is understood to effect medial displacement of the ventricular folds.

ventricle (which cannot be viewed without placing the model camera within the interior of the mesh itself). The aryepiglottic epithelial tissue covers the major portion of the posterior face of the cricoid cartilage, the anterior linings of the piriform fossae, and the arytenoid/corniculate and cuneiform cartilages. The aryepiglottic folds project upwards towards the widest portion of the epiglottis and wrap around the body of the epiglottis to indicate their attachment there. Finally, to provide a visual indication of the exposed surface of the epiglottis, a mucosal covering is rendered. Due to lack of complete data, only the internal surface, which forms part of the epilaryngeal tube, is rendered. Images of these meshes are provided in Figure 6.13.

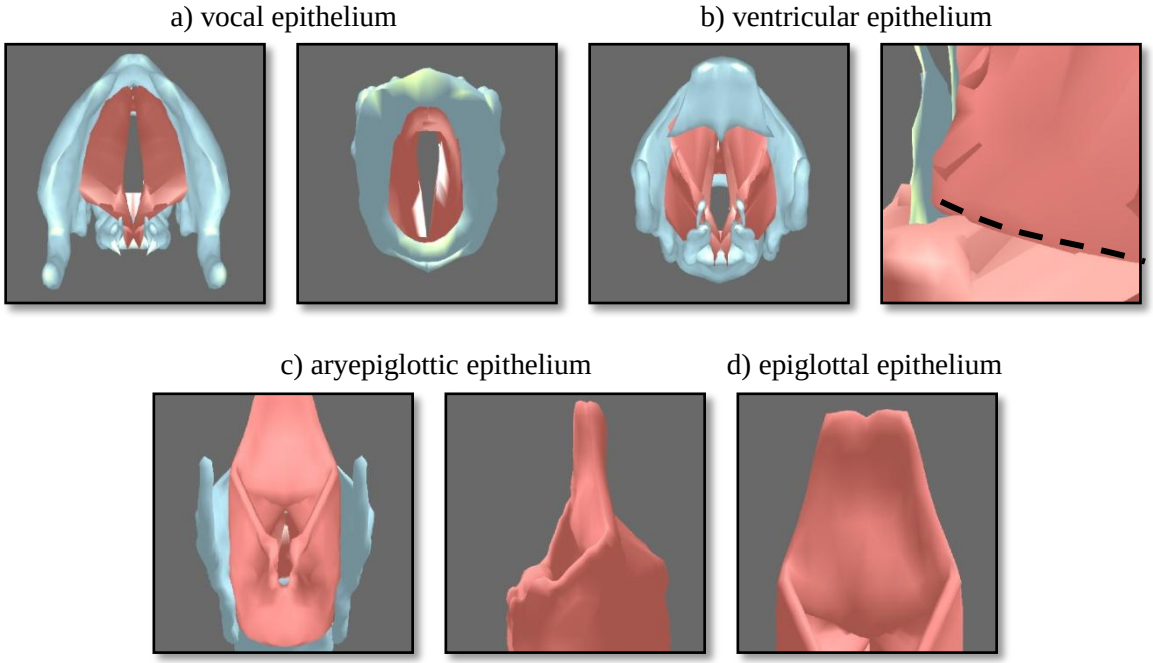


Figure 6.13: The epithelial tissues of the epilaryngeal tube and glottal region as they appear in the model. The epithelial tissues appear in light pink. Images in (a) show the vocal mucosa from a superior (left) and inferior (right) view; note that the epiglottis has been removed for better viewing. The ventricular epithelium is seen in (b) from a superior (left) and internal (right) views. A dotted line in (b) is used to highlight the location of the ventricle, which lies beyond the ventricular fold. The aryepiglottic folds are seen with the addition of the aryepiglottic epithelium in (c). Supero-posterior (left) and right lateral (right) views are provided. Finally, the epiglottal epithelium is displayed in (d), blending into the aryepiglottic folds.

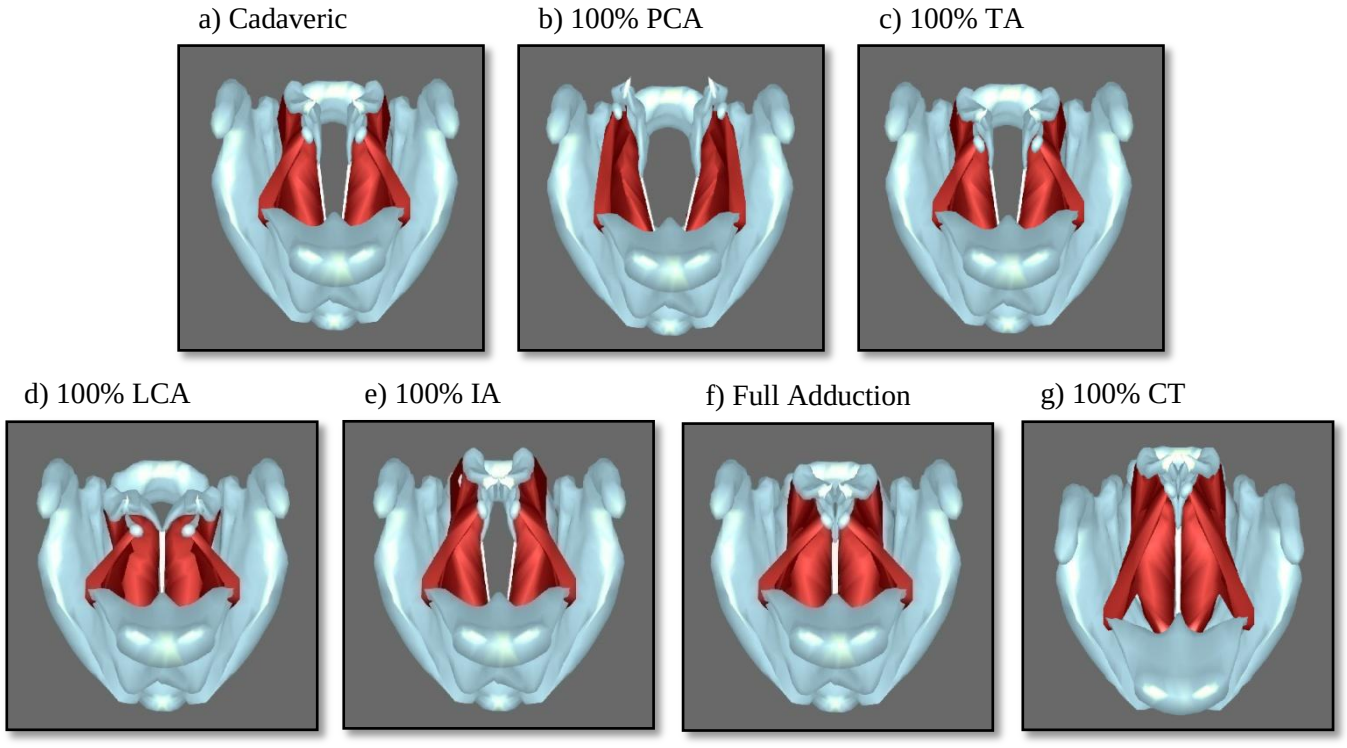
6.3 Movement

This section represents the culmination of research into laryngeal movements discussed in Chapters 2, 3, and 4. Primary vocal fold posturing of the model is exhibited in Section 6.3.1,

while cardinal linguistic movements are examined in Section 6.3.2, for segmental articulations, and Section 6.3.3, for phonation types.

6.3.1 Basic vocal fold posturing

The vocal folds are integral to characterizing laryngeal constrictor states as they are directly responsible for establishing the fundamental distinction of voicing contrast, in addition to more subtle degrees of manipulation of phonation type and pitch control. The purpose of this section is to illustrate the extreme postures of the vocal folds represented in the model, in addition to their default, cadaveric position. The effect of the lateral cricoarytenoids (LCA), interarytenoids (IA), posterior cricoarytenoids (PCA), and thyroarytenoids (TA) is demonstrated by means of isolated and complete contraction of these muscles. Complete adduction with the aid of both the lateral cricoarytenoids and the interarytenoids is illustrated along with the additional action of the cricothyroid (CT) muscles to demonstrate longitudinal tension in a phonatory context. Although only associated with aryepiglottic function, movement of the epiglottis is included here as well. All of these states are presented in Figure 6.14, where only the thyroarytenoid, thyroepiglottic, and aryepiglottic muscles are rendered along with the vocal ligaments for full visual clarity.



h) epiglottis rotation (60.7°)

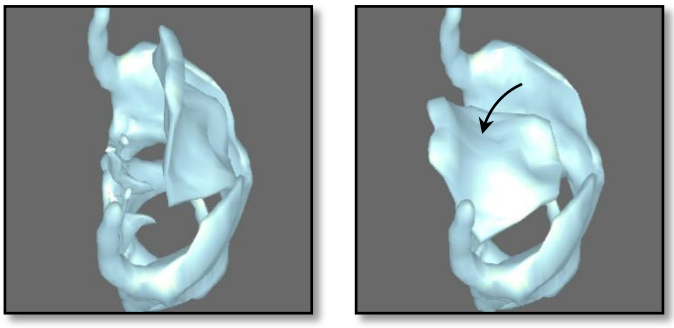


Figure 6.14: Illustration of vocal fold posturing in the model. These images can be compared with the diagrams in Figure 3.9. The red meshes represent the thyroarytenoid and aryepiglottic muscles, while the white mesh represents the vocal ligament. Contraction resulting in full adduction (representing 50% LCA and 90% IA, with 20% TA activity) is shown in (f), while (g) represents the same state with the addition of 100% CT contraction. Image (h) displays the range of epiglottis movement available in the model, totalling 60.7°, which is slightly below the value discussed in Section 3.3.3.

6.3.2 Segmental articulations

The laryngeal constrictor mechanism is identified as both a source and an articulator, in a manner similar to the vocal folds. Indeed, the articulatory action of the vocal folds is inextricably linked to the constrictor mechanism, which facilitates vocal fold damping during glottal stop, [ʔ], through massing the ventricular folds on top of the vocal folds, a manoeuvre described as ventricular incursion (see Section 2.1.4). Glottal fricative, [h], is the least constricted of all of the articulations, with considerably dilated glottal and epilaryngeal lumena. The class of sounds known as pharyngeals is reinterpreted as aryepiglotto-epiglottals primarily, with potential narrowing of the pharynx by dint of tongue root retraction. However, only the primary constriction is represented in this phase of the model. The aryepiglotto-epiglottal place is continuous, leading to complete closure of the epilaryngeal tube, [ʡ], which is identified as a well attested pharyngeal articulation in many of the world's languages (e.g. Esling 1996). The voiceless aryepiglotto-epiglottal fricative (voiceless pharyngeal fricative), [ħ], shares much in common with whispery voice, as evidenced by laryngoscopic videos. The interpretation is that the articulation relies upon reduced activity of the interarytenoid muscles to create a posterior glottal gap, allowing for the markedly increased passage of air. Cuneiform configuration is still moderately adducted as necessitated by the need for constriction of the aryepiglottic folds, which is partially facilitated by lateral cricoarytenoid contraction. The aryepiglotto-epiglottal approximant, [ʕ], represents intermediate contraction of the muscles responsible for aryepiglottic

stricture, and appears as a slightly patent version of aryepiglotto-epiglottal stop. These articulations are illustrated in Figure 6.15. Voiced aryepiglottic trilling is examined in Section 6.4.2.

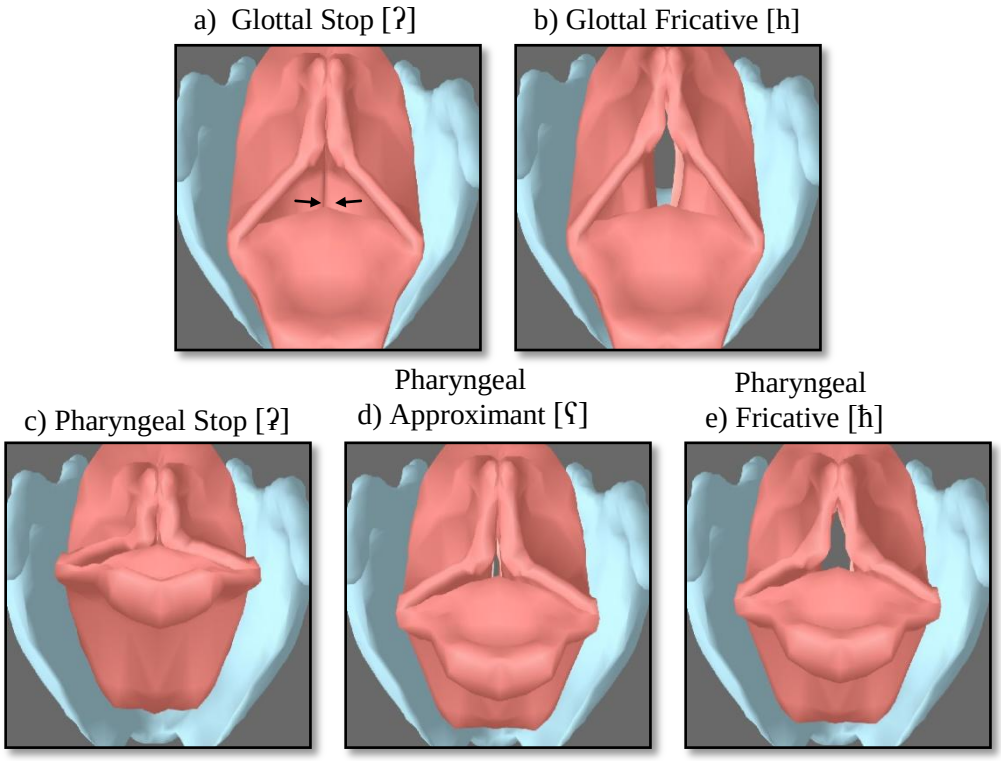


Figure 6.15: Illustration of simulated segmental articulation. In (a) glottal stop illustrates the contribution of the ventricular folds (arrows; V2 in the valve model) through adduction to dampen the movement of the vocal folds. (b) shows glottal fricative, which notably involves very little contribution from the superior valves of the throat (in this case V2 and V3, the ventricular folds and the aryepiglottic folds). Image (c) demonstrates complete closure of the epilaryngeal tube at the aryepiglotto-epiglottal place, while (d) represents variable degrees of aryepiglotto-epiglottal stricture, combined with vocal fold vibration, to produce variations of pharyngeal approximant. Finally, (e) demonstrates fricative production in the larynx with the addition of aryepiglottal stricture. Note that as laryngeal height (V5) is not yet a feature of the model, this has been ignored in the production of the pharyngeal articulations. For all of the pharyngeals, with aryepiglottal stricture, there is a strong tendency for the larynx to raise, although this is not a necessary condition for pharyngeals to be produced.

The measurements obtained in the geometric analysis (discussed in Section 4.2) are the basis for all of the articulations presented in Figure 6.15 (as well as the phonation types in the next section). Muscle activity levels chosen to synthesize the articulations are shown in Figure 6.16. These are the values specified for each of the preset articulations at 100% stricture level (i.e. when the preset slider is placed at its maximum value). Thus, the bar

graph can be interpreted as ratio of contraction (e.g. when the preset slider is controlling glottal stop, the interarytenoids are activated at twice the rate of the lateral cricoarytenoids).

The model was initially set in the open, deep inspiration configuration, which was used as the reference state for determining the parameter values characteristic of each articulatory state. The basic muscle movement parameters were adjusted until the resulting model configuration, determined by the geometric features, reflected the values obtained in the original analysis. To illustrate this, glottal stop is compared with deep inspiration in Figure 6.16. A complete set of the synthetic geometric analysis values is presented in Appendix E.

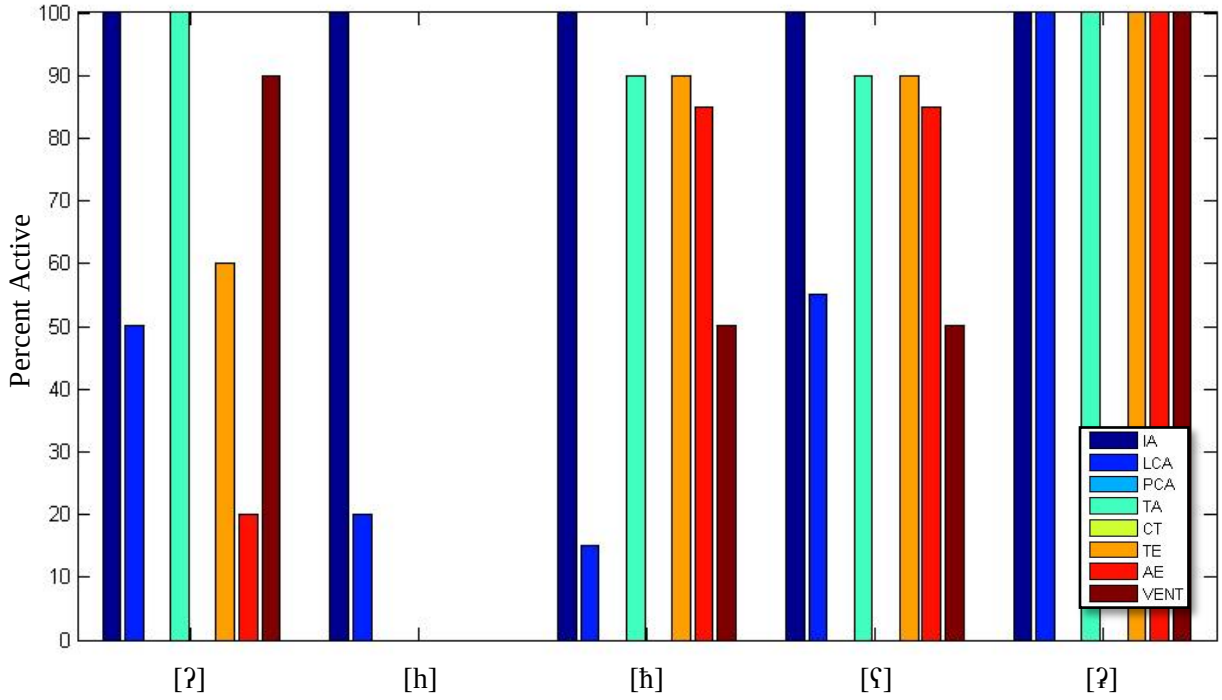


Figure 6.16: Muscle activation levels used in the model to synthesize the segmental articulations (presets).

Table 6.2 Comparison between synthetic and real articulation: geometric features

Type	State	CoT	CuT	C-ET Left	C-ET Right	IA-ET	VP	AE area	VentF area	VF area
Real	Inspiration	100% (87)	100% (126)	100% (127)	100% (130)	100% (143)	100% (62)	100% (18756)	100% (6609)	100% (3431)
	Glottal Stop	14% (13)	50% (64)	33% (43)	57% (75)	48% (69)	0% (0*)	24% (4678)	0% (0*)	0% (0*)
Synthesized	**Inspiration**	100% (38)	100% (66)	100% (150)	100% (150)	100% (183)	100% (83)	100% (20130)	100% (11112)	100% (6177)
	Glottal Stop	0% (0)	22% (15)	51% (77)	51% (77)	66% (121)	0% (0)	29% (6000)	0% (0)	0% (0)

Note: F0 = fundamental frequency (Hz); CoT = distance between the corniculate tubercles; CuT = distance between the cuneiform tubercles; C-ET = cuneiform-epiglottic tubercle distance; IA-ET = interarytenoid notch-epiglottic tubercle distance; VP = distance between the vocal processes of the arytenoid cartilages; AE area = area enclosed by the aryepiglottic folds and base of the epiglottis; VentF area = area enclosed by the ventricular folds; VF area = area of the glottis; LH = larynx height (descriptive: N = normal height; L = lowered; R = raised). Values marked with a <*> have been estimated, while values marked by a <?> were deemed too difficult to assess using this method.

The absolute values of the measurements (in units of pixels or pixels^2) are not relevant to the analysis as these values reflect image scaling differences between laryngoscopy and screen shots of the program. The important values are the percentages, which express the relationships between the various landmarks. Effort was made to make the synthetic articulations correspond as much as possible to what was observed laryngoscopically, and this is largely reflected in the values obtained through image analysis of these articulations. To further compare the degree of accuracy between the real and synthetic articulations, a measure of average stricture is taken, which is simply the average of all percentage values. This measure provides an overall characterization of the degree of the stricture, while conflating individual stricture measurements (i.e. landmark distances and areas). Table 6.3 presents these measurements for all of the articulations.

Table 6.3 Average stricture measurement for segmental articulations (real vs. synthetic)

Articulation	Real	Synthetic
Inspiration	100%	100%
Glottal Stop[ʔ]	25.1%	24.3%
Glottal Fricative [h]	54%	39.8%
AE Stop [ʡ]	16.7%	2.7%
AE Fricative [ħ]	36.2%	20.5%
AE Approximant [ʕ]	11.7%	7.11%

If the data are interpreted to be a reasonable approximation of the overall constriction in the larynx during the articulation, synthetic or real, then the patterns match overall. There is, however, a tendency for the model values to overshoot the degree of constriction observed in real life, as their values tend to be lower. This is particularly notable for the pharyngeals, where larynx height plays a major role in the articulation, which leads to skewed values in the laryngoscopic observation (i.e. they are larger than expected). For these sounds, the synthetic articulations are on average 44% of the real values, suggesting greater constriction, but more likely reflective of image scaling in the laryngoscopic videos. Synthetic glottal fricative strays the most from the real value, indicating that the synthesis is overly constricted for this articulation. Another potential problem is with the reference measurement of deep inspiration. The lower synthetic values may be a reflex of a simulated deep inspiration that is simply too wide open, causing every other measurement to be small by comparison. Figure 6.17 shows the simulation of deep inspiration.

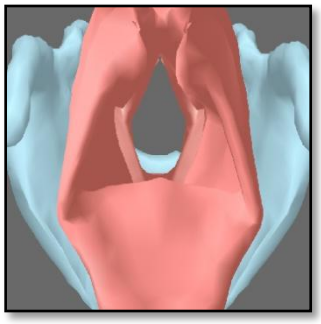
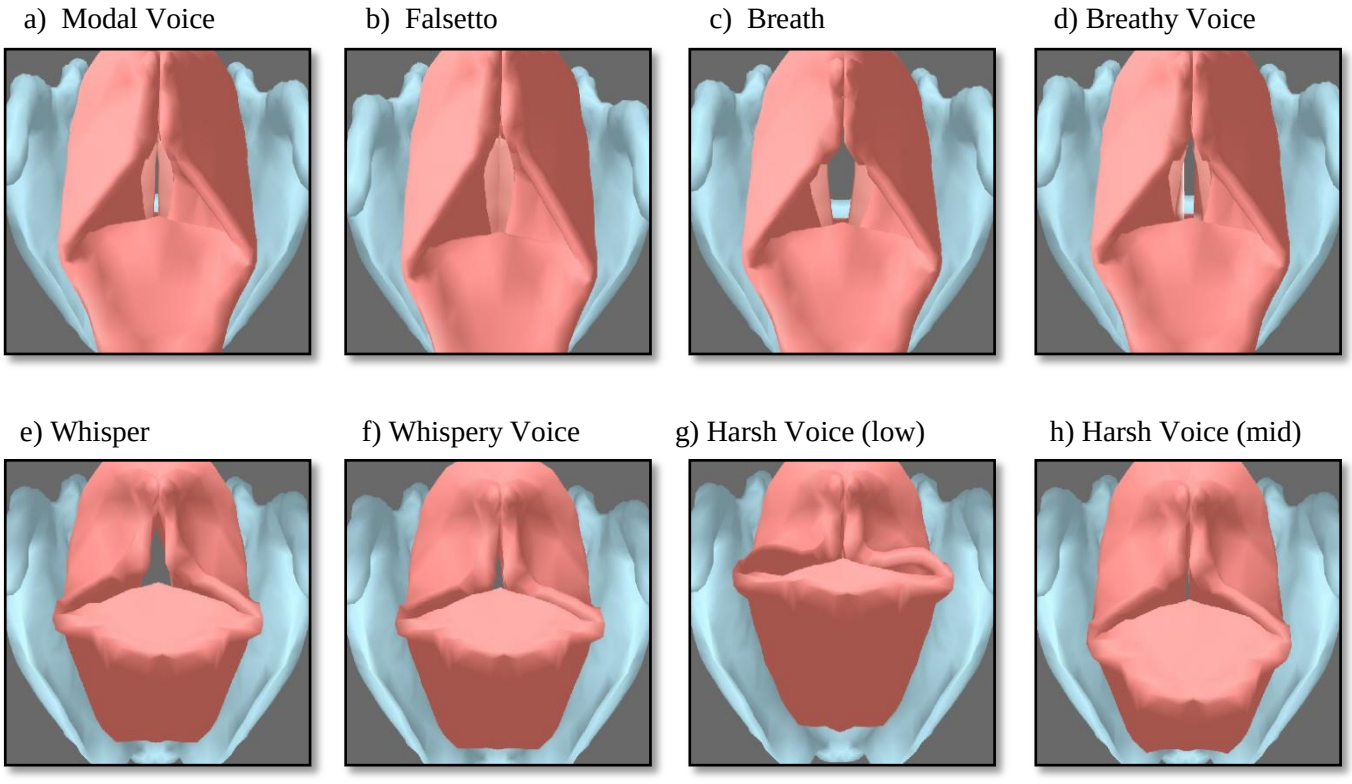


Figure 6.17: Simulated deep inspiration.

6.3.3 Phonation types

The laryngeal constrictor mechanism is identified as a critical element in many phonation types, both linguistically contrastive ones and those used for paralinguistic effect. This section also includes voiceless sources such as whisper, breath, and creak. Phonation types broadly fall into two categories in terms of the activity of the laryngeal constrictor mechanism. The unconstricted, or open category, involves generating phonation, or noisy airflow in the voiceless case, with little to no engagement of the muscles that tense and sphincter the aryepiglottic and ventricular folds; lowered or neutral larynx height is also characteristic of unconstricted sounds. Modal voice, falsetto, breath, and breathy voice are categorized as unconstricted phonation types.

The other category draws on constriction of the aryepiglottic and ventricular folds to increase turbulence in the airflow upstream of the glottis, impeding vocal fold function and generating jitter in the period of vocal fold vibration, epitomized by harsh voice. Perhaps one of the most interesting phonatory states is harsh voice at high pitch, which represents an articulatory conflict between the cricothyroid muscle serving to broaden and open the laryngeal airway, while the aryepiglottic folds and epiglottis are simultaneously working to close it off. The result is a unique[111] constricted state where the vocal folds are elongated and under considerable tension to produce high pitch, but at the same time the ventricular folds are pressed upon the vocal folds and are considerably adducted due to both high thyroarytenoid activity and aryepiglottal constriction. Whisper, whispery voice, creak, creaky voice, harsh voice at low, mid, and high pitch belong to the constricted category. Figure 6.18 illustrates all of the phonatory states represented in the model. Trilling of the aryepiglottic folds in harsh voice at lower pitches is presented in Section 6.4.2. Muscle activation levels used to synthesize the phonation types are presented in Figure 6.19.



[111] Although it shares much in common with glottal stop, distinguished primarily by increased airflow to trigger voicing (Esling, personal communication).

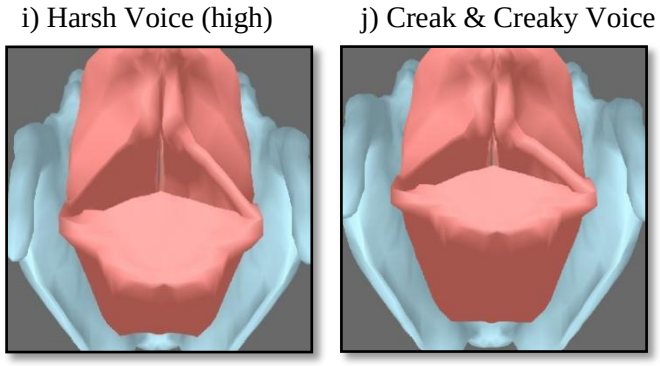


Figure 6.18: Illustration of simulated phonatory states. The top row (a-d) demonstrate unconstricted laryngeal states where aryepiglottic stricture is minimal. The bottom two rows of images (e-i) show constricted state with aryepiglotto-epiglottal approximation. Notably, the aryepiglottic folds tend to form a 180° angle during constricted settings, while they appear to have a 90° angle in unconstricted settings. The major exception to this case is seen in harsh voice at high pitch where the cricothyroid counteracts aryepiglotto-epiglottal approximation. The ventricular folds are seen to be highly adducted during this articulation as a result of aryepiglottal tension. Except for voiceless "phonation" types, vibration of the vocal folds should be assumed to be occurring.

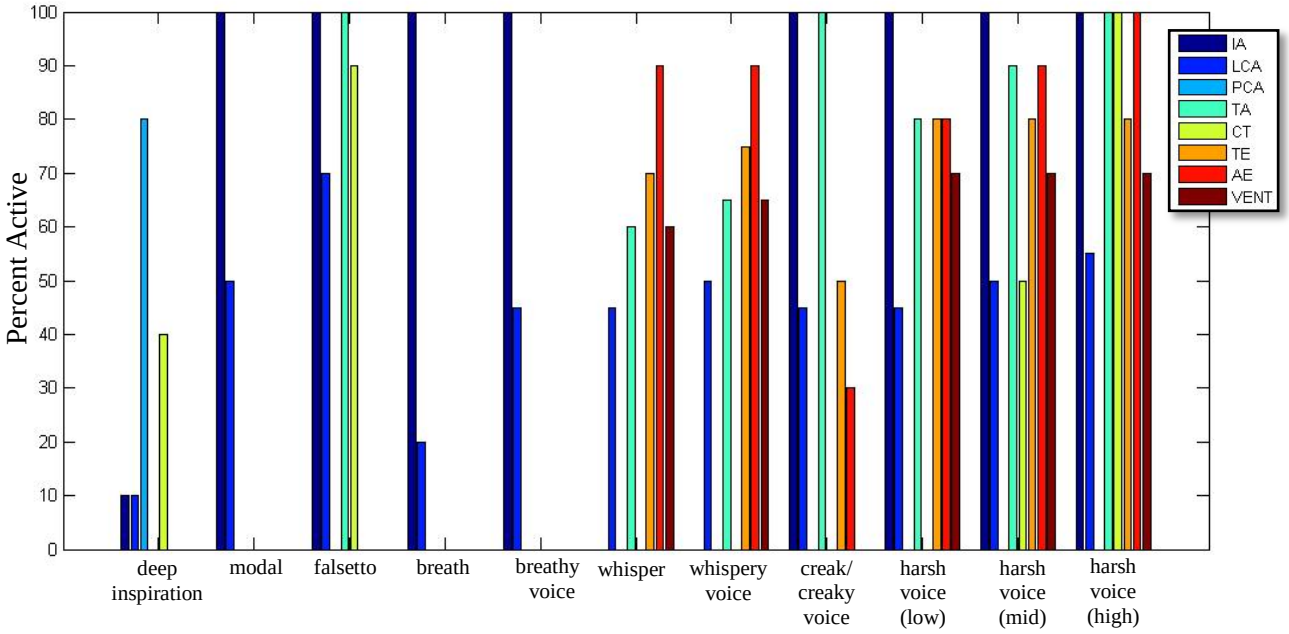


Figure 6.19: Muscle activation levels used in the model to synthesize the phonation types (presets).

Table 6.4 Average stricture measurement for phonation types (real vs. synthetic)

Phonation Type		Real	Synthetic
Unconstricted	Inspiration	100%	100%
	Modal Voice	47.2%	30.6%
	Falsetto	58.5%	40%
	Breath[112]	54%	44.4%
	Breathy Voice	38.4%	35.1%
Constricted	Whisper	36.2%	24.6%
	Whispery Voice	33.3%	12.5%
	Creak/Creaky Voice	20.3%	16.4%
	Harsh Voice (Low)	11.7%	7.4%
	Harsh Voice (Mid)	20.6%	15.8%
	Harsh Voice (High)	24.2%	24.6%

Table 6.4 uses the same technique discussed in Table 6.3 to obtain insight into constriction patterns across the entire set of phonation types. Once again, average stricture shows that the synthesized gestures are overly constricted compared with their natural counterparts. On average, the synthetic unconstricted states are 76.2% less open, while the synthetic constricted states are 71.4% less open. There are a number of mismatches in the overall constriction pattern for the phonatory states. Modal voice should be less constricted than breathy voice, and falsetto must be the most unconstricted phonation type of all, according to the laryngoscopic observations; however, the simulated versions are not consistent with this. It is possible that the error may be attributable to the fact that breathy and modal voice both involve minimal CT activity by default in the model. Setting breathy voice to a lower CT activity level than modal would serve to simulate the reduction of the valve areas (vocal, ventricular, and aryepiglottic) that is inherent when pitch is lowered, as is typical of breathy voice. The synthetic breathy voice would, consequently, be more open due to the increase in vocal fold area obtained when the posterior glottis is opened for breathy airflow. Also problematic is synthetic whispery voice; it exhibits double the amount of constriction, where the real value indicates that only a moderate 9% increase in constriction is required. Apart from these erroneous aspects, the rest of the patterns seem to be roughly consistent across the other phonation types, with the general tendency, again, for the simulated values to be lower (i.e. more constricted). Regardless, the

[112] The values used for breath and harsh voice (low pitch) are the same as those for glottal fricative and voiced pharyngeal (AE) approximant, respectively; this is on the grounds that the gestures used are essentially identical for these sounds.

values should be taken as suggestive rather than rigorously empirical because of the difficulty inherent in comparing the two disparate sources of data.

6.4 Biomechanical simulation

The final model feature to be discussed is the biomechanical simulation of vocal fold, and aryepiglottic fold movement. Three aspects will be discussed for each simulation: aperture area over time, volume velocity over time, and the visual appearance of the model tissue as it is driven by the physical model. Aperture area is an approximation of the area enclosed by the fold and its contact point; the midline in the case of the vocal fold simulation, and the epiglottis in the aryepiglottic fold simulation. For the vocal folds, the left and right fold areas are added together to obtain the total glottal area[113]. Volume velocity demonstrates the simulation of aerodynamic factors in response to changing glottal area and airway resistance and inertance/inductance. In Titze's (1973) model, the Bernoulli equations have volume velocity dependency, as well as many of the other equations involved in the overall system. Finally, the 3D model is intended to be a visual synthesizer of articulation and hence images of the biomechanical simulations taking place during characteristic movement phases are provided.

6.4.1 Vocal fold vibration

The vocal fold simulation is derived from Titze's (1973) Sixteen-mass model of the vocal folds, with some adjustments to allow for the simulation to be interactive. Apart from controlling subglottal pressure, the user can also manipulate longitudinal tension of the vocal folds, which changes the effective mass of the vibrating portion of the fold and its elastic properties. The range of values for tension and mass with respect to cricothyroid activity were provided in Table 5.1. The simulation does not incorporate medial compression as a parameter for modifying vocal fold function, but this is a possible area of development in future phases of the 3D model. Breathy and whispery phonation may both be visually simulated by engaging the biomechanical simulation of the vocal folds while the model is set to display the synthesized breathy and whispery phonation types.

[113] As the model only tracks the area between a given fold and the midline.

Phonation at 100 Hz is demonstrated in Figure 6.20 alongside a plot of the volume velocity. Glottal area ranges from fully closed, to ~0.2 cm^2 at its maximum. The pulse occurs regularly at 10 ms intervals with a rapid opening phase, followed by a slightly less rapid closing cycle. This pattern should not be confused with the measurement of laryngeal impedance via electroglottography; EGG waveforms of modal voice are characterized by having a gradual or two part opening gesture (Esling (1978: 228; Painter 1988: 119-120). A possible interpretation is that the impedance contour of the opening phase is initially slow due to the vertical phase difference of glottal opening: the impedance gradually decreases as the inferior portion opens, and then rapidly decreases with the opening of the superior portion. The glottal area of the model reflects the state of both the inferior portion (the body) and the superior portion (the cover) of the vocal folds and ultimately the contour obtained is the region of overlap in the two area contours. Glottal flow appears to have a delayed rise to its maximal value of ~400 ml/s. This effect is due to vocal tract inertance[114], which reduces the amount of flow and slows its acceleration (Titze 2001, 2006). Overall, the appearance of these plots corresponds well with those presented in Titze (2006: 260), where glottal flow is modeled in detail.

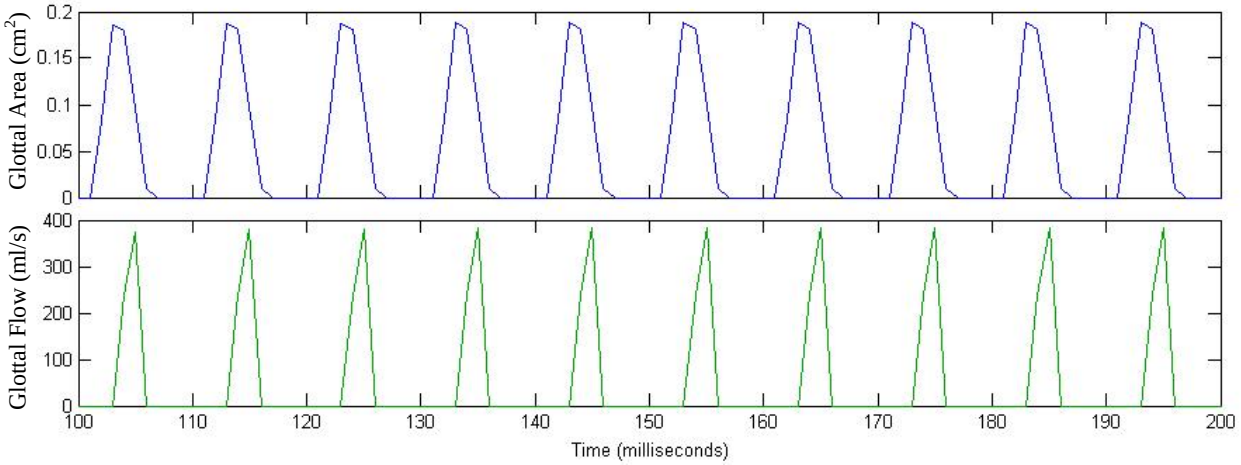


Figure 6.20: Simulated vocal fold vibration at 100 Hz (blue) with corresponding glottal flow (green).

The model is reasonably flexible at simulating various fundamental frequencies, including falsetto voice, with the appropriate adjustment to model parameters (see Table 5.1). Due to the rapidity of the vibration in this state, it is necessary to reduce the time step of the

[114] Due to the analogy with an electric circuit, Titze (1973) uses the term inductance to describe this property of the vocal tract, although describes it as inertance in later publications (Titze 2001, 2006). Inertance is defined as the property of a column of air that describes its resistance to change in velocity see (Titze 2001).

simulation (from 1 ms to 0.1 ms), to prevent the loss of large portions of each glottal cycle. The effect this has is to slow down the appearance of the vibration, despite its actual frequency being considerably higher. The stability of the vibration is also impacted, with variation in amplitude of each pulse, determined by glottal area, appearing to be somewhat random. Figure 6.21 presents simulation of phonation at ~300 Hz.

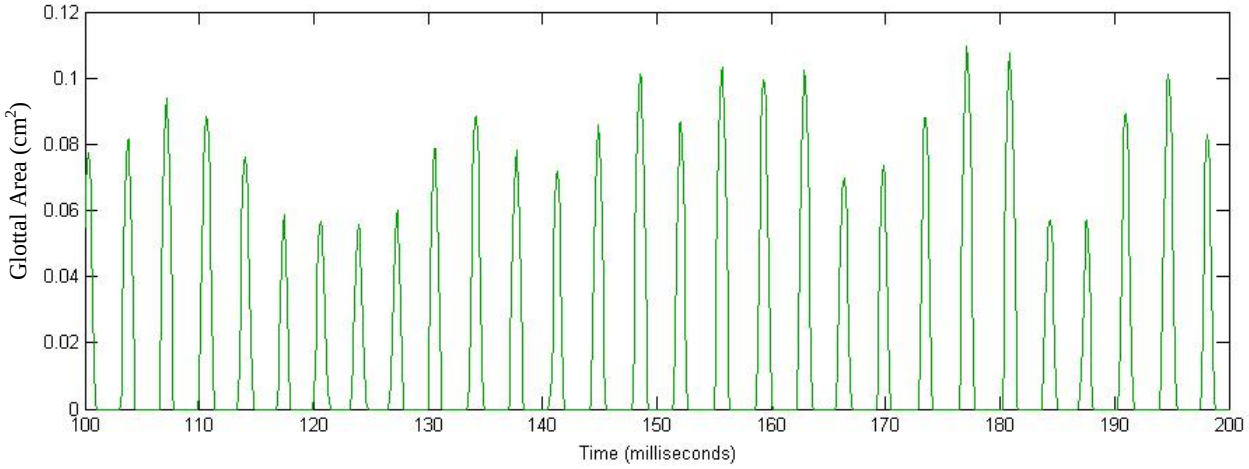


Figure 6.21: Simulated falsetto (~300 Hz).

The vertical phase difference that characterizes vocal fold vibration is also simulated in the model and is illustrated in coronal section in Figure 6.22. The appearance of the vocal folds during phonation is demonstrated in a separate sequence in Figure 6.23.

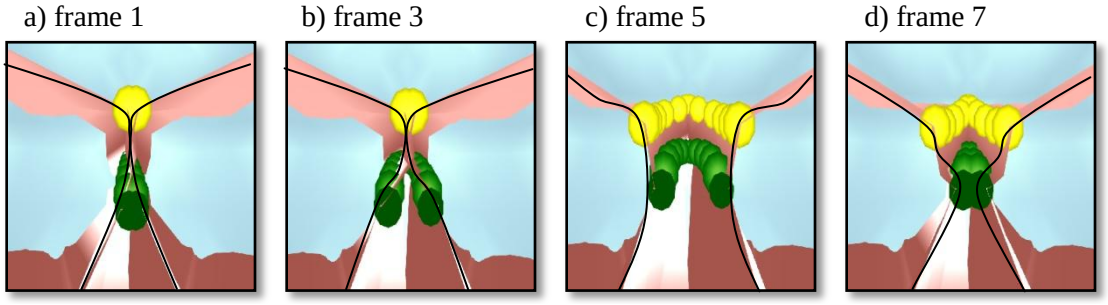


Figure 6.22: An illustration of the simulated vertical phase difference of the vocal folds during modal phonation at 100 Hz. The acquisition of these screen shots involved positioning the camera so that the internal volume of the mucosa of the vocal folds was visible (in an anterior view). To help clarify what the image shows, the particle locations for the body and mucosa of the vocal folds are rendered as green and yellow spheres, respectively. The mesh itself was traced to provide an approximation of the contour it forms in the coronal plane as the particles are displaced. Each image is two frames apart, while the entire pulse lasts approximately 10 frames (i.e. 10 ms). Frame (a) shows full closure, while frame (b) shows opening at the vocalis body. Image (c) shows both particle layers at their point of maximal opening. Finally (d) shows the closure of the vocalis particles, while those of the mucosa are still returning towards the midline. Note that in (d) the mesh does not close entirely, although the particles of the body (green) are in fact touching, indicating full closure. This is likely due to a translation correspondence error between the mesh and the particles.

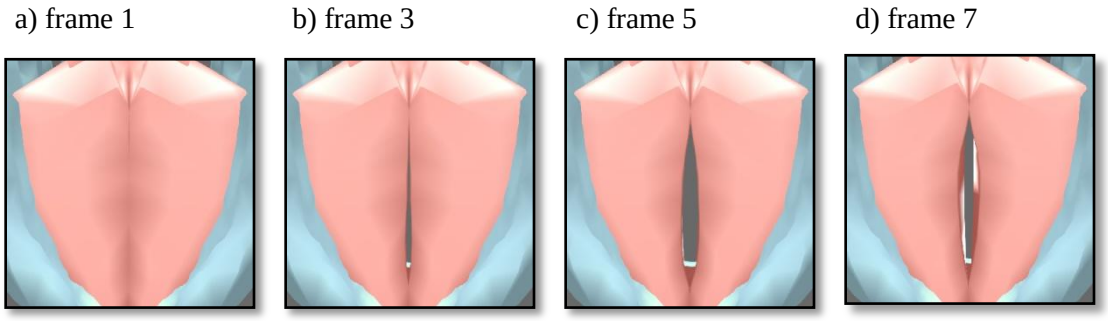


Figure 6.23: Modal phonation as it appears in the model from a superior view. The epiglottis and associated structures have been removed to facilitate viewing of the vocal mucosa. The sequence reflects the glottal states demonstrated in Figure 6.22. The vertical phase difference is most evident in image (d), where the darker region with highlights along the internal edge of the fold is the part of the vocal epithelium associated with the vocalis particles. The edges flanking this internal portion are the slightly dilatory mucosa particles on their return to the midline. There is also an emergent horizontal phase difference evident in (b) where the anterior glottis is exhibiting greater aperture earlier than the posterior portion.

6.4.2 Aryepiglottic fold vibration

The final aspect for consideration is the simulation of aryepiglottic fold vibration. Symmetrical oscillation of the right and left folds is a prediction of the theoretical model developed in Section 5.1, but far more interesting is the condition where the folds move asymmetrically. The physiological explanation for asymmetrical trilling, also discussed in Section 5.1, was that the default cuneiform configuration would give rise to differential contact between the aryepiglottic folds and the tubercle of the epiglottis. This was observed in the high-speed laryngoscopic videos, discussed in Section 4.3, and specifically it was the subject's right fold that made complete contact, while his left fold was patent throughout the trilling sequence. For the sake of uniformity, the present simulation seeks to synthesize the very conditions that gave rise to this individual pattern of trilling, which happened to have a tight right and loose left cuneiform configurations.

The 3D model allows for cuneiform position to be adjusted on the fly, allowing for a tight configuration to form on one end of the continuum, and a loose configuration on the other end. The loose position generally involves cuneiform movement in a latero-posterior direction. The inclusion of these cuneiform position parameters are for the purpose of experimenting with different epilaryngeal configurations, but are not intended to imply that an individual can wilfully manipulate the position of these cartilages. It is suspected that the default cuneiform position during laryngeal constriction is unchangeable in actual larynges. Table 6.5 presents the

parameters used to generate asymmetrical trilling of the aryepiglottic folds, where the right fold is set to a tight configuration and the left fold is in a loose one. Note that the biomechanical parameters for the spring coefficient and damping factors vary between tight and loose configurations, which facilitated producing trilling such that the tight aryepiglottic fold oscillated at $f_0/2$, while the loose fold shared the same frequency as the glottal pulse. Mass of the aryepiglottic fold is expected to increase near the basal portions of the fold; however, the simulation seems to function optimally when the lower particles of the fold are actually lighter (by a factor of 10) than the edge particles. This may indicate an empirical problem for the model, despite its ability to replicate the different trilling patterns. Ultimately, the empirical adequacy of all of these values is subject to future research and the current values are only intended to be expository of how the simulation functions.

Figure 6.24 provides a series of plots that document the aryepiglottic aperture areas and corresponding aryepiglottic volume velocities. The values of these measurements are unattested in actual subjects, and should be regarded, rather, as an index of the simulation conditions. Glottal area is shown at the top for temporal reference. As the green contour indicates, the right aryepiglottic fold is oscillating at half the frequency of the left (red contour) and the glottal pulse itself. Interestingly, the left fold (loose configuration) exhibits cycle-to-cycle variation in amplitude, an effect that is emergent, but parallel to the 'double beating' observed in the high-speed videos, where successive pulses of the left aryepiglottic fold show differing intensities. While these are desirable results, the overall effect is highly periodic, which, as the aperture analysis indicates (see Section 4.3.2), is not entirely the case: the system contains far more jitter and chaos. The actual aerodynamic conditions undoubtedly are more complex. Volume velocity shows a slow acceleration at the onset of aperture formation, as in the glottal simulation, which reflects the inertive properties expressed in the aerodynamic equations. Consequently, the volume velocity appears to be biased to the right. The significance of this is that negative Bernoulli pressures only emerge well into the formation of the aperture. Initially the pressure is still positive and the force is expulsive. As the flow increases, pressure becomes negative, causing collapse of the aperture. The decrease in area only serves to cause the flow to become more rapid until its maximum is reached and ultimately the flow ceases as the glottal aperture below is closed.

Table 6.5 Biomechanical parameter values for voiced aryepiglottic trilling[115]

	Parameters		Voiced Aryepiglottic Trill	
			Tight (Right)	Loose (Left)
Biomechanical	Tension	Edge	10000 g/mm^2	10000 g/mm^2
		Body	10000 g/mm^2	10000 g/mm^2
	Spring Coefficient (k)	Edge	1000 N/m	1800 N/m
		Body	100 N/m	1000 N/m
	Mass	Edge	0.01 kg	0.01 kg
		Body	0.001 kg	0.001 kg
	Damping Factor (ζ)	Edge	0.4	0.5
		Body	1	0.4
Dimensions	Depth[116]	Edge	-	-
		Body	-	-
	Thickness	Edge	0.45 cm	0.45 cm
		Body	0.1 cm	0.1 cm

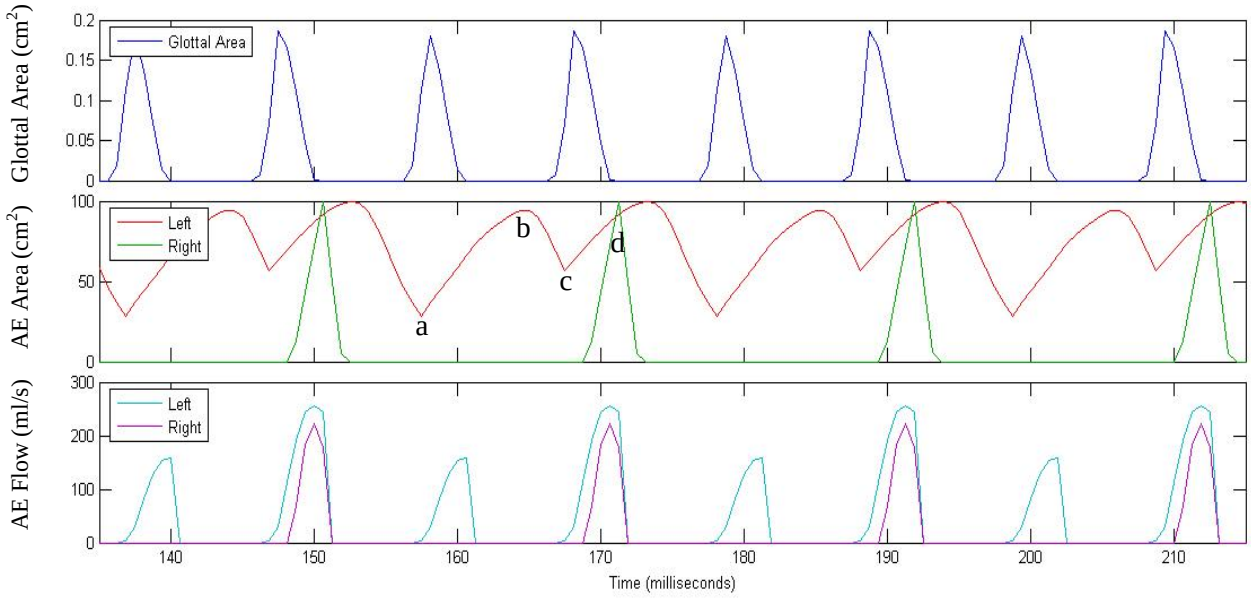


Figure 6.24: Time series plot of asymmetrical aryepiglottic trilling at ~100 Hz (glottal pulse). The middle plot demonstrates the asymmetry between the right pulse (green) at 50 Hz, and the left pulse (red) at 100 Hz. The left fold is also continuously open throughout the simulation on account of the simulated loose cuneiform configuration, which allows for the left pulse to be directly related to the glottal pulse. Aryepiglottic airflow is also considered; again, due to the modeling of vocal tract inertance, flow increases slowly after the opening of the glottis, which means that pressures in the aryepiglottic lumen are not initially negative, but rather are positive until the velocity has had a chance to increase to an appropriate level. The markers (a-d) correspond to the frames used in Figure 6.25.

[115] The remainder of the simulation properties such as aerodynamic and other parameters are presented already in Table 5.1.

[116] Particle volumes are assumed to be sphereical, where thickness represents the diameter of the particle, thus depth is not used.

The appearance of the simulation is demonstrated in Figure 6.25, where a selection of frames across one pulse of the trilling sequence is examined (which are indicated in Figure 6.24). Initially, in Figure 6.25a, the left fold is slightly ajar, which reflects the simulation of the loose cuneiform configuration. On the right side there is complete contact between the aryepiglottic fold and the tubercle of the epiglottis. Trilling of the aryepiglottic folds commences on the left side first, Figure 6.25b, which directly receives the glottal pulse, while the right fold appears static, although the lower portion of the fold (not visually modeled) is being forced open. During the time between the first and second glottal pulse, the loose aryepiglottic fold makes a return towards the epiglottis due to elastic restoring forces (Figure 6.25c). With the second pulse, the right fold finally opens (Figure 6.25d), as the left fold continuous its 1:1 oscillatory relationship with the glottal pulse.

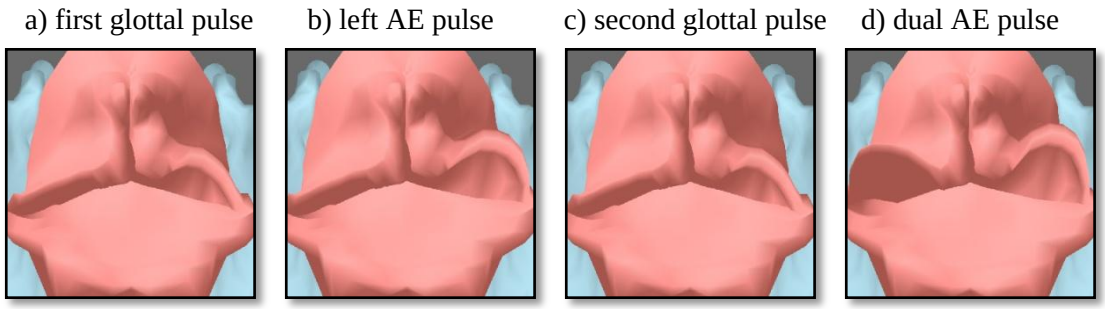


Figure 6.25: Aryepiglottic trilling at 100 Hz (glottal pulse) as it appears in the model from a superior view. The initial frame shows the different cuneiform configurations: the right side is tight and the left side is loose. The pulse of the right fold (i.e. the fold that appears on the left) occurs during the third frame, while the left fold pulse occurs twice in the sequence (maxima during the second and fourth frames). The image indices are matched with those on the middle plot of Figure 6.24 to illustrate where each screen shot lies in the trilling cycle.

6.5 Summary of the 3D LCM

This chapter has provided a broad view of the many features of the 3D Laryngeal Constrictor Model. The overall layout of the model's GUI and the appearance of the mesh representation of laryngeal structures were examined. Measurements taken from studies of real larynges were compared to virtual measurements of the representations of the laryngeal cartilages in the model. The results indicate that the model representations are fairly closely within the range of the measurements provided by the empirical studies. Demonstration of the

visual synthesis of pharyngeal and laryngeal articulations, as well as a variety of phonation types was also provided. These were evaluated using the same measurement technique described in Section 4.2, for the still images extracted from normal speed laryngoscopy of the articulations and phonation types of interest. Once again, there was a reasonable correspondence between measurement values for the simulated articulations, and those for the actual laryngoscopic video frames. Finally, the biomechanical modeling vocal fold and aryepiglottic fold function were evaluated. The glottal model can be manipulated to produce vibrations in the range from 100 Hz to 300 Hz, by varying the degree of cricothyroid activity (and thus thyroid notch-cricoid lamina distance, resulting in elongation of the vocal folds). Volume velocity was also examined and showed the expected right-skewed curve: which was an index of the inertive properties encoded into the aerodynamic modeling. The non-trivial case of aryepiglottic trill was also demonstrated; that is, visual synthesis of asymmetrical aryepiglottic trilling showed that by manipulating cuneiform position, different response patterns could be obtained. The tight cuneiform position showed a pulse rate at half the frequency of the loose cuneiform position, which was maintained at the frequency of the glottal pulse.

Chapter Seven

Conclusion

7.1 Summary

The function of the laryngeal constrictor mechanism has surfaced as an important contemporary issue in the phonetic sciences. The aryepiglottic folds have been identified as the primary articulator in pharyngeal articulations, and are observed to play an important role in shaping the quality of many phonation types. Their action is coupled to the function of the ventricular folds below and the tongue root-epiglottis complex above, and larynx height, enabling the multi-tiered sphincteringof the larynx, serving a vital protective function, and making it possible to shape the laryngeal vocal tract in innumerable ways. Researchers in phonetics have been aware of the importance of the pharyngeal cavity to vowel quality, voice quality, and segmental articulations, but the aryepiglottic folds themselves have rarely been given due consideration of their articulatory contribution or status as a sound source. Esling and his colleagues have strongly advocated for reconceptualizing how the laryngeal vocal tract functions in light of research conducted over the past twelve years (Esling 1996, 1999a, 1999b, 2002, 2005; Esling & Clayards 1999; Esling, Carlson, & Harris 2002; Edmondson & Esling 2005; Esling & Harris 2005; Esling, Fraser, & Harris 2005; Esling, Zeroual, & Crevier-Buchman 2007; Edmondson, Padayodi, Hassan, & Esling 2007). The implications of this reconceptualization run deep to the core of issues in phonetics including the articulatory theory of vowels, infant phonetic learning, the source-filter model of vocal tract acoustics, and the nature of pharyngeal articulation.

The 3D Laryngeal Constrictor Model described in this thesis is intended to encapsulate Esling's revised view of articulation in the laryngeal vocal tract, and places its focus upon the structure and movement of the aryepiglottic folds. Modeling these aspects of the laryngeal vocal tract is assertively rare; treatment is usually given to the better understood aspects of the larynx and pharynx, particularly the vocal folds, larynx height, and the tongue root. Attempting to establish a standard representation of the structures that characterize the laryngeal constrictor mechanism is incredibly difficult as they represent some of the most infrequently discussed, imaged, and studied aspects of the larynx. There is also considerable variation of the structures, which makes it difficult to say with any degree of certainty that the representation will be

universal or even correct, from a statistical perspective. The model is, thus, likely to be the first of its kind, and should be regarded as such. The hope is that it will establish a dialogue among researchers of phonetics, anatomy, and speech engineering to pursue future models that are more elaborate and accurate.

The particular approach adopted in creating model structures drew upon histological data collected by Hirano & Sato (1993) and vertex tracing to segment the structures and ultimately build 3D mesh representations of them. The measurements obtained in Section 6.2 indicate, based on comparison of measurements from Eckel et al. (1994) and Titze (2006) that this technique was highly successful in producing faithful representations of laryngeal structures. This approach has a number of major draw backs, however, not the least of which is that the data set cannot be considered complete in light of the scope of the original project. The current phase of the model lacks representations of the tongue root, pharyngeal wall, hyoid bone and thyrohyoid membrane, in addition to the extrinsic laryngeal muscles because of the unavailability of these structures in the imaging resources. Furthermore, while extraction of the laryngeal cartilages proved to be a simple matter, extraction of soft tissues was complicated by considerable asymmetry in the structures and deformation of the tissues that result from preserving and fixing the tissues in the process of creating histological sections.

Basic movement properties were developed using scientific descriptions of the range and character of movement for the arytenoids, epiglottis, thyroid, and cricoid cartilages. Despite being a necessary feature of the laryngeal model, the benefit of implementing the basic movements, particularly those of the arytenoids, is that the aryepiglottic folds inherit much of their movement from the positioning of the arytenoids, cuneiforms, and epiglottis. The remainder of the movements, particularly those of an articulatory or phonatory nature, were determined using laryngoscopic videos to assess how the various laryngeal landmarks correspond to one another during cardinal articulatory configurations. These data were used to develop movement constraints on the model parameters during preset articulation control. This leads to another innovative feature of the model: it allows the user to have a complete interactive control over its behaviour through the manipulation of parameters that correspond both to muscle activation on the physiological level, and to laryngeal valves on a linguistic level. This dual contextualization serves to embellish the context under which movement takes place and reinforces the link between muscle contraction and linguistic function of articulators.

The final feature of the model, in terms of its application as a visual synthesizer of constricted and unconstricted laryngeal articulation, is the implementation of biomechanical simulations of tissue vibration to simulate various phonation types and aryepiglottic trilling. While the vocal folds have been the subject of numerous studies in biomechanical modeling over the years, the aryepiglottic folds have likely never been modeled. This innovative aspect of the 3D Laryngeal Constrictor Model is founded upon observations of voiced aryepiglottic trilling in high-speed laryngoscopic videos. One of the central insights into how aryepiglottic trill functions obtained from the image analysis is that there are two aryepiglottic fold configurations that yield entirely different movement patterns and consequently, asymmetrical trilling. These configurations are suspected to be primarily determined by the anatomical positioning of the cuneiform cartilages within the aryepiglottic fold relative to the epiglottis during maximal aryepiglottic stricture. If the cuneiform cartilage forces the fold into a patent or "loose" configuration, the aryepiglottic fold trills at and above the same frequency of the vocal folds. If the cuneiform cartilage forces the aryepiglottic fold to tightly press against the epiglottis at the tubercle, half-periodic vibration appears to result. A hypothesis was developed to explain why the tight cuneiform configuration produces vibration at half of the fundamental frequency, based on the observation that a mucosal seal forms each time the aryepiglottic fold comes into contact with the epiglottis. It was assumed that the depth of this seal runs below the superficial most part of the aryepiglottic fold. This means that the lower part of the aryepiglotto-epiglottal contact must be opened before the entire aryepiglottic aperture can open, which takes precisely two glottal pulses to accomplish. The image analysis also allowed for the movement characteristics of aryepiglottic fold vibration to be extracted, and these were also implemented into the biomechanical model in the form of force vectors, to cause the particles in the simulation to move in the observed direction.

The performance of the aryepiglottic model is limited, however, and does not provide a complete explanation of aryepiglottic function in the terms of the myoelastic aerodynamic theory of phonation. This is primarily due to the shared one-dimensional model of airflow through the aryepiglottic apertures based on Titze's (1973) glottal airflow model. It is intended to be functional rather than explanatory, which stands as a goal for future research into the trilling mechanism. It does, however, represent a first step towards a more articulated theory of this

unique articulation/ phonation type and possible consequences for source-filter theory; two aryepiglottic apertures simultaneously and asymmetrically vibrating means potentially two independent vocal sources in addition to the glottal source below.

7.2 Future research and developments

This is the first phase of development for the 3D Laryngeal Constrictor Model. Future phases are planned, which will expand upon the structural elements of the model, providing further context for the simulation of articulation. The obvious components are those that could not be developed in this phase: the tongue root, extrinsic laryngeal muscles and architecture (i.e. the hyoid bone and thyrohyoid membrane), and the pharyngeal wall. Furthermore, the biomechanical simulations require further development, fine tuning, and verification to make the model more robust in terms of its representational verisimilitude. One particular gap concerns voiceless aryepiglottic trilling, and other varieties of aryepiglottic trill. Only the voiced aryeiglottic trill was available for study in the high-speed laryngoscopic video analysis and thus the behaviour of the voiceless trill could not be examined. More research is also required on the subject of aryepiglottic aerodynamics in order for a more felicitous model of trilling to be produced. The complexity of modelling the airflow in two or three dimensions is considerable, but stands as an important goal if a complete model of laryngeal airflow is to be achieved. In terms of geometry, it is unlikely that information about the vertical dimension of trilling, and aryepiglottic constriction in general, will be obtained easily, given that laryngoscopy only informs us about movements the *x*-*z* plane, but this still stands as an area that needs more refinement. These developments are intended to take place alongside the next major foray into synthesis: that of the acoustic output of the model. As a tool intended to illustrate articulation, it is ultimately only as good as the auditory output, which allows for the connection between articulation and sound production to be established. The goal then is for real time acoustic synthesis to be produced according to the parameter settings of the model. The acoustic source generated by the biomechanical models can be filtered by a virtual oral vocal tract to produce recognizable vowel sounds, and provide an accurate representation of pharyngeal and laryngeal segmental articulations.

Other extensions of the 3D Laryngeal Constrictor Model are imaginable. One of the more useful developments would be an investigation of laryngeal embryology to illustrate the

connection between constriction salient in early vocalizations of infants and laryngeal constriction in the adult larynx. More medically oriented applications can be conceived, such as vascular structure and laryngeal innervation. The latter is also a topic that is suitable for linguistic investigation to explore the timing and sequencing of neuromuscular events responsible for laryngeal movements.

Practical applications of the model also exist. One of the most obvious is to employ the model as a teaching tool for students of phonetics and linguistics, and also for medical students and the interested layperson. An ideal mode of presentation of the model is through the internet, although this would require migrating much of the program code to a format that is compatible with internet protocall and a wider range of computer platforms[117]. The model may also be used for entirely non-linguistic applications, such as vocal training. Visualization of the larynx may help singers become more aware of laryngeal structures, particularly the aryepiglottic folds, which can dramatically impact the quality of musical vocalization. The beauty of the model is that it is digital and, consequently, highly modifiable, extensible, and adaptable to new ideas and research.

[117] One such platform is Java 3D developed by Sun Microsystems.

BIBLIOGRAPHY

Agarwal, M., Scherer, R. C. and Hollien, H. 2003. The False Vocal Folds: Shape and Size in Frontal View During Phonation Based on Laminagraphic Tracings. *Journal of Voice* 17 (2). 97-113.

Ali, L.H. & Daniloff, R.E. 1972. A contrastive cinefluorographic investigation of the articulation of emphatic-non-emphatic cognate consonants. *Studia Linguistica* 26. 81-105.

Al-Nassir, A. 1993. *Sibawayh the Phonologist*. London: Kegan Paul.

Badin, P., Borel, P., Bailly, G., Reveret, M., Baciu, M., & Segebarth, C. 2000. Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images. *5th Speech Production Seminar*, München.

Baker, K., Ramig, L., Sapir, S., Luschei, E., & Smith, M. 2001. Control of vocal loudness in young and old adults. *Journal Speech, Language and Hearing Research* 44. 297–305.

Barry, W. 1997. Another [r]-tickle. *Journal of the International Phonetics Association* 27 (1-2). 35-45.

Bartlett, D. 1989. Respiratory Functions of the Larynx. Physiological Reviews 69 (1). 33-57.

Benner, A., Grenon, I., & Esling, J. H. 2007. Infants' phonetic acquisition of voice quality parameters in the first year of life. In *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken, Germany. 2073-2076.

Berg, Jw. van den. 1954. Sur les theories myo-elastique et neurochronaxique de la phonation. *Revue de Laryngologie*. 495-512.

Berg, Jw. van den. 1955. On the role of the laryngeal ventricle in voice production. *Folia Phoniatrica* 7:57-69.

Berg, Jw. van den. 1960. Vocal ligaments versus registers. Current Problems in Phoniatrics and Logopedics 1. 19.

Berg, Jw. van den & Moll, J. 1955. Zur Anatomie des meschlichen Musculus vocalis. *Zeitschift fur Anatomie und Enticklungsgeschichte* 118. 465-470.

Beskow, J. 2003. *Talking heads – models and applications for multimodal speech synthesis.* Doctoral dissertation, KTH, Stockholm, Sweden.

Blankenship, B. 1997. The Time Course of Breathiness and Laryngealization in Vowels. UCLA Doctoral dissertation.

Bird, S., Caldecott, M., Campbell, F., Gick, B., Shaw, P. 2008. Oral–laryngeal timing in glottalised resonants. *Journal of Phonetics*.

Birkholz, P. & Jackel, D. 2003. A three-dimensional model of the vocal tract for speech synthesis. In *Proceedings of the 15th International Congress Phonetic Sciences*. Barcelona, Spain. 2597–2600.

Birkholz, P. & Kröger, B. J. 2007. Simulation of vocal tract growth for articulatory speech synthesis. In the *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken, Germany. 377-380.

Brancatisano, A., Dodd, D. S. & Engel, L. A. 1991. Posterior cricoarytenoid activity and glottic size during hyperpnea in humans. *Journal of Applied Physiology* 71. 997-982.

Bukshaisha, F. A. M. 1985. *An Experimental Phonetic Study of Some Aspects of Qatari Arabic*. Ph.D. Dissertation. University of Edinburgh.

Burden, R.L. & Faires, J. D. 2001. *Numerical Analysis*. Seventh edition. Thomson Learning

Butcher, A. & Ahmad, K. 1987. Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic. *Phonetica* 44. 156-172.

Carlson, B. F., Esling, J. H., & Harris, J. G. 2004. A laryngoscopic phonetic study of Nlaka'pamux (Thompson) Salish glottal stop, glottalized resonants, and pharyngeals. In D. B. Gerdts, & L. Matthewson (Eds.), *Studies in Salish linguistics in honor of M. Dale Kinkade, occasional papers in linguistics* No. 17. Missoula: Universityof Montana Press. 58–71.

Carlson, R. 1995. Models of speech synthesis. In the *Proceedings of the National Academy of Science* 92. 9932-9937.

Catford, J. C. 1964. Phonation types: the classification of some laryngeal components of speech production. In *In honour of Daniel Jones,* Abercrombie, D., Fry, D. B., MacCarthy, P. A. D., Scott, N. C., and J. L.M. Trim (eds.).

Catford, J. C. 1968. The articulatory possibilities of man. In Malmberg, B. (ed.), *Manual of Phonetics/* Amsterdam: North-Holland. 309-333

Catford, J. C. 1977a. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Catford, J. C. 1977b. Mountain of tongues: the languages of the Caucasus. *Annual Review of Anthropology 6*. 283-214.

Catford, J. C. 1983. Pharyngeal and laryngeal sounds in Caucasian languages. In Bless, D. M. & Abbs, J. H. (eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. San Diego: College Hill Press. 344-350.

Catford, J.C. 2001. On Rs, rhotacism and paleophony. *Journal of the International Phonetics Association* 31 (2). 171-185.

Coker, C. & Fujimura, O. 1966. Model for Specification of the Vocal-Tract Area Function. *Journal of the Acoustical Society of America* 40. 1271.

Colarusso, J. 1985. Pharyngeals and Pharyngealization in Salishan and Wakashan. *International Journal of American Linguistics* 51 (4). 366-368.

Cremers, D. and Soatto, S. 2005. Motion Competition: A Variational Approach to Piecewise Parametric Motion Segmentation. *International Journal of Computer Vision* 62 (3). 249–265.

Delattre, P. 1971. Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English. *Phonetica* 23. 129–155.

Eckel, H.E., Sittel, C., Zorowka, P. & Jerke, A. 1994. Dimensions of the laryngeal framework in adults. *Surgical Radiological Anatomy* 16. 31-36.

Edmondson, J. A. & Esling, J. H. 2005. The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology* 23. 157-191.

Edmondson, J. A., Padayodi, C. M., Hassan, Z. M., & J. H. Esling. 2007. The laryngeal articulator: source and resonator. In the *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken, Germany. 2065-2068.

Ekberg, O. & Pokieser, P. 1997. Radiologic evaluation of the dysphagic patient. *European Radiology* 7 (8). 1285-1295.

Ekberg, O. & Sigurjónsson, S. V. 1982. Movement of the epiglottis during deglutition: A cineradiographic study. *Abdominal Imaging* 7 (1). 101-107.

Engwall O. 1999. Vocal tract modeling in 3D. *KTH TMH-QPSR* 1-2. 31-38.

Engwall, O. 2000. A 3D tongue model based on MRI data. In Yuan, B., Huang, T., & Tang, X. (eds.), *Proceedings of the 6th International Conference on Spoken Language Processing*. Beijing. 901-904.

Elze, C. 1925. Anatomie des Kehlkopfes und des Tracheobronchialbaumes. In Handbuch der Hals-, Nasen-, und Ohrenheilk, Vol. 1, Denker und Kahler (eds.). Spinger & Bergmann, Berlin-München.

Esling, J. H. 1978. Voice Quality in Edinburgh - A Sociolinguistic and Phonetic Study. Ph.D. Dissertation. University of Edinburgh.

Esling, J. H. 1996. Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association* 26. 65-88.

Esling, J. H. 1999a. The IPA Categories "Pharyngeal" and "Epiglottal": Laryngoscopic Observations of Pharyngeal Articulations and Larynx Height. *Language and Speech* 42 (4). 349-372.

Esling, J. H. 1999b. Voice quality settings of the pharynx. In the *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, USA. 2449-2452.

Esling, J. H. 2002. Laryngoscopic analysis of Tibetan chanting modes and their relationship to register in Sino-Tibetan. Proceedings of the *7th International Conference on Spoken Language Processing* (2). Denver, Colorado. 1081-1084.

Esling, J. H. 2005. There are no back vowels: the laryngeal articulator model. *Canadian Journal of Linguistics* 50. 13-44.

Esling, J. H., Carlson, B. F., & Harris, J. G. 2002. A laryngoscopic phonetic study of Nootka and Salish glottal stop, glottalized resonants and pharyngeals. Presented at *The Society for the Study of the Indigenous Languages of the Americas,* San Fransisco, Jan. 5, 2002.

Esling, J. H. & Clayards, J.A.W. 1999. Laryngoscopic analysis of pharygeal articulations and layrnx-height voice quality settings: *Advances in Phonetics. Proceedings of the International Phonetic Sciences Conference (IPS).* Braun, A. (ed.). Stuttgart: Franz Steiner Verlag. 24-33.

Esling, J. H., Fraser, K. F., & Harris, J. G. 2005. Glottal stop, glottalized resonants, and pharyngeals: a reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics* 33. 383–410.

Esling, J. H. & Harris, J. G. 2005. States of the glottis: an articulatory phonetic model based on laryngoscopic observations. In *A figure of speech: a Festschrift for John Laver*. Hardcastle, W. J. and J. M. Beck (eds.). Mahwah, NJ: Erlbaum. 347–383.

Esling, J. H., Zeroual, C. & Crevier-Buchman, L. 2007. A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. In the *Proceedings of the 16th International Congress Phonetic Sciences,* Trouvain J. and W.J. Barry (eds.). Saarbrücken: Universität des Saarlandes, Germany. 585-588.

Etienne-Cummings, R. 2001. Biologically Inspired Visual Motion Detection in VLSI. *International Journal of Computer Vision* 44 (3). 175–198.

Fant, G. 1992. Vocal tract area functions of Swedish vowels and a new three- parameter model. In Proceedings of the *ICSLP-92*, 1. 807-810.

Fant, G. 1993. A new three-parameter model of VT area functions. *ESPRIT BR, SPEECHMAPS (6975), report WP2 year 1, delivery 5*, 3. 1-28.

Fant, G. & Båvegård, M. 1997. Parametric model of VT area functions: vowels and consonants. *TMH-Q: Progress Status Report* 31. 1–20.

Fels, S.S., Lloyd, J.E., van den Doel, K., Vogt, F., Stavness, I. & Vatikiotis-Bateson. 2006. Developing Physically-Based, Dynamic Vocal Tract Models using ArtiSynth. In *Proceedings of the 7th International Seminar of Speech Production*. Ubatuba, Brazil. 419-426.

Flanagan, J. L. 1971. Personal communication between Titze and Flanaga. See references of Titze (1974: 17).

Flanagan, J. L. & Landgraf, L. L. 1968. Self-oscillating source for vocal-tract synthesizers. In *IEEE Transactions on Audio and Electroacoustics* AU-16. 57-64.

Fujimura, O. & Sawashima, M. 1971. Consonant sequences and laryngeal control. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 5. University of Tokyo. 1-13.

Fujisaki, H., Ohno, S.,& Gu W. 2004. Physiological and Physical Mechanisms for Fundamental Frequency Control in Some Tone Languages and a Command-Response Model for Generation of Their *F*0 Contours. *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages Beijing, China.*

Garrison, D. H. & Hast, M. H. 1993. Andreas Vesalius on the larynx and hyoid bone: an annotated translation from the 1543 and 1555 editions of *De humani corporis fabrica*. *Medical History* 37. 3-36.

Gerratt, B. R. & Kreiman, J. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29. 365-381.

Ghazeli, S. 1977. *Back Consonants and Backing Coarticulation in Arabic*. Unpublished Ph.D. Dissertation. University of Texas, Austin.

Gick, B. 2002. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32 (2). 113-121.

Goerttler, K. 1950. Die Anordnung, Histologie und Histogenese der quergestreiften Muskulatur im menschlichen Stimmband. *Zeitschift fur Anatomie und Enticklungsgeschichte* 115. 352-401.

Gordon, M. & Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29. 383-406.

Granat, J., Boë1, L., Badin, P., Pochic, D., Heim, J., Peyre, É. & Benoît, R. 2007. Prediction of the ability of reconstituted vocal tracts of fossils to produce speech. In *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken, Germany. 381-384.

Grenon, I., Benner, A. & Esling, J. H. 2007. Language-specific phonetic production patterns in the first year of life. In *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken, Germany. 1561-1564.

Guida, H. L. & Zorzetto, N. L. 2007. Morphological and Histochemical Analysis of the Human Vestibular Fold. International Journal of Morphology 25 (3). 537-543.

Han, J., Kim, J., Lee, D., Lee, J., Kim, J., Park, K., & Kang, H. 2002. Three-dimensional Visualization of the Human Vocal Organ on the Web. *Journal of Digital Imaging* 15 (1). 264-266.

Harrison, D.F.N. 1995. The anatomy and physiology of the mammalian larynx. Cambridge University Press.

Hassan, Z. M. & Esling, J. H. 2007. Laryngoscopic (articulatory) and acoustic evidence of a prevailing emphatic feature over the word in Arabic. In the *Proceedings of the 16th International Congress Phonetic Sciences*, Trouvain J. and W.J. Barry (eds.). Saarbrücken: Universität des Saarlandes, Germany. 1753-1756.

Heinz, J. M. & Stevens, K.N. 1965. On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech. In Proceedings of the *5th International Conference on Acoustics*. A44.

Heman-Ackah , Y. & Barr, A. 2007. The Value of Laryngeal Electromyography in the Evaluation of Laryngeal Motion Abnormalities . *Journal of Voice* 20 (3). 452 – 460.

Henton, C. G. & Bladon, A. 1988. Creak as a sociophonetic marker. In *Language, speech and mind*: *studies in honour of Victoria A. Fromkin.* Hyman L. and C. Li (eds.). London: Routledge. 3-29.

Heselwood, B. 2007. The 'tight approximant' variant of the Arabic 'ayn. *Journal of the International Phonetic Association* 37. 1-32

Heselwood, B. & Al-Tamimi, F. Y. A. 2006. A nasoendoscopic study of epiglottal activity in the laryngeal, pharyngeal and pharyngealised consonants of Jordanian Arabic. British Association of Academic Phoneticians' Colloquium, Queen Margaret University College, Edinburgh, April 10–12, 2006.

Hirano, M. 1974. Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica* 26. 89-94.

Hirano, M. & Sato, K. 1993. Histological color atlas of the human larynx. San Diego: Singular Publishing Group.

Hirose, H. 1995. Investigating the physiology of laryngeal structures. In *The handbook of phonetic sciences*. Hardcastle, W. and J. Laver (eds.). Cambridge, MA: Blackwells. 116-136.

Hockett, C.F. 1958. *A Course in Modern Linguistics*. New York: The Macmillan Company.

Hollien, H. 1974. On Vocal Registers. *Journal of Phonetics* 2. 125-143.

Hollien, H., Moore, P., Wendahl, R. W. & Michel, J. 1966. On the nature of vocal fry. *Journal of Speech and Hearing Research* 9. 245-247.

Hong, K. H., Kim, H. K., & Kim, Y. H. 2001. The Role of the Pars Recta and Pars Oblique of Cricothyroid Muscle in Speech Production. *Journal of Voice* 15 (4). 512–518.

Honda, K., Takemoto, H., Kitamura, T., Fujita, S., & Takano, S. 2004. Exploring Human Speech Production Mechanisms by MRI. *The Institute of Electronics, Information and Communication Engineers (IEICE): Transactions on Information and Systems* E87-D (5). 1050-1058.

Husson, R. 1950. Étude des Phenomenes Physiologiques et Acoustiques Fondamentaux de la Voix Cantée. University of Paris, Paris. Thesis.

Hunter, E. J., Titze, I. R., & Alipour, F. 2004. A three-dimensional model of vocal fold abduction/adduction. *Journal of the Acoustical Society of America* 115. 1747–1759.

International Phonetic Association (IPA). 1993. Council actions on the revision of the IPA. *Journal of the International Phonetic Association* 23. 32-34.

Ishizaka, K. & Flanagan, J.L. 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Systems Technical Journal* 51. 1233-1268.

Jones, S. 1934. Somali [ħ] and [ʕ]. *Le Maître Phonétique* 49. 8-9.

Kahrilas, P.J., Lin S., Chen, J. & Logemann, J. A. 1995. Three-dimensional modeling of the oropharynx during swallowing. *Radiology* 194. 575-579.

Kaplan, H. M. 1971. Anatomy and Physiology of Speech (2nd ed). McGraw-Hill.

Kier, W. M. & Smith, K. K. 1985. Tongues, tentacles and trunks: The biomechanics of movement in muscular hydrostats. *Zoological Journal of the Linnean Society* 83. 307-324.

Kingston, J., & Nichols, J. 1986. Pharyngealization in Chechen. *Journal of the Acoustical Society of America* 80 (S1). S62.

Kitamura, T., Honda, K., & Takemoto, H. 2004 Individual variation of the hypopharyngeal cavities and its acoustic effects. *Acoustic Science & Technology* 26, 1. 16-26.

Klatt, D. H. & Klatt, L. C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 88. 820-857.

Kokawa, T., Saigusa, H., Aino, I., Matsuoka, C., Nakamura, T., Tanuma, K., Yamashita, K., & Niimi, S. 2006. Physiological Studies of Retrusive Movements of the Human Tongue. *Journal of Voice* 20, 3. 414-422.

Kotby, M.N., Kirchner, J.A., Kahane, J.C., Basiouny, S.E., & EL-Samaa, M. 1991. Histo–anatomical structure of the human laryngeal ventricle. *Acta Otolaryngologica* 111. 396–402.

Krmpotic, J. 1957. Le cartilage de Wrisberg, ses variations et sa relation avec l'aryténoide. *Revue de laryngologie, otologie, rhinologie* 78 (11-12). 945-951.

Ladefoged, P. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

Ladefoged, P. & Maddieson, I. 1996. *The sounds of the world's languages*. Cambridge, MA: Blackwells.

Laradi, W. 1983. *Pharyngealisation in Libyan (Tripoli) Arabic*. PhD, Univ. Edinburgh.

Laufer, A. 1991. Does the 'voiced epiglottal plosive' exist? *Journal of the International Phonetic Association* 21 (1). 44-45.

Laufer, A. & Condax, I. D. 1979. The epiglottis as an articulator. *UCLA Working Papers in Phonetics* 45. 60-83.

Laufer, A. & Condax, I. D. 1981. The function of the epiglottis in speech. Language and Speech, 24, 39-62.

Laufer, A. & Baer, T. 1988. The emphatic and pharyngeal sounds in Hebrew and Arabic. *Lang. Language and Speech* 24, 39-61.

Laver, J. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Lehn, W. 1963. Emphasis in Cairo Arabic. *Language* 39 (1). 29-39.

Lindau, M. 1978. Vowel Features. *Language* 54 (3). 541-563.

Lindau, M. 1985. The Story of /r/. In V. A. Fromkin (ed.), *Phonetic Linguistics*. Orlando & al., Academic Press. 157-168.

Lindestad, P.A.,Sodersten, M., Merker, B., & Granqvist, S. 2001. Voice source characteristics in mongolian "throat singing" studied with high-speed imaging technique, acoustic spectra, and inverse filtering. *Journal of Voice* 15 (1). 78–85.

Logemann, J. A. 1993. *Manual for the videofluoroscopic study of swallowing* (2nd ed.). Austin, Texas: PRO-ED.

Lufkin, R.B., Hanafee, W.N., Wortham, D. & Hoover, L. 1986. Larynx and hypopharynx: MR imaging with surface coils. *Radiology* 158. 747-754.

Maeda, S. 1972. Conversion of Midsagittal Dimensions to Vocal Tract Area Function. *The Journal of the Acoustical Society of America* 51. 88.

Mayet, A. 1955. Zur funktionellen Anatomie der meschilichen Stimmlippe. *Zeitschift fur Anatomie und Enticklungsgeschichte* 119. 87-111.

Mayet, A. & Muendnich, K. 1958. Beitrag zur Anatomie und zur function des M. Cricothyroideus und der Cricothyreiodgelenke. *Acta Anatomica* 33. 273-288.

Meller, S. M. 1984. Functional anatomy of the larynx. *Otolaryngology Clinics of North America* 17. 3–12.

Mermelstein, P. 1973. Articulatory model of speech production. *The Journal of the Acoustical Society of America* 53. 1070-1082.

Miller, A. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics*. 56-84.

Millington, I. 2007. *Game physics: engine development*. Elsevier Inc.

Mithun, M. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.

Montagnoli, A. N., Rubert, J. B., Guido, R. C. & Pereira, J. C. 2006. Vocal folds vibrations with a three-dimensional deformable model. *Proceedings of the Eighth IEEE International Symposium on Multimedia*. 674 – 678.

Moore, G, P. 1971. Subtle Laryngeal Adjustments. *Transactions of the American Laryngological Association* 92.

Moisik, S. R. & Esling, J. H. 2007. 3D auditory-articulatory modeling of the laryngeal constrictor mechanism. In *Proceedings of the 16th International Congress Phonetic Sciences*. Saarbrücken, Germany. 373-378.

Moisik, S.R. & Esling, J.H. 2008. Modelling pharyngeal (aryepiglottic) stricture and trilling. Paper presented at the British Association of Academic Phoneticians Colloquium, University of Sheffield, April 2008.

Namdaran, N. 2006. Retraction in St'a´t'imcets: An ultrasonic investigation. MA thesis, University of British Columbia.

Negus, V. E. 1949. The comparative anatomy and physiology of the larynx. New York: Hafner Publishing Company.

O'Haver, T. 2006. Peak finding and measurement. Custom scripts for the MATLAB platform. http://www.wam.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm. Website accessed June 10th, 2008.

Paik, N. Kim, S. J., Lee, H. J., Jeon, J. Y., Lim, J., & Han, T. R. 2008. Movement of the hyoid bone and the epiglottis during swallowing in patients with dysphagia from different etiologies. *Journal of Electromyography and Kinesiology* 18. 329–335.

Painter, C. 1986. The laryngeal vestibule and voice quality. *Archives of Oto-Rhino-Laryngology* 243. 329-337.

Painter, C. 1988. Electroglottogram waveform types. *European Archives of Oto-Rhino-Laryngology*, 245 (2). 116-121.

Palmer, J. M. 1993. Anatomy for Speech and Hearing. 4th edition. Lippincott Williams & Wilkins Inc.

Palo, P. 2006. *A Review of Articulatory Speech Synthesis.* Master's Thesis. Helsinki University of Technology.

Pierrehumbert, J. & Talkin, D. 1992. Lenition of /h/ and glottal stop. *Papers in laboratory phonology II. Gesture, segment, prosody*. Docherty G. J. and D. R. Ladd (eds.) Cambridge: Cambridge University Press. 90-119.

Poirier, P.S., Charpy, A., & Nicolas, A. 1903. Traité d'anatomie humaine. 2e éd. Paris, Masson et Cie.

Polana, R. & Nelson, R. C. 1997. Detection and Recognition of Periodic, Nonrigid Motion. *Internation Journal of Computer Vision* 23 (3). 261-282.

Pracy, R. 1984. The functional anatomy of the immature mammalian larynx. A study with comparative anatomy of the herbivore, carnivore and primate larynx. Ph.D. Dissertation, University of London.

Pretterklieber, M. L. 2003. Functional Anatomy of the Human Intrinsic Laryngeal Muscles. *European Surgery* 35 (5). 250-258.

Protopapas, A. & Lieberman, P. 1997. Fundamental frequency of phonation and perceived emotional stress. *Journal of Acoustical Society of America* 101(4). 2267-77.

Pullum, G. K. & A. Ladusaw, W. 1996. Phonetic Symbol Guide. 2nd Edition. The University of Chicago Press.

Reidenbach, M. M. 1998. The muscular tissue of the vestibular folds of the larynx. *European Archives of Otolaryngology* 255. 365-367.

Roach, P. J. 1979. Laryngeal-oral coarticulation in glottalized English plosives. *Journal of the International Phonetic Association* 9. 2-6.

Saeed, J. I. 1999. *Somali*. Amsterdam & Philadelphia: Benjamins.

Sakakibara, K. I., Kimura, M., Imagawa, H., Niimi, S., & Tayama, N. 2004a. Physiological study of the supraglottal structure. *International Conference on Vocal Fold Physiology and Biomechanics* 2004. Marseille.

Sakakibara, K. I., Fuks, L., Imagawa, H., & Tayama, N. 2004b. Growl voice in pop and ethnic styles. In *Proceedings of the International Symposium on Musical Acoustics*.

Saunders, W. H. 1964. The larynx. *Clinical Symposia* 16. 67-99.

Selbie, S., Gewalt, S. L., & Ludlow, C. L. 2002. Developing an anatomical model of the human laryngeal cartilages from magnetic resonance imaging. *Journal of the Acoustical Society of America* 112 (3). 1077-1090.

Shabana, A. A. 1991. Theory of Vibration (Volume 1: An Introduction). Springer-Verlag New York, Inc.

Sonesson, B. 1959. Die funktionelle Anatomie des Cricoarytaenoidgelenkes. *Zeitschrift für Anatomie und Entwicklungsgeschichte* 121. 292-303.

Sonesson, B. 1960. On the anatomy and vibratory pattern of the human vocal folds. *Acta Oto-Laryngologica, Supplement* 156.

Stevens, K. N. 1977. Physics of Laryngeal Behavior and Larynx Modes. *Phonetica* 34. 264-279.

Stevens, K. N. 1988. Modes of vocal fold vibration based on a two-section model. In *Vocal Physiology: Voice Production, Mechanisms and Functions*. Fujimura, O. (ed.). Raven Press, New York. 357-367.

Stevens, K. N. 1999. Acoustic Phonetics. Cambridge, Mass.: MIT Press.

Stewart, J.M. 1967. Tongue root position in Akan vowel harmony. *Phonetica* 16. 185-204.

Stone, M. 1990. A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America* 87. 2207-2217.

Stone, M. & Lundberg, A. 1996. Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America* 99 (6). 3728-3737.

Tiede, M. K. 1996. An MRI-based study of pharyngeal volume contrasts in Akan and English. *Journal of Phonetics* 24. 399-421.

Titze, I. R. 1973. The human vocal cords: a mathematical model, part I. *Phonetica* 28, 129-170.

Titze, I. R. 1974. The human vocal cords: a mathematical model, part II. *Phonetica* 29, 1-21.

Titze, I. R. 1995. Definitions and nomenclature related to voice quality. *Vocal fold physiology. Voice quality control.* Fujimura,O. and M. Hirano (eds.). San Diego, CA: Singular Publishing Group, Inc. 335-342.

Titze, I. R. 2001. Acoustic interpretation of resonant voice. *Journal of Voice*, 15 (4). 519-528.

Titze, I. R. 2006. The Myoelastic Aerodynamic Theory of Phonation. The National Center for Voice and Speech. Iowa City, Iowa.

Traill, A. 1985. Phonetic and Phonological Studies of !Xo (Bushman). *Quellen zur Khoisan-Forschung* 1. Hamburg: Helmut Buske Verlag.

Traill, A. 1986. The Laryngeal Sphincter as a Phonatory Mechanism in !Xóõ (Bushman). In R. Singer and J. K. Lundy (Eds.), *Variation, Culture and Evolution in African Populations: Papers in Honour of Dr. Hertha de Villiers.* 123-131. Johannesburg: Witwatersrand University Press.

Van Daele, D., Perlman, A., & Cassell, M. 1995. Intrinsic Fibre Architecture and Attachments of a Human Epiglottis and Their Contributions to the Mechanism of Degluttion. *Journal of Anatomy* 186. 1-15.

Van Gelder, A. 1995. Efficient computation of polygon area and polyhedron volume. In: A.W. Paeth (Ed.), *Graphics Gems V.* Academic Press. 35–41.

Vennard, W. 1967. *Singing: The Mechanism and the Technic* (rev. ed.) New York: Carl Fischer.

Vilkman, E., Pitkanen, R., & Suominen, H. 1987. Observations on the structure and biomechanics of the cricothyroid articulation. *Acta Otolaryngology (Stockh)* 103. 117–126.

Von Leden H. & Moore, P. 1961. The mechanics of the cricoarytenoid joint. *Archives of Otolaryngology* 81. 616-625.

Whalen, D. H. 2004. Measurements and modeling in speech production research. *Proceedings of From Sound to Sense: 50+ years of discoveries in speech communication*. June 11th-13th. Cambridge, MA. B107-B117

Wustrow, F. 1952. Bau und Funktion des meschliche Musculus Vocalis. *Zeitschift fur Anatomie und Enticklungsgeschichte* 116. 506-522.

Yehia, H. C. & M. Tiede. 1997. A parametric three-dimensional model of the vocal-tract based on MRI data. In the *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. 1619–1625.

Zemlin, W.R. 1997. *Speech and Hearing Science: Anatomy and Physiology*. 4th edition. Allyn & Bacon, Boston, Mass.

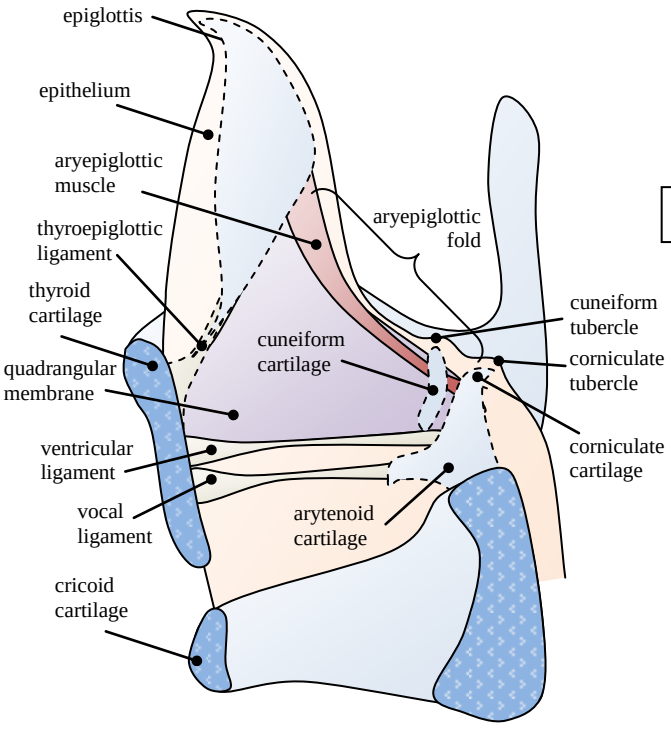
Zheng, J. & Chau, L. 2003. A Motion Vector Recovery Algorithm for Digital Video Using Lagrange Interpolation. IEEE Transactions on broadcasting 49 (4). 383-389.

APPENDIX A

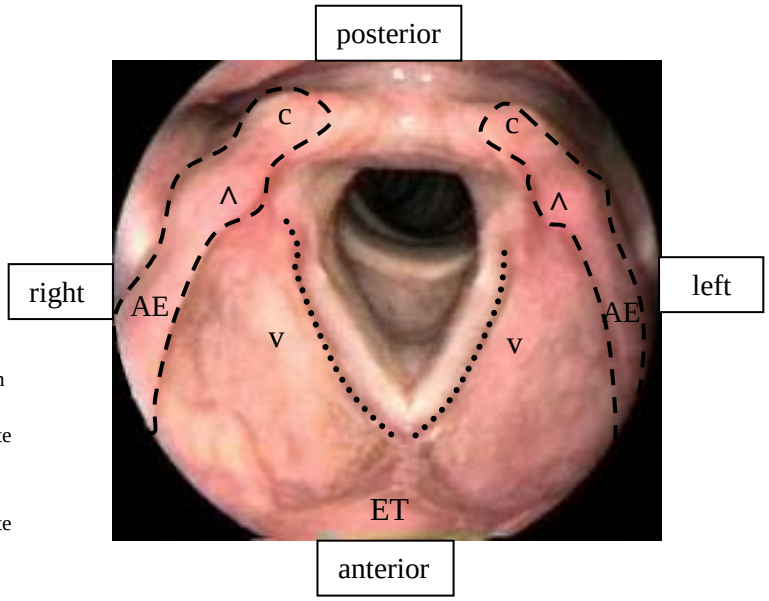
Aryepiglottic Anatomy

The aryepiglottic folds are paired structures arising from the apex & corniculate tubercle of the arytenoids and merging with the mucosa covering the epiglottis at its lateral margins. Together with the base region of the epiglottis (particularly the tubercle of the epiglottis) they enclose the space referred to as the laryngeal vestibule, and the entire structure is sometimes referred to as the epilaryngeal tube or the laryngeal collar. Their external walls form the anterior-medial walls of the piriform fossae. The body of the aryepiglottic folds is comprised of the quadrangular membrane, whose inferior margin terminates as a thickened ligamentous tissue, which forms the ventricular ligament and gives structure to the external region of the ventricular fold. Due to this shared anatomical structure, the aryepiglottic folds and the ventricular folds are intimately linked together physiologically. Constriction of the aryepiglottic folds will bring about medial movement and compaction of the ventricular folds. Within the posterior most portion of the aryepiglottic fold is the cuneiform cartilage, which stiffens the fold and assists in maintaining its sail-like shape when the larynx is not in a constricted posture.

Cut-away Side View



Laryngoscopic View (deep inspiration)



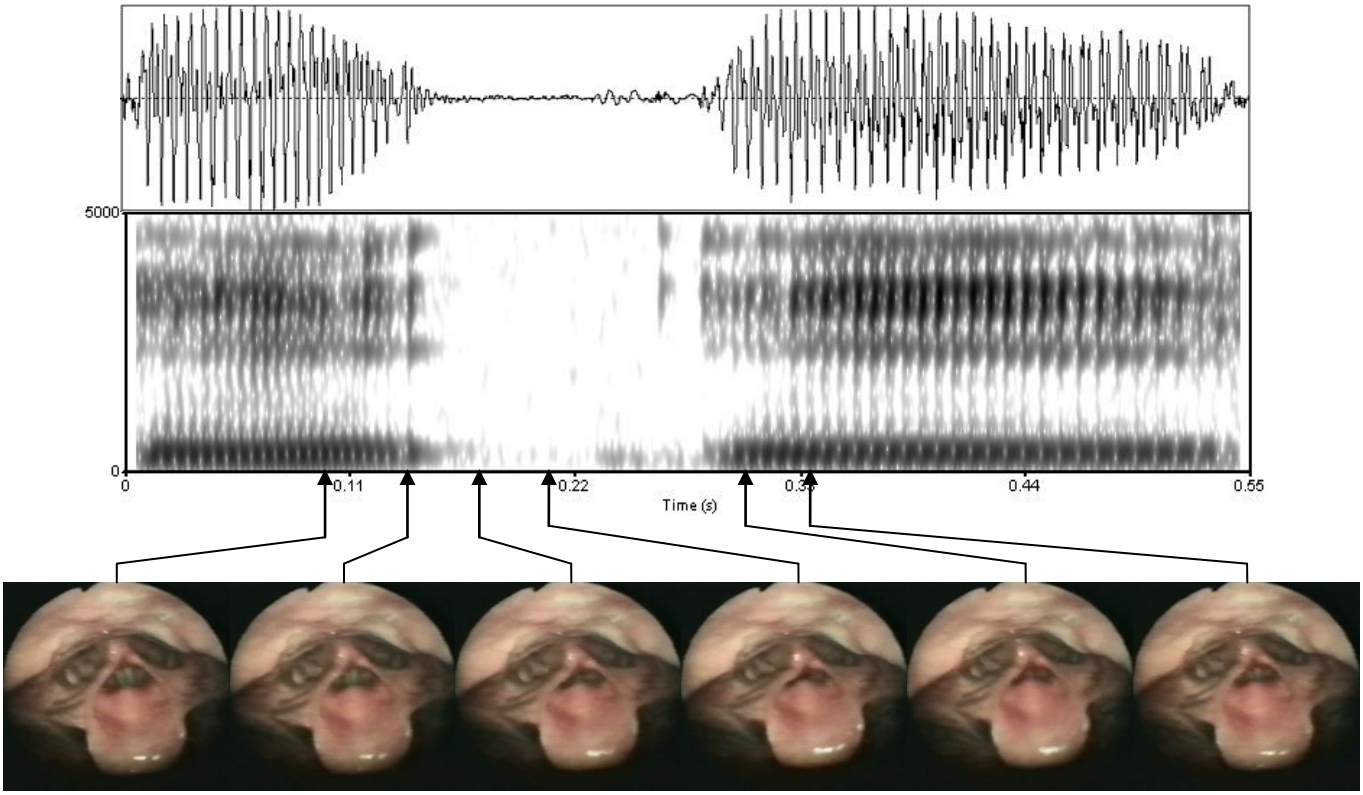
c = corniculate tubercle
^ = cuneiform tubercle
v = ventricular fold (dotted line)
AE = aryepiglottic fold (dashed line)
ET = tubercle of the epiglottis

APPENDIX B

Normal speed laryngoscopy of laryngeal and pharyngeal articulations

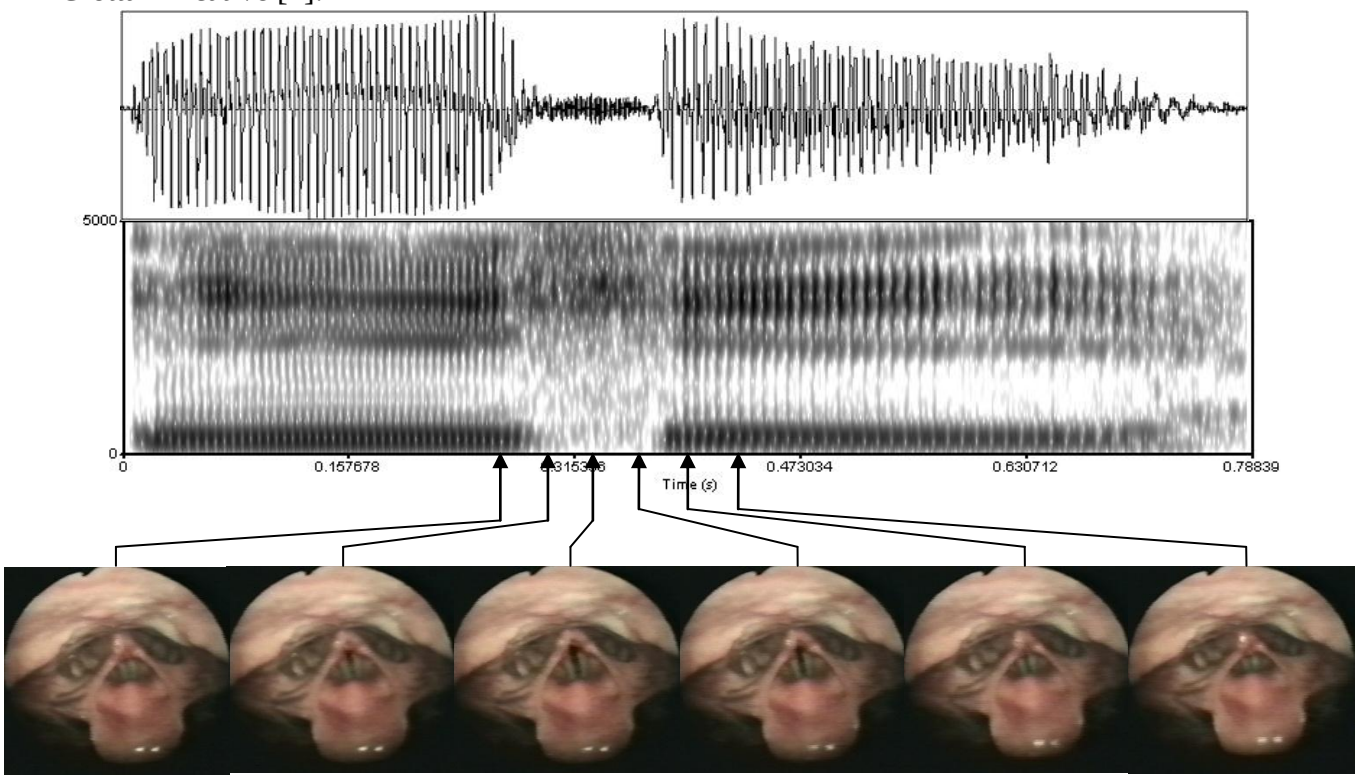
The figures in this appendix provide a more detailed look at the articulatory movements during laryngeal and pharyngeal articulations as seen from normal speed laryngoscopy. In creating the 3D model and its synthetic articulations, these images were considered to be exemplary of canonical type of movement that occurs with each sound. All of these articulations are performed by a linguist and represent careful phonetic production. The sound wave and spectrogram accompany each of the seven sounds examined, [ʔ h ħ ʕ ʡ ʜ ʢ].

Glottal Stop [ʔ]:



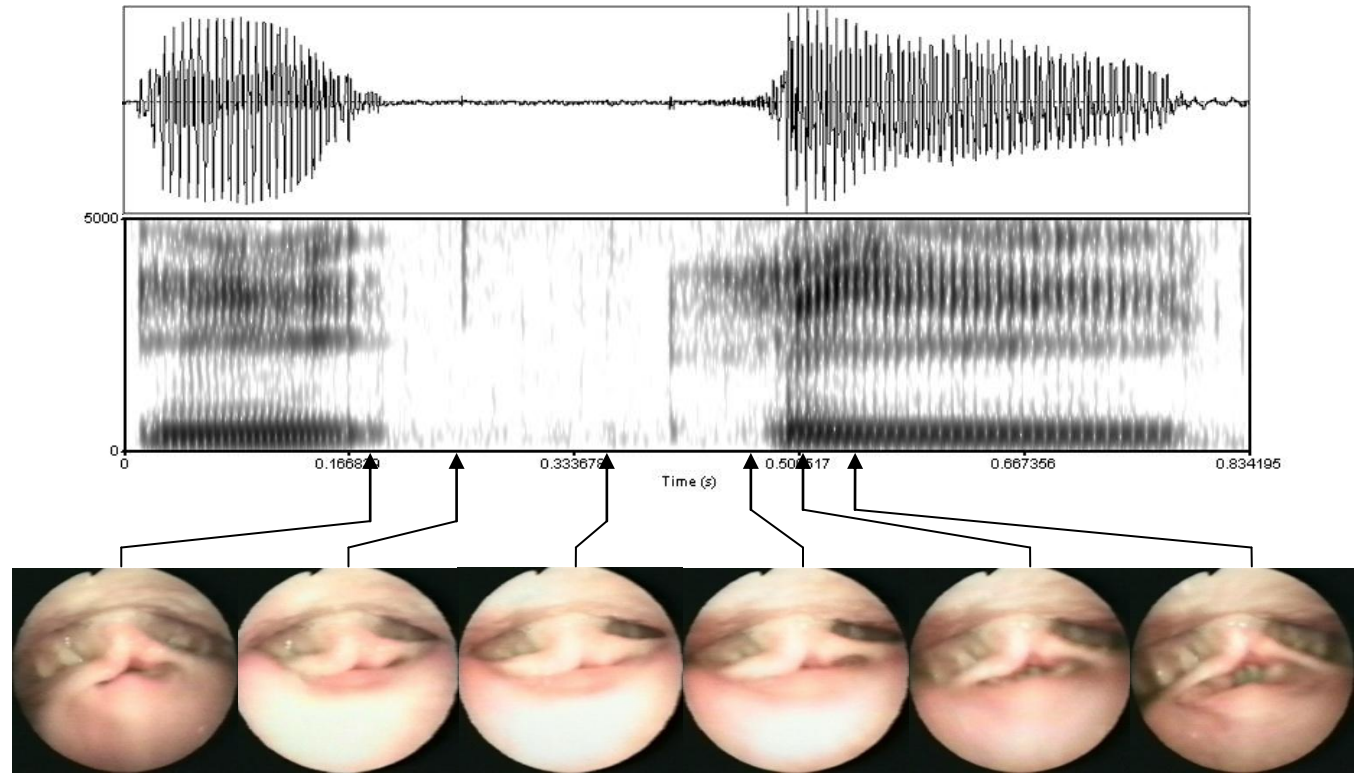
frame: 3; 100ms frame: 4; 134ms frame: 5; 167ms frame: 6; 201ms frame: 9; 303ms frame: 10; 336ms

Glottal Fricative [h]:



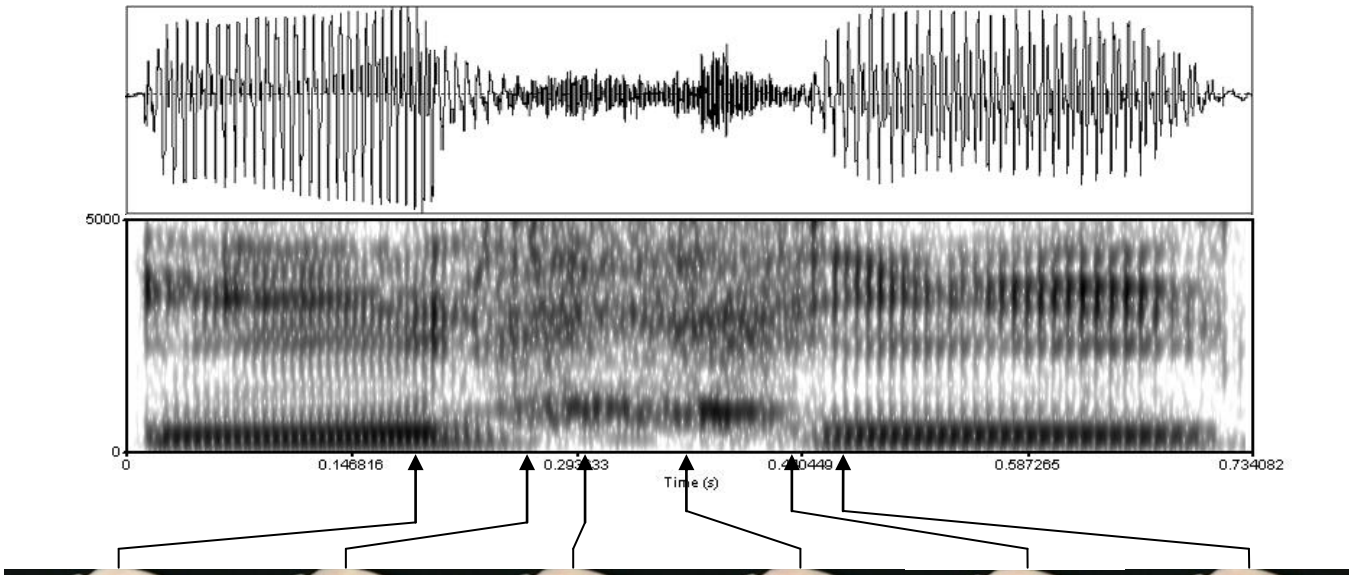
frame: 64; 268ms frame: 65; 302ms frame: 66; 335ms frame: 67; 369ms frame: 68; 402ms frame: 69; 436ms

Pharyngeal (Aryepiglotto-epiglottal Stop) [ʔ]:



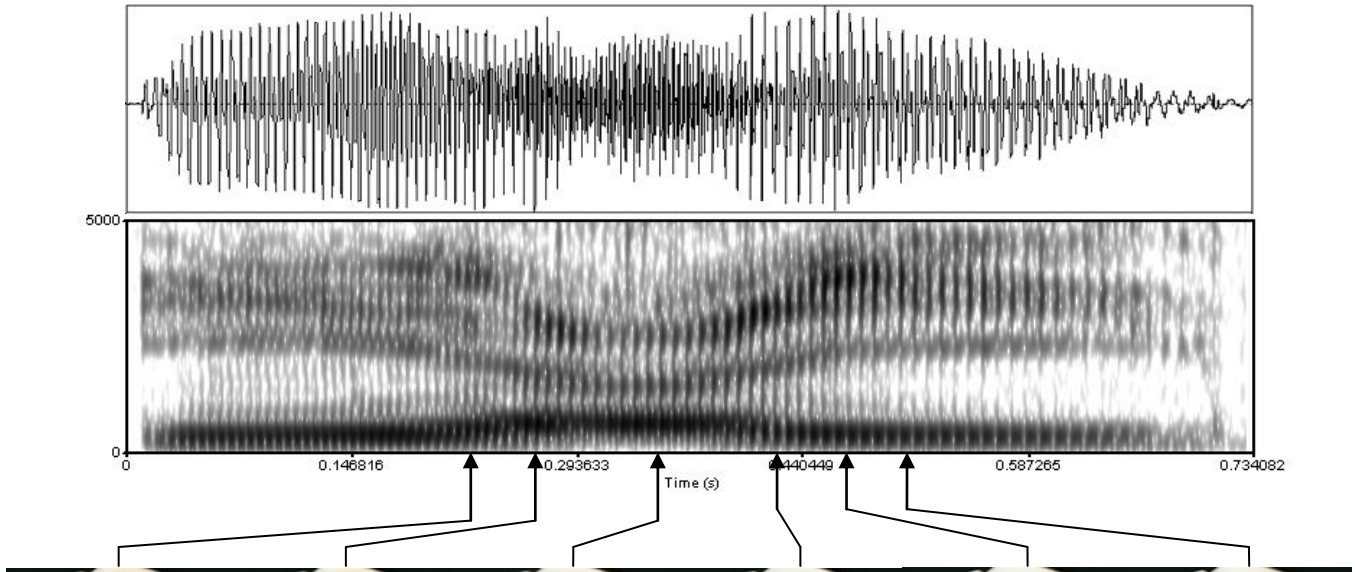
frame: 6; 201ms frame: 8; 268ms frame: 11; 368ms frame: 14; 469ms frame: 15; 502ms frame: 16; 536ms

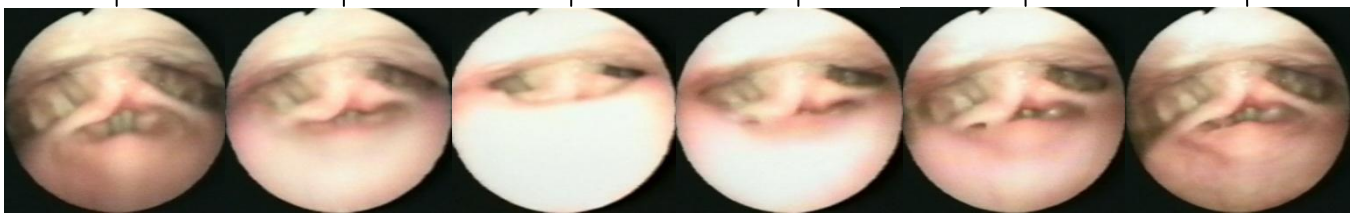
Voiceless Pharyngeal Fricative (Aryepiglotto-epiglottal fricative) [ħ]:



frame: 6; 201ms frame: 8; 268ms frame: 9; 302ms frame: 11; 368ms frame: 13; 435ms frame: 14; 469ms

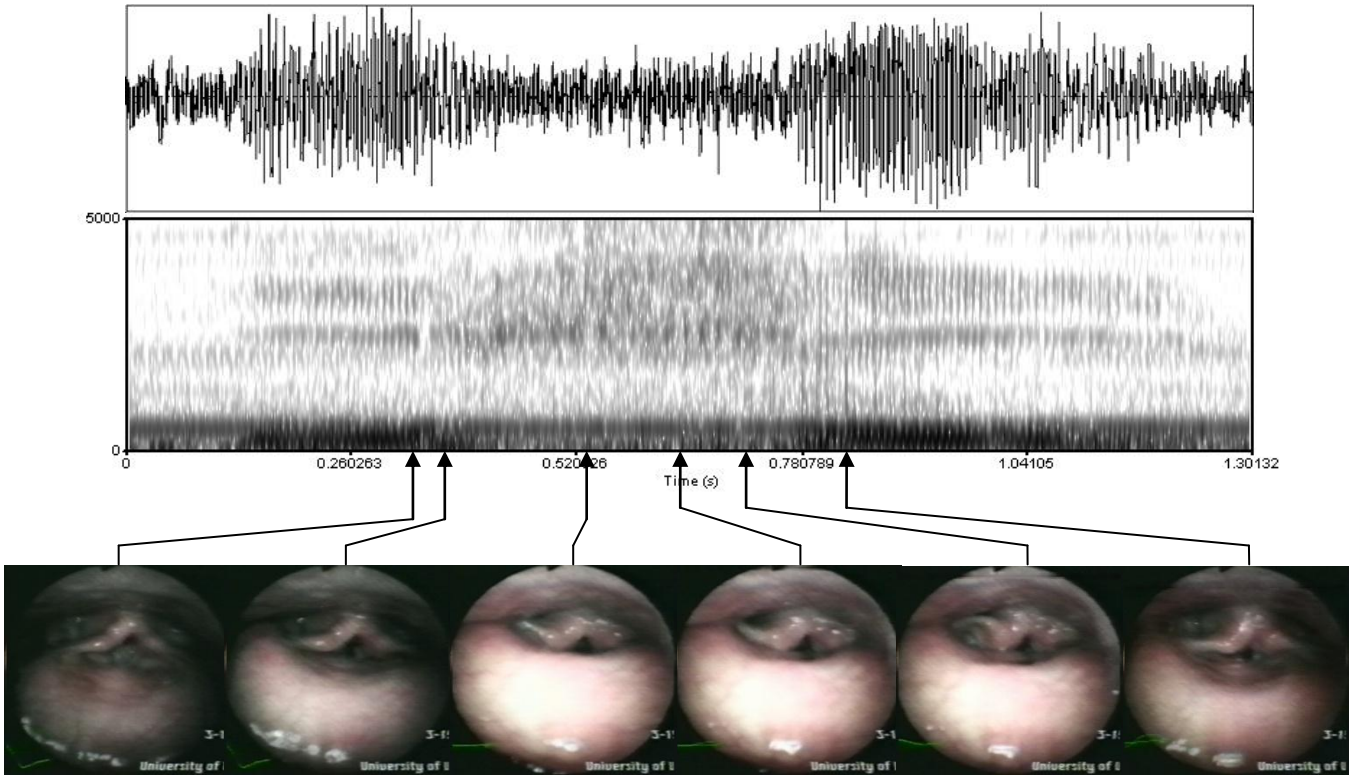
Voiced Pharyngeal Approximant (Aryepiglotto-epiglottal Approximant) [ʕ]:





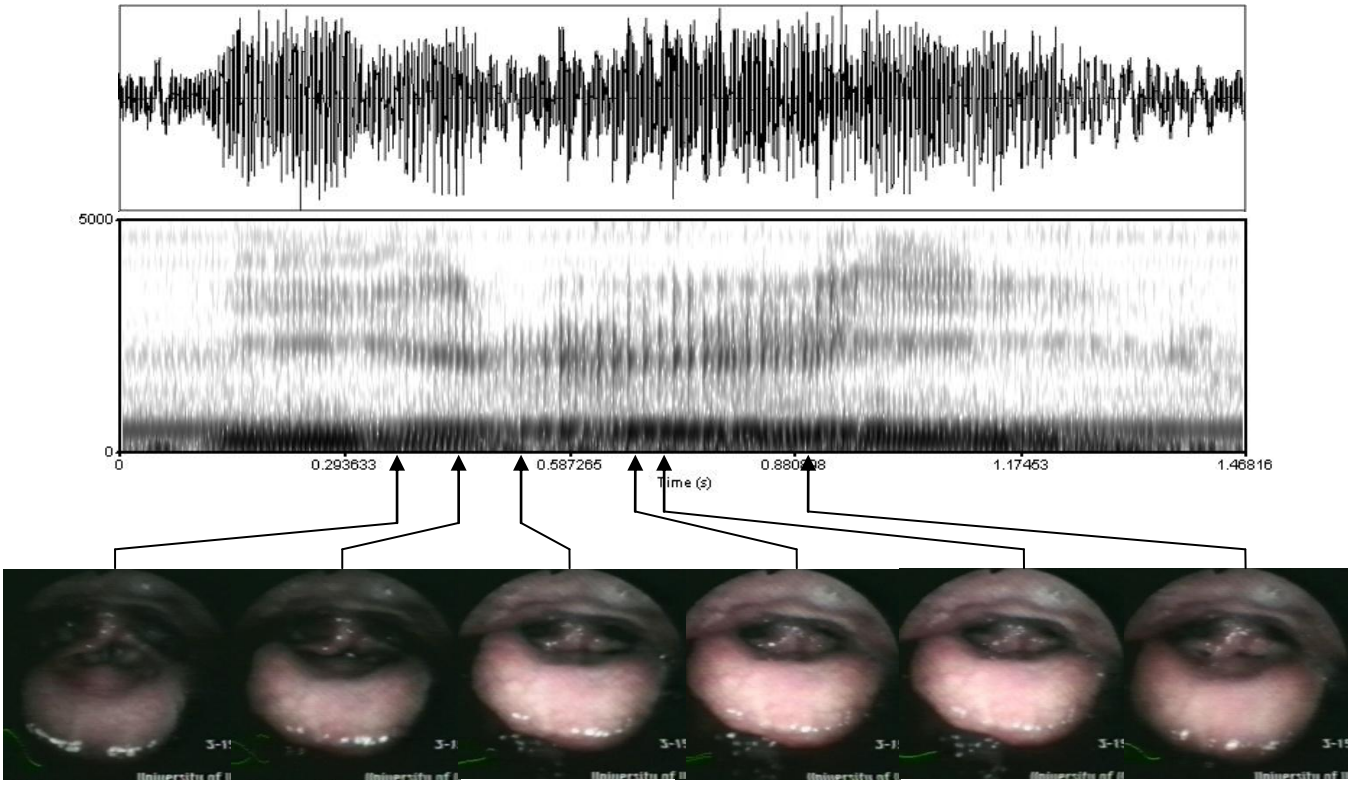
frame: 7; 238ms frame: 8; 272ms frame: 11; 368ms frame: 13; 435ms frame: 14; 469ms frame: 15; 502ms

Voiceless Pharyngeal Trill (Aryepiglotto-epiglottal Trill) [ʜ]:



frame: 11; 368ms frame: 12; 406ms frame: 16; 541ms frame: 19; 642ms frame: 21; 708ms frame: 24; 809ms

Voiced Pharyngeal Trill (Aryepiglotto-epiglottal Trill) [ʕ]:



frame: 11; 368ms frame: 13; 402ms frame: 15; 506ms frame: 19; 640ms frame: 20; 675ms frame: 27; 904ms

APPENDIX C

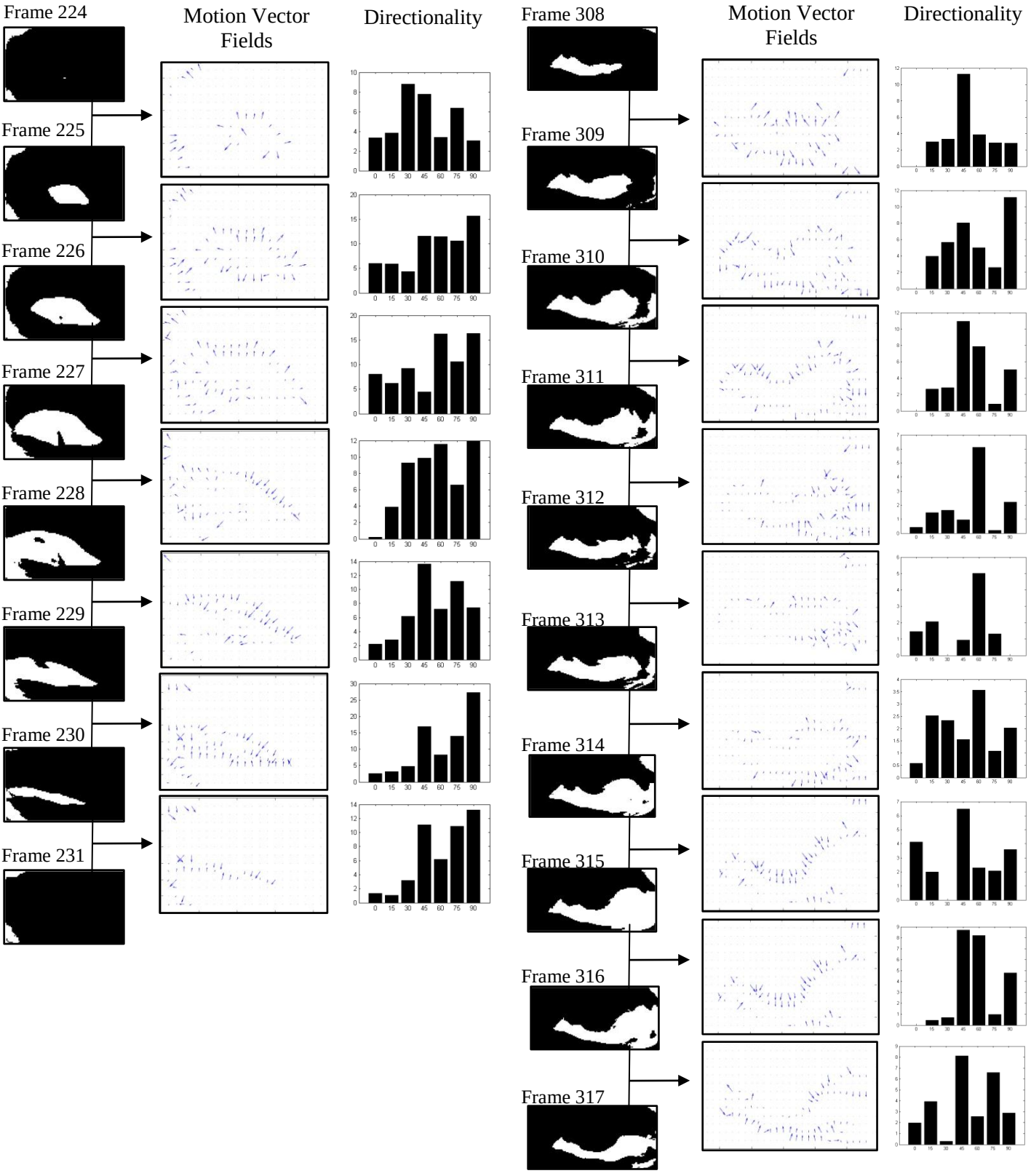
MATLAB m-file code for the motion vector analysis

```
function [ FlipU FlipV OutBar] = motionvectors( binarycurrent, binarynext,  blockdim )
%MOTIONVECTORS

[width height] = size(binarycurrent);

CoordCorrection = int8(blockdim / 2);

hitcount = 0;

BlockRows = int8(width / blockdim);
BlockCols = int8(height / blockdim);

CurrentCentroids = cell(BlockRows, BlockCols);
NextCentroids = cell(BlockRows, BlockCols);

U = zeros(BlockRows, BlockCols);
V = zeros(BlockRows, BlockCols);

for phase=1:2;
    for BR=0:BlockRows;
        for BC=0:BlockCols;

            %reset for array space count
            hitcount = 0;
            for y=1:blockdim; %rows are on y-axis
                for x=1:blockdim; %columns are on x-axis
                    %if the current pixel is 'on', augment hitcount
                    if (phase == 1);
                        if (((y + (BR * blockdim)) <= size(binarycurrent,1)) && ((x + (BC * blockdim)) <=size(binarycurrent,2)));
                            Xtest = y + (BR * blockdim);
                            Ytest = x + (BC * blockdim);
                            Ztestvalue = binarycurrent(y + (BR * blockdim), x + (BC * blockdim));
                            if (binarycurrent(y + (BR * blockdim), x + (BC * blockdim)) == 1);
                                hitcount = hitcount + 1;
                            end;
                        end;
                    else
                        if (((y + (BR * blockdim)) <= size(binarynext,1)) && ((x + (BC * blockdim)) <= size(binarynext,2)));
                            if (binarynext(y + (BR * blockdim), x + (BC * blockdim)) == 1);
                                hitcount = hitcount + 1;
                            end;
                        end;
                    end;
                end;
            end;

            %don't bother doing anything if there are no active pixels in this block
            if (hitcount > 0);

            %now that hitcount is known, preallocate the arrays for this block
            xCoords = zeros(hitcount,1);
            yCoords = zeros(hitcount,1);
            %reset hit counter
            hitcount = 1;
               for y=1:blockdim; %rows are on y-axis
                    for x=1:blockdim; %columns are on x-axis
                        %if the current pixel is 'on', add its coordinates to X & Y coord arrays
                        if (phase == 1);
                            if (((y + (BR * blockdim)) <= size(binarycurrent,1)) && ((x + (BC * blockdim)) <=size(binarycurrent,2)));
                                if (binarycurrent(y + (BR * blockdim), x + (BC * blockdim)) == 1);
                                    xCoords(hitcount) = x - CoordCorrection;
                                    yCoords(hitcount) = -y + CoordCorrection; %addition to flip y-axis
                                    hitcount = hitcount + 1;
                                end;
                            end;
                        elseif (phase == 2);
                            if (((y + (BR * blockdim)) <= size(binarynext,1)) && ((x + (BC * blockdim)) <= size(binarynext,2)));
                                if (binarynext(y + (BR * blockdim), x + (BC * blockdim)) == 1);
                                    xCoords(hitcount) = x - CoordCorrection;
                                    yCoords(hitcount) = -y + CoordCorrection; %addition to flip y-axis
                                    hitcount = hitcount + 1;
                                end;
                            end;
                        end;

                    end;
                end;

                %Obtain centroid for this search block (using cell arrays)
                xCenter = 0;
                yCenter = 0;

                for x=1:size(xCoords, 1);
                   xCenter = xCenter + xCoords(x);
                end
                for y=1:size(yCoords, 1);
                   yCenter = yCenter + yCoords(y);
                end
```

```
                %store x & y components of current image block's centroids
                %the 1 or 0 in the 3rd place is intended to indicate whether a
                %vector of [0,0] was obtained by all 1s, or all 0s
                if (phase == 1);
                    CurrentCentroids(BR + 1, BC + 1) = {[(xCenter / size(xCoords,1)) (yCenter / size(yCoords,1)) 1]};
                elseif (phase == 2);
                    NextCentroids(BR + 1, BC + 1) = {[(xCenter / size(xCoords,1)) (yCenter / size(yCoords,1)) 1]};
                end;
            else
                if (phase == 1);
                    CurrentCentroids(BR + 1, BC + 1) = {[0 0 0]};
                elseif (phase == 2);
                    NextCentroids(BR + 1, BC + 1) = {[0 0 0]};
                end;
            end;
        end;
    end;
end;

%Calculate Motion Vectors
for BR=1:BlockRows;
    for BC=1:BlockCols;
        %U stores X component & V stores Y component
        if (CurrentCentroids{BR,BC}(1) == 0 && CurrentCentroids{BR,BC}(3) == 0);
            %invert the vector if currentcentroid is 0
            U(BR, BC) = -NextCentroids{BR, BC}(1) - CurrentCentroids{BR,BC}(1);
        elseif (NextCentroids{BR, BC}(1) == 0 && NextCentroids{BR,BC}(3) == 0);
            U(BR, BC) = NextCentroids{BR, BC}(1) - -CurrentCentroids{BR,BC}(1);
        else
            U(BR, BC) = NextCentroids{BR, BC}(1) - CurrentCentroids{BR,BC}(1);
        end;

        if (CurrentCentroids{BR,BC}(2) == 0 && CurrentCentroids{BR,BC}(3) == 0);
             V(BR, BC) = -NextCentroids{BR, BC}(2) - CurrentCentroids{BR,BC}(2);
        elseif (NextCentroids{BR, BC}(2) == 0 && NextCentroids{BR,BC}(3) == 0);
            V(BR, BC) = NextCentroids{BR, BC}(2) - -CurrentCentroids{BR,BC}(2);
        else
             V(BR, BC) = NextCentroids{BR, BC}(2) - CurrentCentroids{BR,BC}(2);
        end;
    end;
end;

%the matrices are upside-down, this is corrected with the flip(up-down) function
FlipU = flipud(U);
FlipV = flipud(V);

%count the number of non-zero vectors in (U, V)
entries = 0;
for x=1:size(FlipU,1);
    for y=1:size(FlipU, 2);
        if ((FlipU(x,y) ~= 0) || (FlipV(x,y) ~= 0));
            entries = entries + 1;
        end;
    end;
end;

MovementDir = zeros(entries, 3);
entries = 1;
%calculate the movement angles (based on the sine of the vector)
for x=1:size(FlipU,1);
    for y=1:size(FlipV,1);
        if ((FlipU(x,y) ~= 0) || (FlipV(x,y) ~= 0));
            MovementDir(entries, 1) = abs(asin(FlipV(x,y) / sqrt(FlipU(x,y)^2 + FlipV(x,y)^2)));
            MovementDir(entries, 2) = MovementDir(entries, 1) * (180 / pi);
            MovementDir(entries, 3) = sqrt(FlipU(x,y)^2 + FlipV(x,y)^2);
            entries = entries + 1;
        end;
    end;
end;

Bins = (0:pi/14:(pi / 1.9));

WeightedBins = zeros(length(Bins) - 1,1);
for x = 1:length(Bins) - 1;
    %Numerator = 0;
    %Denominator = 0;
    for y = 1:length(MovementDir(:,1));
        if (MovementDir(y,1) >= Bins(x) && MovementDir(y,1) <= Bins(x + 1));
            %Numerator = Numerator + (MovementDir(y,1) * MovementDir(y,3));
            %Denominator = Denominator + MovementDir(y,3);
            WeightedBins(x, 1) = WeightedBins(x, 1) + MovementDir(y,3);
        end;
    end;
end;
OutBar = WeightedBins;
```

APPENDIX D

Left vs. Right Aryepiglottic Motion Vector Plots



APPENDIX E

Geometric feature measurements of synthesized articulations

Type	State	CoT	CuT	C-ET		IA-ET	VP	AE area	VentF area	VF area
				Left	Right					
Real	Inspiration	100% (87)	100% (126)	100% (127)	100% (130)	100% (143)	100% (62)	100% (18756)	100% (6609)	100% (3431)
	Glottal Stop	14% (13)	50% (64)	33% (43)	57% (75)	48% (69)	0% (0*)	24% (4678)	0% (0*)	0% (0*)
Synthesized	Inspiration	100% (38)	100% (66)	100% (150)	100% (150)	100% (183)	100% (83)	100% (20130)	100% (11112)	100% (6177)
	Glottal Stop	0% (0)	22% (15)	51% (77)	51% (77)	66% (121)	0% (0)	29% (6000)	0% (0)	0% (0)
	Glottal Fricative	0% (0)	22% (15)	66% (99)	66% (99)	75% (139)	32% (27)	42% (8507)	29% (3318)	27% (1720)
	AE Stop	0% (0)	0% (0)	6% (10)	6% (10)	13% (52)	0% (0)	0% (0)	0% (0)	0% (0)
	AE Fricative	0% (0)	31% (21)	14% (22)	14% (22)	47% (86)	32% (27)	12% (2440)	15% (1725)	20% (1286)
	AE Approximant	0% (0)	16% (11)	13% (20)	13% (20)	13% (85)	0% (0)	6% (1382)	3% (409)	0% (0)
	Modal Voice	0% (0)	22% (15)	55% (83)	55% (83)	79% (146)	0% (0)	43% (8752)	22% (2459)	0% (0)
	Falsetto	0% (0)	21% (14)	79% (119)	79% (119)	93% (172)	0% (0)	60% (12244)	28% (3127)	0% (0)
	Breath	0% (0)	27% (18)	57% (86)	57% (86)	81% (149)	40% (34)	53% (10766)	46% (5201)	39% (2412)
	Breathy Voice	0% (0)	19% (13)	60% (91)	60% (91)	78% (143)	13% (11)	46% (9460)	31% (3500)	9% (584)
	Whisper	13% (5)	37% (25)	20% (31)	20% (31)	51% (95)	33% (28)	14% (2832)	13% (1486)	21% (1339)
	Whispery Voice	0% (0)	21% (14)	9% (14)	9% (14)	43% (80)	10% (9)	14% (2868)	4% (442)	3% (227)
	Creaky Voice	0% (0)	18% (12)	24% (37)	24% (37)	53% (98)	0% (0)	25% (5068)	4% (439)	0% (0)
	Harsh Voice (low)	0% (0)	18% (12)	8% (12)	8% (12)	32% (59)	0% (0)	1% (163)	0% (0)	0% (0)
	Harsh Voice (mid)	0% (0)	39% (26)	21% (32)	21% (32)	50% (92)	0% (0)	10% (2158)	2% (197)	0% (0)
	Harsh Voice (high)	0% (0)	15% (10)	48% (73)	48% (73)	73% (135)	0% (0)	33% (6678)	5% (567)	0% (0)