The Voice Box: A Fast Coupled Vocal Fold Model for Articulatory Speech Synthesis

by

Arvind Vasudevan

B.Tech, Electrical and Electronics Engineering, S.R.M. University, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2017

Abstract

Speech is unique to human beings as a means of communication and many efforts have been made towards understanding and characterizing speech. In particular, articulatory speech synthesis is a critical field of study as it works towards simulating the fundamental physical phenomena that underlines speech. Of the various components that constitute an articulatory speech synthesizer, vocal fold models play an important role as the source of the acoustic simulation. A balance between the simplicity and speed of lumped-element vocal fold models and the completeness and complexity of continuum-models is required to achieve time-efficient high-quality speech synthesis. In addition, most models of the vocal folds are seen in a vacuum without any coupling to the vocal tract model.

This thesis aims to fill these lacunae in the field through two major contributions. We develop and implement a novel self-oscillating vocal-fold model, composed of an 1D unsteady fluid model loosely coupled with a 2D finite-element structural model. The flow model is capable of handling irregular geometries, different boundary conditions, closure of the glottis and unsteady flow states. A method for a fast decoupled solution of the flow equations that does not require the computation of the Jacobian matrix is provided. The simulation results are shown to agree with existing data in literature, and give realistic glottal pressure-velocity distributions, glottal width and glottal flow values. In addition, the model is more than order of magnitude faster than comparable 2D Navier-Stokes fluid solvers while better capturing transitional flow than simple Bernoulli-based flow models.

Secondly, as an illustrative case study, we implement a complete articulatory speech synthesizer using our vocal fold model. This includes both lumped-element and continuum vocal fold models, a 2D finite-difference time-domain solver of the vocal tract, and a 1D tracheal model. A clear work flow is established to derive model components from experimental data or user-specified meshes, and run fully-coupled acoustic simulations. This leads to one of the few complete articulatory speech synthesizers in literature and a valuable tool for speech research to run time-efficient speech simulations, and thoroughly study the acoustic outcomes of model formulations.

Lay Abstract

Building a speech synthesizer that can generate natural-sounding high quality speech has been a long-term goal. One of the critical components of these synthesizers is a model of the vocal folds, colloquially known as the vocal cords. These are two fleshy slits inside our larynx that vibrate quasi-periodically, and act as the sound source for speech. This buzzing sound (called phonation) then propagates through our vocal tract, which can take different shapes, giving rise to speech sounds at the outlet of our mouth. Most previous models either made coarse approximations of the complex vocal fold structures (as rectangular interconnected masses) or were mathematically intricate and computationally difficult (taking days to simulate one second of speech). In this thesis, we implement and validate a 2-dimensional vocal fold model combined with an 1-dimensional airflow model that aims to find a balance between complexity and computational expense for speech synthesis applications.

Preface

This thesis was part of the Oral, Pharyngeal and Laryngeal Complex (OPAL) project.

Most of the contributions and results described in Chapter 3 have been previously presented in the publication [P1]. I was the main author and contributor to the design, implementation and testing of the vocal fold model described in the [P1], under the supervision of Dr. Sidney Fels. Dr. Victor Zappi assisted with the implementation of the 2D finite-difference time-domain (FDTD) simulation that was used for the vocal tract coupling. Dr. Peter Anderson helped with the validation of the 1D unsteady fluid model that was used to drive the vocal folds.

Chapter 4 has been partially published in the literature [P2]. I developed implementations of two exemplar lumped-element vocal fold models, the two-mass and body-cover model, on the Graphical Processing Unit shader. This enabled coupling with the real-time GPU-based vocal tract solver designed by Dr. Victor Zappi, who is the primary author of this paper. Dr. Andrew Allen and Dr. Nikunj Raghuvanshi assisted with the GPU-based implementation of the 2D wave-equation solver.

Peer-Reviewed Conference Paper Accepted for Publication

[P1] **Vasudevan A**, Zappi V, Anderson P, Fels S, 2017. A Fast Robust 1D Flow Model for a Self-Oscillating Coupled 2D FEM Vocal Fold Simulation. INTERSPEECH. (Accepted)

Journal Manuscripts Accepted for Publication

[P2] Zappi V, **Vasudevan A**, Allen A, Raghuvanshi N, Fels S, 2067. Towards real-time two-dimensional wave propagation for articulatory speech synthesis. Proceedings of Meetings on Acoustics (POMA). (Accepted)

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1D: One Dimensional
2D: Two Dimensional
3D: Three Dimensional
BC/BCs: Boundary Condition/Boundary Conditions
CFD: Computational Fluid Dynamics
CFL: Courant—Friedrich—Levy Condition
CSA: Cross Sectional Area
CT: Computed Tomography
EGG: Electroglottography
FEM: Finite Element Method/ Finite Element Model
FVM: Finite Volume Method
FFT: Fast Fourier Transform
FSI: Fluid Structure Interaction
MRI: Magnetic Resonance Imaging
OSA: Obstructive Sleep Apnea
Pa: Pascal (unit of pressure)
UA: Upper Airway
VF: Vocal Folds
VT: Vocal Tract

Acknowledgements

I express my deepest gratitude to my supervisor, Dr. Sidney Fels; this thesis would have just remained an idea without his guidance, mentorship and above all, patience. His belief in my ability and the academic freedom that he provided, are aspects of my Master's that I will always cherish. I have learned valuable lessons on staying hungry, being humble and having a sense of humour that I will take forward in my career; it has been an honour working with you and learning from you.

I've had the pleasure of working with a number of fine researchers over the course of this degree, foremost being Dr. Victor Zappi, who was friend, guide and all-round amazing co-worker. I truly enjoyed the times we spent breaking our heads over the most intractable of problems, and the lessons I learned about perseverance and perspective from them. Many thanks to Dr. Peter Anderson, who apart from being an extremely distinguished researcher, ranks among the nicest people I've met.

The great memories I take of this degree is due, in part, to my fellow colleagues at the Human Communication Technologies (HCT) Lab and the OPAL modelling group. I thank them for the great conversation and even better company; they made every day (and some nights!) at the lab, truly enjoyable.

I'm lucky to have some truly amazing friends and family who've given me incredible love and support through my degree; you guys are truly one of a kind. Finally, last but definitely not least, what more can I say about my parents except that, this isn't my achievement so much as it's ours! Thank you for everything.

Dedication

To my parents, for being you

To science, for being magical

Chapter 1

Introduction

Speech is a form of communication that is unique to human beings, and has been a topic of intensive study for many centuries. Through the years significant efforts have been made towards building systems for speech recognition and more importantly, speech synthesis. Of the different methods that have been proposed for speech synthesis, articulatory speech synthesis is one of the most challenging and promising. Articulatory synthesis attempts to simulate the physiological processes and physics that occur in the human body to generate speech output. This involves creating anatomically accurate models of the upper airway, a biomechanical simulation of articulation movements and an acoustic model representing the air wave production and propagation through the vocal tract.

The flow of pressurized air from the lungs to the mouth-opening generates sound and, as a consequence, speech. To simulate speech, a critical component that all articulatory synthesizers require is a glottal model that acts as the excitation source to the acoustic simulation. The glottis is the slit-like opening between the two fleshy vocal folds within the larynx. The passing of air from the trachea into the upper airway is controlled by the vocal folds that vibrate periodically in a process known as phonation. This is the source of sound into the vocal tract, and it's position within the larynx gives the latter its colloquial name, the voice box.

Initially, these models were based on the linear *source-filter theory* that posits that speech production is a two-step process: a sound-source is generated from the glottis, and articulators in the vocal tract shape this waveform similar to a resonant filter [42]. However, the underlying assumption that this model is predicated on, that the source and filter are independent of each other, has been challenged in recent work. It is now well-established that the sound-source from the glottis and filtering by the vocal-tract articulators are non-linearly coupled [119]. Thus, improved glottal models that can capture this phenomenon better are critical to achieving high quality articulatory speech synthesis.

There are a multitude of vocal fold models in the literature, implemented with different combinations of structural and flow models. Of these, there are two main classes of models: lumped-element and continuum models. Lumped-element models are conceptually and computationally simple, representing the vocal fold structure through lumped-masses interconnected through springs. Continuum models on the other hand, solve the internal continuum mechanics of the vocal fold structure through numerical methods such as the finite-element method. This provides a rich and complex characterization of vocal fold vibration, but are computationally extremely expensive and are often unstable when run as part of a larger acoustic simulation. Thus, articulatory synthesizers have been forced to using lumped-element models, despite the possibilities of higher-quality synthesis and physical realism offered by continuum models. A balance between the simplicity and speed of the lumped-element vocal fold models and the completeness and complexity of continuum models is required for speech synthesis.

Equally, surveys of the field have noted that while multiple models exist, each with their unique model formulations, there is a lack of understanding of how these modelling decisions translate to acoustic outcomes [26]. Two issues in particular have exacerbated this problem: firstly, most models are built using proprietary or commercial finite-element modelling toolkits which makes it extremely difficult for speech researchers to compare models. Secondly, many of the continuum vocal fold models can take days to simulate a short portion of acoustic output. When combined with an acoustic solver for the vocal tract, this makes it practically impossible for researchers to run multiple simulations and do comparative analyses. The present thesis investigates a solution to these issues through two main pathways: firstly, through developing and implementing a novel self-oscillating vocal fold model that attempts to find a balance between computational cost and complexity. This model is then used to drive a 2D finite-difference time-domain (FDTD) simulation of the vocal tract, as part of a coupled stable acoustic simulation to demonstrate the feasibility of the model as a potential tool for speech researchers.

1.1 Contributions

This thesis targets a specific lacunae in the field of articulatory speech synthesis, and vocal fold modelling in particular. The contributions of the thesis are given below:

Implemented and validated a time-efficient vocal fold model for articulatory speech synthesis

- We identified that fluid simulation of glottal flow remains a major roadblock towards creating time-efficient and complex continuum models of the vocal folds for articulatory speech synthesis.

- We put forward a novel solution framework of the 1D unsteady Navier-Stokes equations, for a time-efficient decoupled numerical solution. This solution does not required the computation of the Jacobian, and is faster than the coupled simulation as well as and other 2D Navier-Stokes models.

- A novel vocal fold model was built comprising of a 2D finite-element based structural model, loosely coupled with the 1D unsteady flow model.

- The model performance was compared with published literature and experimental data, to validate its efficacy for a set of standard tests.

Demonstrated the potential of the proposed model in building an articulatory speech synthesizer

We demonstrated the feasibility of our vocal fold model by building an illustrative articulatory speech synthesizer. The synthesizer included the vocal fold model, a 2D finite-difference time-domain vocal tract acoustic solver and a model of the trachea. A sample framework is included to drive the model using different geometries. Coupled acoustic simulations are performed for two vowel shapes derived from literature as an illustrative example.

Chapter 2

Background and Previous Work

With the increase in simulation and computational capabilities, articulatory speech synthesis has once again become an active area of research. Through articulatory speech synthesis, researchers can gain insight into a more fundamental understanding and characterization of speech that no other method offers [90]. Articulatory synthesis also allows for the possibility of building natural looking graphical talking-head models, where the biomechanics can drive the acoustics of the system [23]. Finally, better models of the human body for speech will provide clinicians with insights to help predict pathologies and functional outcomes of surgical interventions [133] [116]. For these reasons, articulatory speech synthesis is a challenging, but promising, area of research.

This chapter provides a bird's-eye view of the field, reviewing advances in vocal fold models in the context of developments in articulatory synthesis. Section 2.1 provides a self-contained overview of the physiology of human phonation and a detailed description of the exemplar theories of phonation. As these theories facilitated a more thorough understanding of the process of human phonation, many models of the vocal folds were created. These started with lumped-element models that represented the vocal fold structure through masses interconnected with springs and dampers. These masses are driven by airflow to give rise to the oscillation that is characteristic of vocal fold vibration. Current advancements in computational capabilities have made it possible to create more complex models that solve the continuum mechanics of the vocal fold structure and the airflow driving it. These models give rise to Partial Differential Equations (PDE's), that are solved using numerical techniques such as the Finite-Element Method (FEM) or Finite-Volume Method (FVM).

Section 2.2 reviews the wealth of lumped-element models and identifies their strengths, limitations and areas of improvements. The following Section 2.3 does the same for continuum models solved using PDE's. For both groups of models, the different types of flow models used to drive the structural models are highlighted and discussed. While the sheer number of models are overwhelming, there is a clear pattern to the model combinations that have made in literature. This is explored in section 2.4, which isolates exemplar structural and flow models. By looking at these models in isolation, we can better understand what gap exists in models of the field and how we can potentially fill it. A particularly interesting sub-problem in the field is the handling of collisions which is covered in subsection 2.4.2 Finally, subsection 2.4.3 talks about the coupling between the vocal folds and vocal tract, and how this is handled by models in literature.

The ultimate goal of speech mechanics is to generate sound. Building an articulatory speech synthesizer requires many different model components to work together in tandem: biomechanical models to represent the anatomy of the vocal tract and movement of articulators, acoustic solvers for the vocal tract and a glottal excitation to act as the source to the simulation. While there are a range of individual components, complete articulatory synthesizers are rare due to the complexity inherent in building an entire system. Section 2.5 provides an overview of the different acoustic and biomechanical models in literature, and reviews the current trend in articulatory synthesis.

Finally, advancements in data acquisition techniques have enabled researchers to gain improved knowledge on observing and quantifying articulatory processes. As a thumb rule, experimental measurements of the glottis are difficult due to the general inaccessibility of the vocal folds. This requires specialized instrumentation, or even fabrication studies in lieu of direct measurements. Medical imaging methods such as Magnetic Resonance Images (MRI) and Computed Tomography (CT) scans have enabled clearer visual depiction of oropharyngeal structures in general, that complement other acoustic recordings. Section 2.6 gives an insight into the experimental data in literature and the methods used to obtain them.

2.1 Overview of Human Phonation

Human phonation is a complex multi-step process that requires an understanding of the anatomy of the human airway and the physics that leads to phonation. In the following sections, we shall briefly look at the anatomy of the upper airway, with specific attention paid to the vocal folds. Then we shall understand the various theories of phonation, and what lessons we learn from them in designing better models of the vocal folds.

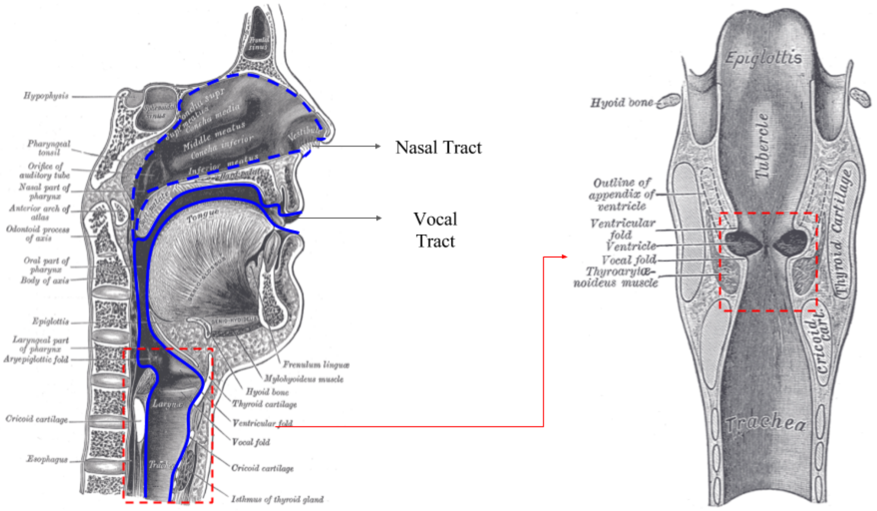


Figure 2.1: Basic Anatomy of the Human Airway: Saggital view of the head (left) and enlarged coronal view of the laryngeal complex (right), ©Wikimedia Commons, adapted from [102]

2.1.1 Physiology of the Human Airway

Figure 2.1 details the basic physiological structure of the human airway. The periodic expansion and contraction of the lungs leads to the expelling of pressurized air through the trachea, also known as the windpipe. At the end of the trachea is the larynx which connects it to the vocal tract (highlighted in blue in figure 2.1) and contains the vocal folds (commonly known as the vocal cords). The vocal tract splits at the soft palate to lead to the nasal tract above (highlights in dotted blue lines in figure 2.1). The vocal fold regulates the flow of air from the trachea to the vocal tract in a periodic manner. The upper airway contains multiple articulators that can be individually controlled to take up multiple poses; this modifies the sound from the glottis.

The vocal folds are contained inside the larynx complex (highlighted in dotted red lines in figure 2.1), which includes various adjustable muscles, ligaments and cartilage. The larynx is suspended between the trachea below and pharyngeal complex leading to the vocal tract above [51]. It is important to note that the glottis forms a constriction in the airway tube. The ventricular folds, also known as the false vocal folds, can also be seen above the true vocal folds. They are two thick folds of mucosal membrane

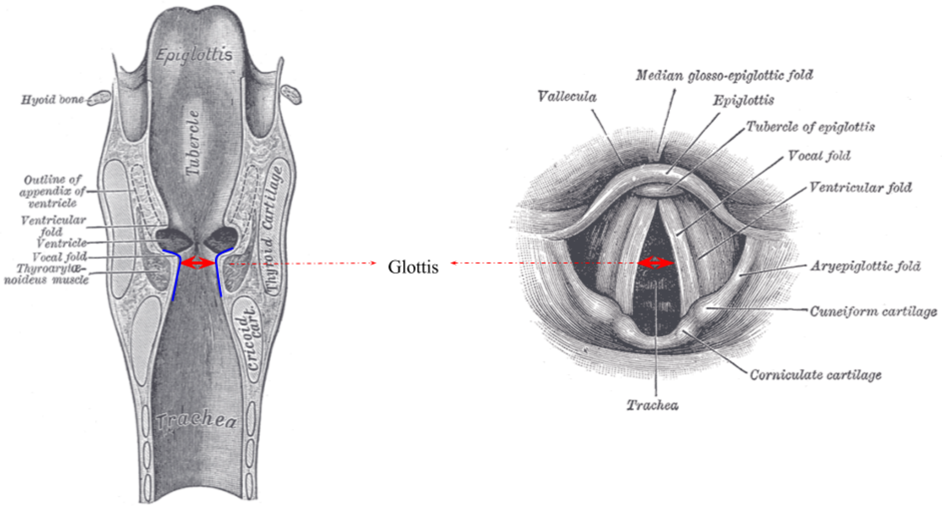


Figure 2.2: Coronal (left) and superior (right) view of the laryngeal complex (right), 🅭Wikimedia Commons, adapted from [102]

with narrow bands of fibrous tissues [102]. They generally do not play a major role in primary phonation, but can be used for phonation in certain forms of singing and for patients who lose the ability to phonate with their true vocal folds.

2.1.2 Vocal Fold Structure

Figure 2.2 shows a closeup coronal view and superior view of the larynx, with the glottis between the vocal fold highlighted. The vocal folds can be generally assumed to be symmetric about the glottal mid-line (highlighted in the figure 2.2); this approximation is not valid in the case of vocal fold pathologies that causes polyps to form on individual vocal folds. The vocal folds can be adducted (brought together) and abducted (taken apart), by the muscles of the larynx. The latter creates a triangular slit opening that is called the glottis. The vocal folds are attached posteriorily to the arytenoid cartilages and anteriorily to the thyroid cartilage. The vocal folds are stretched horizontally from back to front across the larynx, similar to a rubber band. Thus, when air flows over its surface after it is expelled from the lungs, it vibrates similar to a stretched membrane. This rhythmic opening and closing of the vocal folds create the buzzing sound known as phonation.

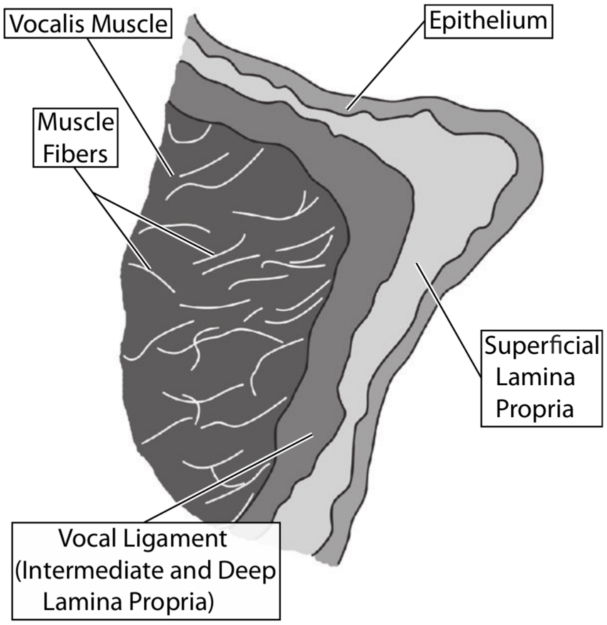

Vocalis Muscle
Epithelium
Muscle Fibers
Superficial Lamina Propria
Vocal Ligament (Intermediate and Deep Lamina Propria)

Figure 2.3: Layers of the vocal fold, ©Wiley Publishing, Gick et al [51]

Hirano et al showed that the vocal fold is comprised of multiple layers, each with unique structural and material properties [59]. As a general rule of thumb, as we go from superficial outermost layers of the vocal folds to the core, the material becomes less flexible and more rigid. Figure 2.3 shows the coronal view of the internal structure of the vocal folds. Medially to laterally, the vocal fold is made of multiple layers including the epithelium and the lamina propia before leading to the vocalis muscle. The lamina propia itself can be divided into three constituent parts: superficial, intermediate and deep. Hirano et al also put forward the *Body-Cover Theory* that classified the layers into two main groups: the cover and the body. Figure 2.4 explains the correlation between the layers to the cover-body theory. The loose cover enables the propagation of the vertical travelling wave that is critical to self-sustained vocal fold oscillation and phonation. The cross-section of the vocal folds varies in the anterior-posterior direction and is not uniform.

2.1.3 Theories of Phonation

The process of phonation has fascinated speech researchers from the late 18th century, with a range of theories proposed to explain the self-oscillation

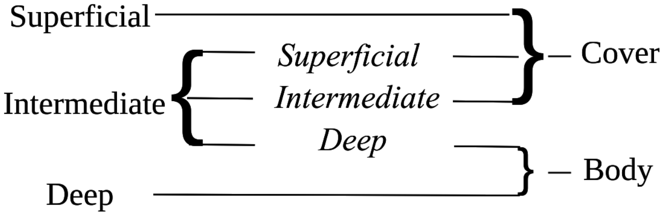


Figure 2.4: Composition of the Vocal Folds according to Body-Cover theory [59]

of the vocal folds. The most commonly accepted theory today is the *Myoelastic Aerodynamic theory of Phonation* [125], which was augmented by the *Body-Cover Theory* that was previously mentioned. The former theory is a amalgamation of two concepts: the *myoelastic* nature of the vocal folds, coupled with the *aerodynamics* of the flow passing through the glottis. With the compression of the lungs, air flows through the trachea to the closed glottis. This continuously increases the subglottal pressure (P_{sg}) that acts against the bottom surface of the vocal folds. At a certain threshold pressure called the *onset pressure*, the subglottal pressure overcomes the elastic properties of the vocal folds, and forces the vocal folds open.

It is important to note here that the glottis still represents a constriction in the upper airway. Bernoulli's principle along with the laws of *conservation of mass* and *conservation of energy* tells us that the airflow through a constriction is faster, with an associated drop in pressure inside the constriction. This negative air pressure difference combined with the elastic nature of the vocal folds pulls the individual folds back together, thus closing the glottal opening. Figure 2.5 shows the different poses taken up by the vocal folds from the coronal view, during a cycle of phonation. This is opening-closing pattern is repeated continuously giving rise to a periodic glottal signal. It is important to note the importance that the transfer of energy between the flow and the vocal fold structure plays in the self-oscillation process. The vertical travelling wave also seen in the figure 2.5 is a representation of this process.

Later work, extended the *myoelastic aerodynamic theory* to investigate the mucosa that coats the vocal fold surface, challenging the underlying laminar flow assumption of the theory. Another seminal work, was Ishizaka et al's work on the *flow separation theory*, which looked at the important role of turbulences and eddies in the phonation process. Flow models that can capture these aspects of the flow can help to better characterize the glottal

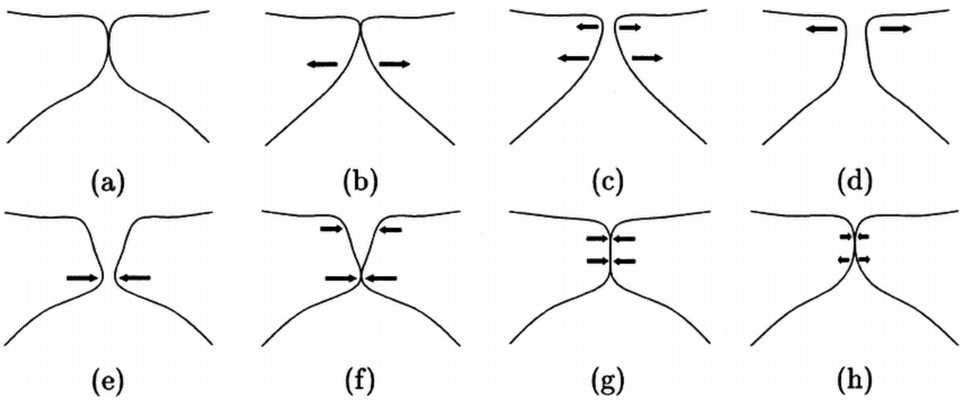


Figure 2.5: Idealized shapes of the vocal fold during a single cycle of vibration. Note that the lower part of the vocal fold leads the upper part and the vertical travelling wave on the vocal fold surface. Reproduced with permission from Story, [105]

flow, and help with potentially achieving better acoustic output. Readers are referred to [51] for a non-technical introduction to the theories of phonation, and different types of phonation.

The vibratory behaviour of the vocal folds can be also modified by the activation of the intrinsic and extrinsic muscles of the vocal folds. The properties that affect vocal fold motion are the tissue mass, stiffness and viscosity. The constriction of the cricothyroid muscle can alter vocal fold tension while the thyroarytenoid muscle allows changing of the internal stiffness [58]. The fundamental frequency of oscillation is a function of the current properties of the vocal folds; thus, creating models that connect muscle activation to the material properties of the vocal folds will be useful in achieving different voice registers. For a more in-depth discussion of phonatory control and vocal fold physiology, the reader is refered to the paper by Jiang et al [67].

2.2 Lumped-Element Models of the Vocal Folds

The earliest models of the vocal folds were those that approximated the vocal fold structure as lumped masses interconnected by springs and dampers. The section looks at some seminal models that have shaped the field over the past 50 years. For a more comprehensive survey of the different models, organized by mechanical design, aerodynamic simulation and application, we refer the reader to the survey by Birkholz et al [26].

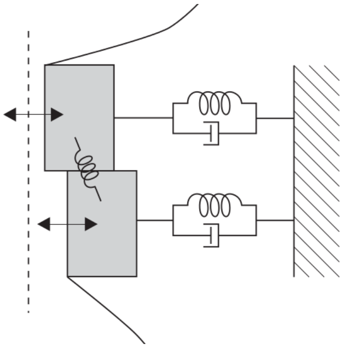


Figure 2.6: Coronal view of the two-mass model. Reproduced with permission from Peter Birkholz, [26]

2.2.1 One-Mass Models

The earliest model of vocal folds was the simple one-mass block model introduced in 1968 by Flanagan and Landgraf [45]. This model consisted of a single mass, that had 1 degree of freedom. The model could oscillate in the medial-lateral direction in the coronal plane. However, to obtain a true self-oscillating vocal fold model, the system requires the constant imparting of energy to compensate for internal friction losses. This is achieved through asymmetric pressure loading, something that a one-mass model cannot do due to its time invariant glottal orientation. This model achieves self-oscillations through acoustic loading of the supraglottal tract; later studies by Zanuartu et al [131] analyzed the importance of the acoustic loading by using a non-square geometry for the vocal fold.

The lack of time-varying geometry remained the major issue with this model along with the inability to replicate experimentally observed phase differences between the inferior and superior edge of the vocal folds. There were attempts made to introduce an extra degree of freedom, in the translational and rotational directions [74][62] and more recently in the parallel-perpendicular direction of the airflow [1]. Almost all these models uses a flow model that based on a simple steady Bernoulli principle flow model. The exception is the work by Horacek et al [62] which use the unsteady 1D Euler equation to solve the pressure profile and flow. However, due to their significant drawbacks, one-mass models are no longer used in most application areas and are not promising as a potential model for speech synthesis.

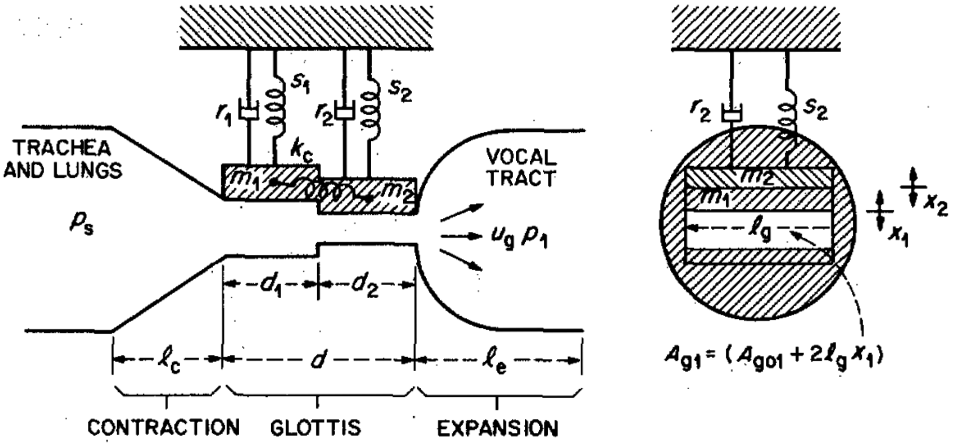


Figure 2.7: Coronal (left) and superior (right) view of the vocal folds with quantities highlighted. Ishizaka et al, [64]

2.2.2 Multi-Mass Model

Multi-Mass models added extra mass elements to each vocal fold to overcome some of the drawbacks of the single-mass model. The classic two-mass model introduced by Ishizaka and Flanagan [64] added a second mass superior to the one-mass model (shown in fig.2.6), while restricting motion to medial-lateral translation. While the airflow fluid loading occurs only on the lower mass, this configuration allows for the experimentally observed phase difference between the inferior and superior edges of the vocal folds. Due to the importance of this model in terms of understanding the field of vocal fold modeling better, let us briefly look at the mechanics behind this model.

Two-Mass Model

Figure 2.7, shows the different quantities involved in the two-mass vocal fold structure in the coronal view. The displacements of the two vocal folds define the vocal fold shape, and as a consequence, the flow at every time instant. The lateral displacements x_1 and x_2 are the solution to the differential equation:

$$m_1\ddot{x_1} + r_1\dot{x_1} + s_1 + k_c(x_1 - x_2) = f_1 \tag{2.1}$$

$$m_2\ddot{x_2} + r_2\dot{x_2} + s_2 + k_c(x_2 - x_1) = f_2 \tag{2.2}$$

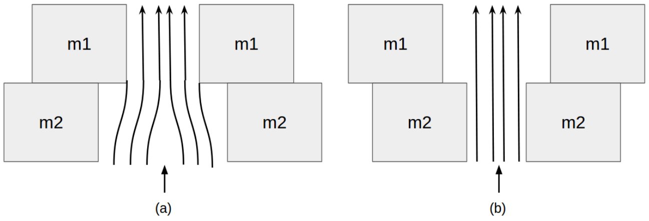


Figure 2.8: Bernoulli flow in the two-mass model in two cases a) Bernoulli flow in lower region and jet flow in upper region b) jet flow in both lower and upper regions

where m_1, m_2 are the masses, r_1, r_2 are the damping values, s_1, s_2 are the restoring spring forces and k_c is the stiffness of the coupling spring. f_1, f_2 are the aerodynamic forces that act on the vocal folds and varying based on the flow model used. The simplest model to represent flow is the well-known Bernoulli's equation. For a steady, inviscid, incompressible flow it can be written as:

$$p + \frac{1}{2}\rho u^2 + \rho g z = \text{const} \tag{2.3}$$

where u is the flow speed in a point of the streamline, p is the pressure at the chosen point, ρ is the density of the fluid, g is acceleration due to gravity and z is the elevation with respect to the reference 1D plane.

Equation 2.3 can be seen as the statement of conservation of energy-momentum for a fluid, and implies the classic *Bernoulli's Effect* when used in conjunction with the statement of mass conservation ($Au = \text{const}$). This would mean a reduction of pressure and increase in speed when a fluid goes through a tube constriction. The most-common implementation of the flow differs from the original paper by [64], and is based on a modified version of Bernoulli-based flow. As shown in figure 2.8, steady Bernoulli flow is assumed till the point of minimum glottal diameter at which flow is assumed to detach. A constant diameter jet exists in the region from the flow-separation point till the glottal exist, where pressure is assumed to be constant. A pressure recovery after the glottal exit is also included in the equation. This simplifies to the following equation:

$$P_b = P_s - (P_s - P_i)(a_m/a)^2 \tag{2.4}$$

$$P_j = P_i \tag{2.5}$$

where P_b is the pressure in the Bernoulli regime and P_j is the pressure in the jet regime. Here, P_s and P_i are the subglottal and supraglottal (epilaryngeal) pressures respectively. a_m is the minimum glottal area and a is the area at the location we are calculating the pressure at. The glottal flow U_g can now be calculated based on the glottal shape and the pressures incident on the vocal folds.

This model has been one of the most used VF models and many variations have been proposed [17] including models with added asymmetries in the masses [103]. An additional vertical DOF was added to the model to account for inferior-superior vocal fold motion in later models [46][63]. However, this model fails to simulate the difference in registers and transitions between them; to this end, a third superior mass was added to solve this issue [122]. While this helped the two-mass model better simulate different voice qualities, the inability of the model to simulate the relative closure and the smooth continuum between registers remains a drawback in it's use for speech synthesis. Finally, Titze et al [121][117] extended the model, allowing for control through muscle activation as well. However, while the low-dimensional nature of the lumped-element model is attractive for control, there is less physiological accuracy when assigning material properties to these models.

A seminal model for articulatory synthesis was the triangular glottis model was introduced by Birkholz et al [25]. The model used inclined masses (shown in fig. 2.9 to model glottal abduction and closure as part of the governing equations, to better capture different voice qualities. The model has been used as part of a complete articulatory synthesizer and shown to synthesize breathy, normal and pressed phonation types. Finally with the increase in computational capabilities, these systems have been extended to 6-mass with 3D capabilites [98], to even a 25-mass model [130]; the latter model was designed to capture more intricate geometry details for video endoscopy applications. However, the question arises whether continuum mechanics structural model would be preferable instead at such an expensive computational cost.

All the previous model use the Bernoulli-based flow model which is usually enough to characterize the flow for the purpose of lumped-element models. This requires the prior assumption or computation of the flow-separation point for the flow. The location of flow-separation has been shown to play

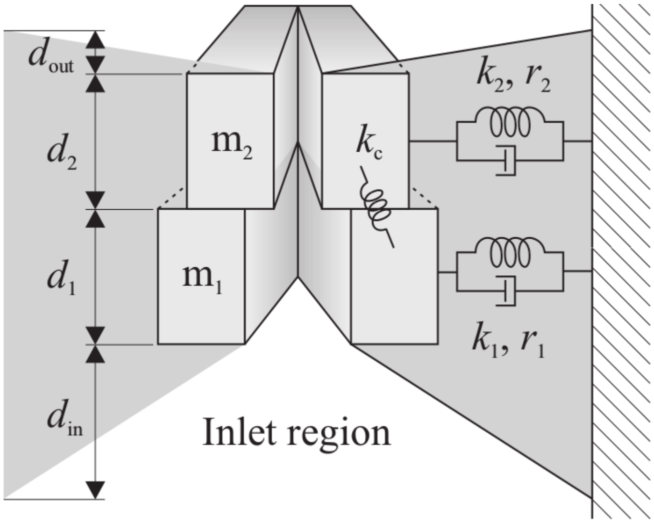


Figure 2.9: Pseudo 3D view showing the inclined masses of the modified two-mass model that can represent the triangular glottal opening, reproduced with permission from Birkholz et al [25]

a critical role in the self-oscillation process as it helps decide the pressure distribution inside the glottis and, as a consequence, the aerodynamic forces on the masses. Generally models assume that the flow separation is either at the location of smallest diameter inside the glottis or at a ratio of it (Eg. $1.2 * a_{min}$). Pelorson et al [91] put forward a model to estimate the flow separation model instead of a fixed location like most previous papers in literature. In terms of flow modelling, an significant exception is the effort by LaMar et al [72] who use a quasi-one-dimensional Euler system to drive a symmetric two-mass model. This model is pertinent as it was shown to be a more accurate treatment than the traditional Bernoulli-based systems, with a lower computational cost than full 2D Navier-Stokes solvers.

2.2.3 Body-Cover Models

As mentioned in Section 2.1, Hirano et al [59] explored the structural differences of the vocal fold layers. Story et al introduced the Body-Cover model that added an extra mass to represent the different layers and capture the cover vibration better [107]. As shown in figure 2.10, this differentiation allowed for defining varied structural properties to the body and cover, and remains one of the most popular models in literature till today. A classic

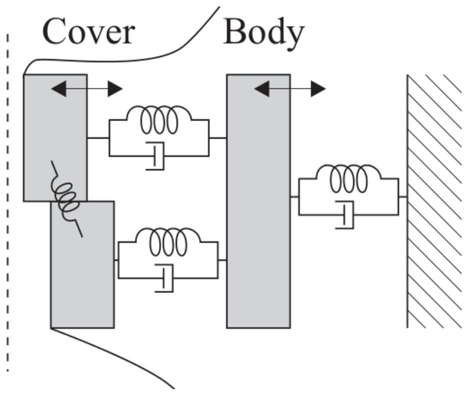


Figure 2.10: Coronal view of the body-cover model of the vocal folds, reproduced with permission from Birkholz et al [25]

Bernoulli-flow model was used to drive the outer cover masses directly, and as a consequence the inner body mass. This was later extended to a 4-mass model to study different voice registers and their transitions [123]. Finally, an extremely complex 128-mass model that explored 7 different layers with divisions in the saggital plane was introduced [118]. This system provided extensive data on differing voice types and was effective in capturing small voice distinctions.

2.3 Continuum Models of the Vocal Folds

While higher dimensional lumped-element models helped to better model the vocal folds, considerable doubt existed if the geometry and viscoelastic properties of the vocal folds adequately. Thus, continuum models of the vocal folds have been a major goal of speech synthesis for the last two decades. As efficient computing capacity evolved, the possibility of highly complex Partial Differential Equation (PDE) based models became feasible. These models are based on the fundamental laws of continuum mechanics where all the processes that contribute to phonation need to be integrated. This include the PDE equations of the aeroacoustic flow along with the biomechanical model of the vocal fold structure. This enables a direct relationship between these characteristics and the glottal waveform that is generated by the system.

2.3.1 Eigen-Analysis Models

While early continuum models failed due to improper application of boundary conditions to study of vocal-fold resonance [34], later work characterized the vibration of the vocal folds in terms of its eigenmodes and eigenfrequencies [19][20]. These empirical studies using two-dimensional models of the vocal folds. also threw up interesting findings where the eigenfunctions in studies incorporating non-linearities of aerodynamic flow was similar to those obtained from linear eigenmode systems. This suggested that, in contrast to the prevailing school of thought at the time, linear eigenmodes were of potentially greater importance in the vibration of the vocal folds than the modelling of the (possibly) non-linear aerodynamic flow. It also established the intrinsic role played by the vertical phasing eigenmode, pointing to a direct control over the opening and closing of the glottis. Structural and visco-elastic parameters were also presented for normal vibration. The latter of these studies also matches data from existing in vivo studies but however, did not include aerodynamic forces.

2.3.2 Self-Oscillating Continuum Models

While the previous studies used the continuum models to study the physiology of phonation, DeVries et al [38] used a 3D Finite Element Model (FEM) of the vocal fold structure to determine realistic parameters for a lumped two-mass model. A symmetric model with each vocal fold having 3000 elements (shown in figure 2.11) was considered with detailed descriptions of the vocal fold's material properties and geometry. By matching the dynamic behaviour of the FEM model and lumped element models for a given pressure flow model, the parameters for the lumped element model was obtained.

The first model that used the FEM for computation of vocal fold dynamics was by Alipour et al [3]. This model adds the ligament as well to the body-cover differentiation that was used by DeVries [38] when formulating the vocal fold tissue structure. The model has tissue mechanics modelled with the finite element method combined with a Bernoulli-based fluid solver. The biomechanical model is quasi 3D with the vocal folds being divided into 15 layers along the coronal plane, with figure 2.12 showing the trajectories of nodes in layer 8. The model by Alipour [3] made some simplifying assumptions:

- Small deformation and linear-elasticity
- Single plane deformation

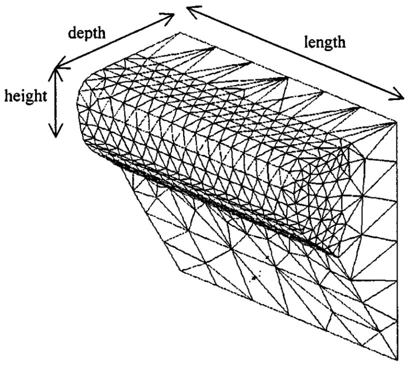


Figure 2.11: 3D view of FEM model by De Vries et al, reproduced with permission from [38]

- Transverse isotropy perpendicular to the tissue fibers
- Fixed control volume for integration

Similar to the two-mass model for the class of lumped-element models, this model also served as the template for other continuum models. We shall look at it's mechanics below. As part of the assumption of a linear material with transverse isotropy, a constitutive equation can be written as:

$$\sigma = [S]\epsilon \tag{2.6}$$

where σ is the stress tensor, ϵ is the strain tensor and $[S]$ is a *stiffness matrix.* For transverse isotropy, we have 5 independent elastic consonants out of 21 in the $6X6$ symmetric matrix $[S]$. These are the Young's modulus and Poisson's ratio (E, ν) in the plane transverse to the fibers, as well as the Young's modulus, Poisson's ratio and shear modulus (E', ν', μ'). The relationship in the transverse plane is written as:

$$\mu = \frac{E}{2(1+\nu)} \tag{2.7}$$

When generalized for the transversely isotropic material in question.

$$\epsilon_x = \frac{1}{E}(\sigma_x - \nu\sigma_z) - \frac{\nu'}{E'}\sigma_y \tag{2.8}$$

$$\epsilon_z = \frac{1}{E}(\sigma_z - \nu\sigma_x) - \frac{\nu'}{E'}\sigma_y \tag{2.9}$$

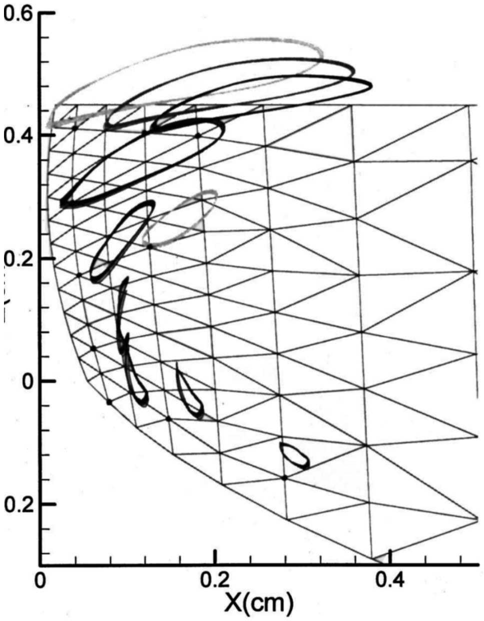


Figure 2.12: Trajectories of nodes in layer-8 of quasi-3D FEM model, reproduced with permission from [3]

$$\epsilon_y = -\frac{\nu'}{E'}(\sigma_x + \sigma_z) + \frac{1}{E'}\sigma_y \tag{2.10}$$

$$\gamma_{xy} = \frac{1}{\mu'}\tau_{xy}, \gamma_{yz} = \frac{1}{\mu'}\tau_{yz}, \gamma_{zx} = \frac{1}{\mu}\tau_{zx} \tag{2.11}$$

In this model we are solving the displacement vector ψ in the x(medial-lateral) and z(inferior-superior) planes. The y plane is the longitudinal plane in the dorsal-ventral direction. Using the assumption of planar strain, a displacement vector is defined as

$$\psi = u(x,y,z,t)\mathbf{i} + w(x,y,z,t)\mathbf{k} \tag{2.12}$$

where u and w are the lateral(x) and vertical(z) components of the displacement vector. Using our other assumption of linear elasticity, we define the relationship between the strain vector and the displacement vector as

$$\epsilon_{ij} = \frac{1}{2}(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) \quad \text{where} \quad (i,j = x,y,z) \tag{2.13}$$

where $u_x = u$, $u_y = v$, $u_z = w$. The previous set of equations enable a direct relationship between the stress and the displacements. These equations describe the continuum and the equations to be solved using numerical methods. The method of choice here is the Finite-Element Method (FEM); the FEM is commonly used in solving PDE's over complex domains by splitting the domain up into smaller finite elements. The equivalent problem is now solved over this element, where the field variable is interpolated inside the element from the values at the nodes of the element. Applying the FEM to the above problem we get:

$$M\ddot{\psi} + D\dot{\psi} + K\psi = F \tag{2.14}$$

where M is the mass matrix, $[D]$ is the damping matrix and $[K]$ is the stiffness matrix all of the size $6X6$. F is the forcing vector that is derived from the aerodynamic forces on the surface of the FEM vocal fold mesh. This paper used the simple Bernoulli formulation similar to the model explained in section 2.2.2.

While the individual spatial problem is in 2D, multiple layers of 2D layers are stacked together to achieve a 3D solution. There is a *string force*, that accounts for the forces between the different layers. This enables a structural solution that captures maximum complexity without becoming too computationally unwieldy. Cook et al [33], put forward a 2D/3D hybrid structural model of the vocal folds that preserves 3D effects of vocal fold length and longitudinal stresses, while maintaining the 2D computational domain. Many of the assumptions mentioned above are used by other models in literature; however, recent work has suggested that the linear elasticity assumption might be sacrificing accuracy at the altar of computational cost [31].

This formulation led to the development of multiple models with variations to either the structural model and its dimensionality, the fluid solver used (ranging from Bernoulli to 2D Navier Stokes[10]) and the application or model-focus in question (phonation[3], pathologies [52], flow separation computation[8] and bulging factor[7]).

While the finite-element model by Alipour et al [3] provided the first FEM structural solution for vocal fold phonation modelling, it used an extremely coarse and simple Bernoulli-based flow model. One of the first models with coupled fluid-solid interaction was by Tao et al.; this included FEM models of both the vocal folds and the airflow along with a proper collision model [113]. The first fully 3D model including the entire larynx geometry, the false vocal folds and laryngeal ventricles was put forward by Rosa et al [37]. The method used shared mesh nodes at the fluid-solid (ie. airflow-muscle) interface to have a coupled simulation. Despite the use of coarse meshes (2600 for airflow and 3000 for tissue) this method still had significant computational complexity and demonstrated the self-oscillation of vocal folds.

Later methods focused on completely resolved flow computation using an immersed-boundary method for the fluid-structure interface [79][41]. Luo et al [79] used two finite difference-based discretizations of the Navier-Stokes equation (for aerodynamics) and viscoelastic equations (for the solid mechanics), which are coupled for solving the FSI problem. Another interesting field of research is capturing fluid-structure-acoustic interactions. Link et al [75] created a 2D FEM scheme where Lighthill's analogy is used to describe the fluid-acoustic interaction as well as capture the Coanda effect. Finally, while most of the previous models used 2D Navier-Stokes models and the small deformation assumption, Zheng et al[135][128], extended the work of Rosa [37] to a 3D unsteady Navier-Stokes flow solver coupled with a 3D structural FEM solver. The simulation results were comprehensive, with insights into the glottal waveform, pressure and velocity distributions inside the vocal folds along with jet dynamics. However, a comparable resolved 3D simulation of flow is more than 100 times more expensive that a 2D flow model.

Recent work has also looked at the role that the false vocal folds play in the phonation process [129]. Zheng et al, created a version of their previous model with 2D flow, to study the computational effects of the false vocal folds in the phonation process [136]. Alipour et al, also extended their original FEM model to include the false vocal folds [9]. The time-dependent pressure and velocity distributions inside the glottal region were reported as part of the results, which an extremely valuable tool to benchmark simulations against.

2.4 Model Coupling

In the previous section, we have looked at a wide range of vocal fold models in literature. Each of the models had a structural and a fluid component. While the number of models tailored to individual applications is vast, they can be segmented by classifying these individual structural and fluid components. Subsection 2.4.1 gives an insight into what combinations exist in literature, and the lacunae in the field that needs to be filled for our application of articulatory speech synthesis. Subsection 2.4.2, talks about standard collision models that are used across structural simulations. Finally, Subsection 2.4.3 looks at which models are actually coupled to the vocal tract/trachea and the challenges with running a full simulation.

2.4.1 Combination of Structural and Flow Models

From the previous two sections, vocal fold models for articulatory speech synthesis can be divided into few major categories.

Structural Models (in increasing order of complexity)

- 2-Mass Models
- Body-Cover Models
- High-dimensional Lumped Models
- 2D/Quasi-3D Continuum Models
- 3D Continuum Models

Fluid Models (in increasing order of complexity)

- Bernoulli-Model
- 1D Inviscid Euler

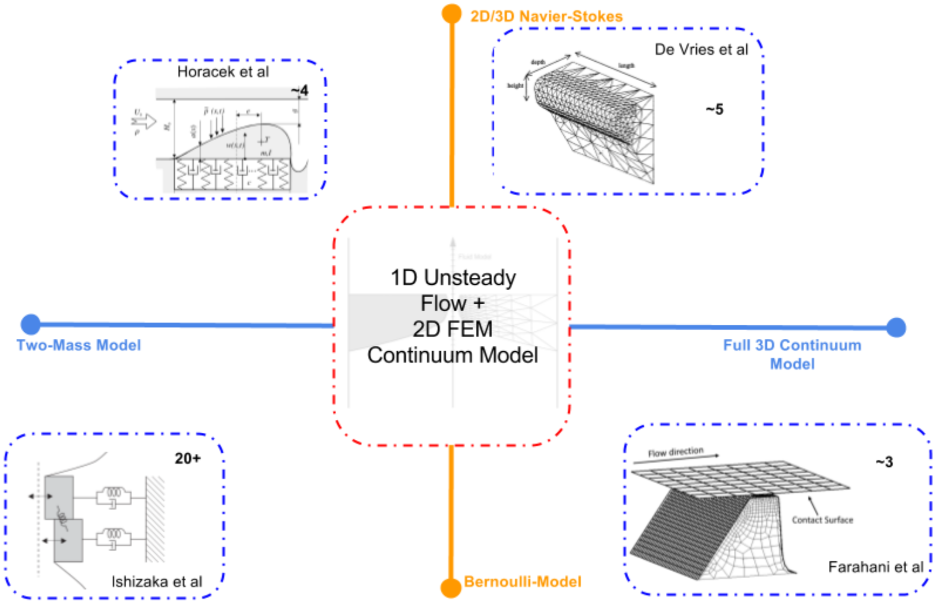


Figure 2.13: Visual depiction of the different combination of structural and flow models in the literature. Horizontal axis represents increasing spatial complexity from left to right. Vertical axis represents increasing flow complexity from bottom to top

- 2D Navier-Stokes
- 3D Navier-Stokes

Figure 2.13, gives a visual illustration of the different model combinations in literature. Going from left to right on the X-Axis we go from the 2-Mass to 3D continuum models in increasing spatial complexity. Going from bottom to top along the Y-Axis, we go from Bernoulli-based flow models to a 3D unsteady Navier-Stokes simulation. Images of sample models are given in each quadrant with their references [62][38][64][44] with the quantity of each type of model gives next to it. Our proposed model is filling a lacunae seen at the centre of the graph. (Note: The graph is not meant to be exhaustive, but rather provide a visual depiction of the standard combinations that exist in literature to better understand potential gaps in the field)

Figure 2.13 leads to some striking observations:

- There exists a multitude of lumped-element models in literature, mainly coupled with a Bernoulli-based solver. These models have generally been the most common models used for speech synthesizers, for their conceptual and computational simplicity. While there exist some combinations of lumped-element models with an unsteady flow model, these have been to benchmark the performance of quasi-steady Bernoulli-based models with the unsteady flow models.

- With increased computational capabilities, the combination of 2D/Quasi-3D continuum models driven by 2D Navier-Stokes models has been applied to different sub-problems. The main advantage of this approach is a synergy of dimensionality between the flow and structure models and the computation of an entire velocity field. Most of the 2D flow models are solved either using the Finite-Volume Method or the Finite-Element Method, while 3D flow models use the latter.

- There is a lacunae of 1D unsteady flow models used in conjunction with continuum vocal fold models. This could particularly be useful in cases where the bulk pressure distribution is of more importance than a fully-resolved flow computation. In addition these models would be computationally much cheaper than comparable 2D models. However, we need 1D flow models that can estimate the flow separation point and the glottal flow properly, unlike previous Bernoulli models.

2.4.2 Collision Handling

Collisions between the vocal folds presents a different prospect to both lumped-element as well as continuum models. Since the lumped-elements could not come to a sudden unnatural stop on contact (zero-glottis condition), it was proposed to apply a non-linear spring at the time of impact to prevent further movement in conjunction with increasing the damping ratios of the masses [64]. This remains an open problem to be solved in continuum models with many different ways of handling collisions suggested. We will look at some significant types of contact handling in literature.

Many of the continuum models assumed mid-line symmetry between the right and left vocal folds, to reduce computational costs by simulating just a single vocal fold. This moves the plane of contact to the glottal midline for many models. Alipour et al [3] in their first model removed a degree of freedom when any node reached the plane of contact. In terms of full 3D structural models, Rosa et al [37] calculated a force to avoid interpenetration of nodes that are in contact while Tao et al [113] use the Augmented Lagrangian Method. As mentioned previously, structural models which use a sharp-interface immersed boundary method are directly coupled to the flow solvers.

An important drawback of these class of problems, especially in the case of 2D or 3D fluid models, is the possibility of zero-area and zero-volume elements. This is a major stability issue with regards to the overall system with the responsibility falling on the structural solver to ensure a minimum glottal diameter at all points. Thus, there is never complete glottal closure in many of these models; for example in quasi-3D models, not all the layers can be completely closed at the same time [9]. This can be a drawback in using these models when stability is important, in applications such as a coupling articulatory synthesizer.

2.4.3 Vocal-Tract Coupling

Through the previous sections we have gained a bird's eye view into the complexities and intricacies that goes into building a vocal fold model. As mentioned previously, a lot of early models were based off the *source-filter theory* [42]; this meant that early model were used as simple exciters to the resonant vocal tract. However, as the non-linear coupling of the vocal folds and the vocal tract has been established in literature [119], it is important to design vocal fold models that are coupled to the vocal tract load.

The glottal flow (U_g) is the output of the vocal fold model that is fed into the vocal tract. This is the source to the vocal tract simulation; the pressure recorded at the outlet of the vocal tract would be the equivalent sound pressure that a listener would hear. There is also feedback from the vocal tract to the vocal fold model: the supraglottal or epilaryngeal pressure (P_e) is acts as a feedback to the fluid simulation of the vocal folds. Similarly the pressure output at the end of the trachea, the subglottal pressure (P_{sg}), will act as the input pressure to the vocal fold model. The supraglottal pressure along with the subglottal pressure act as the boundary conditions to the fluid simulation in the vocal folds. Also, the structural characteristics of both the attached subglottal and supraglottal tracts have shown to affect the vocal fold vibration [115].

However, if we look at the vocal fold models in literature we can see that the vast number of them are not actually coupled. In fact in literature, less than 5 continuum models have actually been coupled to a vocal tract for simulation. Most models give a static boundary condition on each end based on experimental values without running a vocal tract simulation as part of a larger system. It is also pertinent to note that it is easier to couple lumped-element models with quasi-steady Bernoulli-based solvers than un-

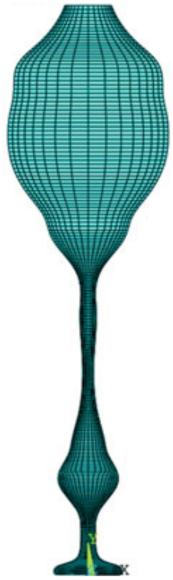


Figure 2.14: View of the idealized 2D vocal tract shape for Czech vowel *[a:]* coupled with the 2D vocal tract model, ©Springer, reproduced with permission from Hajek et al [53]

steady solvers. In fact, both the exemplar two-mass and body-cover models have been coupled as part of larger synthesizers [64][101][126][107]. There are two main potential reasons we can put forward for the lack of coupled models in literature: firstly, running a full synthesizer is extremely conceptually difficult to design and implement, and computationally expensive to run. Till recently, the computing capabilities available made it difficult to run anything others than a 1D wave-reflection based vocal tract model. Secondly, the feedback from the vocal tract can change rapidly at certain times, despite the scale differences between the vocal folds and vocal tract simulations. This changes the boundary conditions of the flow model of the vocal folds, often pushing the vocal folds into regions of instability.

Recently, efforts have been made to couple more complex vocal folds models to vocal tract models. Alipour et al have demonstrated a quasi-3D vocal fold model, driven by an unsteady 2D Navier-Stokes flow model coupled to a 1D wave-reflection analog vocal tract and trachea [7][9]. Recently, efforts to combine more complex vocal tract models have been undertaken; the 2D Navier-Stokes equation was solved in a complete computational channel inspired by the vocal tract [94]. This included the vocal folds and a small trachea tract as well providing an insight into the flow-fields inside the channel. Recent Hajek et al [53], created a 2D finite element solution of the self-oscillating vocal folds, connected to a trachea and vocal tract model (figure 2.14. Idealized vocal tract shapes were created for standard Czech vowels from MRI data and the flow was solved using the Navier-Stokes equations, using the Arbitary Lagrangian-Eulerian approach for boundaries. The first two formants of the generated outlet pressure was compared with published data and was found to be in relatively good agreement. The most complex model of the fluid-structure-acoustics interaction in the vocal tract was a 3D model using realistic laryngeal and vocal tract geometries by Jiang et al [68]. This model achieved self-sustained oscillations and demonstrated that a range of voice-related quantities were within normal physiological ranges. The model also showed the likelihood of strong source-filter coupling from its results; the main drawback of the model was the immense computational cost to simulate the system.

2.5 Articulatory Synthesizers

One of the major goals of vocal fold modelling is articulatory speech synthesis. Similar to vocal fold models, we can think of the articulatory synthesizer having two main components apart from the vocal fold model: the structural model that represents the physiology of the upper airway and the biomechanics of the different articulators; and the acoustic model solving the flow through the vocal tract channel. Figure 5.1 illustrates the different components that are part of an articulatory synthesizer. Since this is a massive field in it's own right, we shall only look specifically at important attempts in literature to model different articulatory structures and acoustic phenomena in the vocal tract.

2.5.1 Structural Models

As seen in 5.1, the structural model of the vocal tract includes two main components: biomechanical models with articulator control, and extraction of the vocal fold mesh for the acoustic simulation. There is a bi-directional coupling between the upper airway structures and the aeroacoustic flow; however, very few models capture this phenomenon. As mentioned before very few articulators have all the components mentioned in 5.1, and often the control of articulators to shape the structural vocal tract is not included.

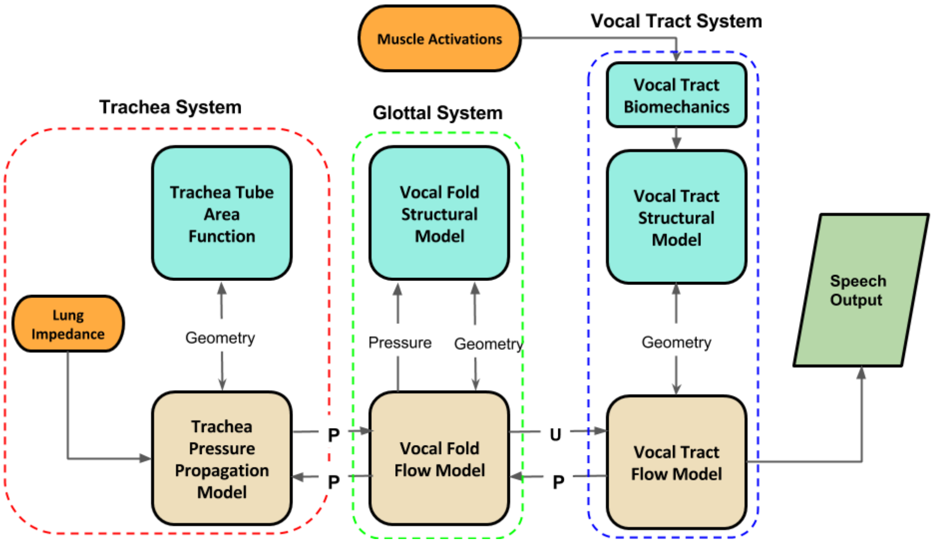


Figure 2.15: Components of an Articulatory Synthesizer

Teixeira et al implemented a 2D midsaggital anatomical model of the vocal tract (figure 2.16), where positions of individual articulators can be controlled using data files [114]. This was an improved implementation of two seminal models of the field [84][95]. It is important to note that this is not strictly a biomechanical model in the true sense of the word, since there are no physics that are associated with the simulation. Rather, the movement of articulators can be derived from existing data, most likely medical imaging data of the vocal tract.

Dang et al introduced FEM-based models of articulators to replicate 2D midsaggital regions of the vocal tract and simulate articulatory movements [36]. Fully 3D models of articulators of the vocal tract were created from medical imaging data by Badin et al [16], with biomechanically interpretable data used to control the system. In the same vein, a comprehensive speech synthesizer with a 3D structural model was put forward by Birkholz et al as part of the *VocalTractLab* software system. Here a gestural score was computed for every utterance that the user wants to achieve; this gestural score was then used to compute the vocal tract shape to run the acoustic simulation (figure 2.17).

One of the major advances in the field has been the development of the

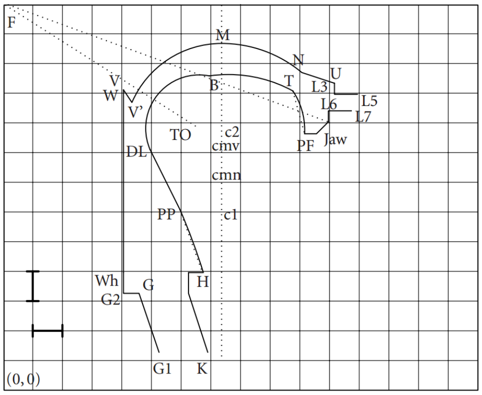


Figure 2.16: 2D midsaggital anatomical model of the vocal tract, reproduced with permission from Texeira et al [114]

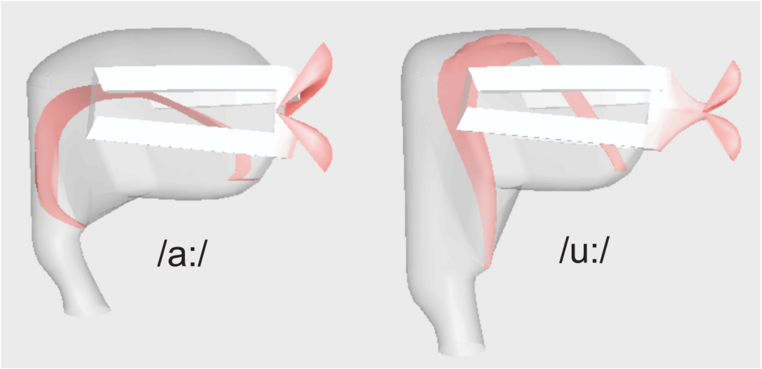


Figure 2.17: Biomechanical vocal tract model used to calculate the vocal tract shape, *VocalTractLab*, reproduced with permission from Birkholz et al [24]

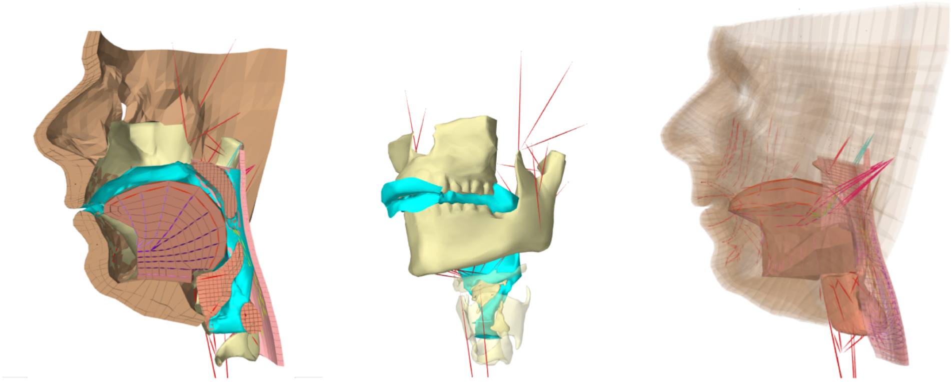
Name	**Value(s)**
Face	FEM
Tongue	FEM
Jaw, Hyoid, Maxilla	Rigid
Soft-Palate	FEM
Pharynx	FEM
Larynx	FEM
Larynx Cartilages	Rigid

Table 2.1: Summary of components in FRANK and component types. Rigid = rigid body, FE = finite element. Adapted from [112]

comprehensive biomechanical toolkit *ArtiSynth* by Lloyd et al [77], geared towards simulations of the upper airway of the human body. Based on an open-source model, the system provides tools to build biomechanical models and simulate them using an internal physics engine. This engine is capable of simulating combine multibody and finite element simulations with collision handling, connectors and numerical solvers build-in. A particular distinguishing feature of the system, is the ability to model both line and FEM muscles, and control them using muscle activations for forward and inverse simulations. This is critical for building a complete speech synthesizer, as it provides a strong physiological link for control of articulators rather than gestural scores that are a function of medical imaging data. Many different models of upper-airway articulators have been put forward in literature [27][50][54][78][92][100][127] but often included only portions of the upper-airway complex. To enable a complete study of the upper-airways complex across fidelities, a Functional Reference ANatomical Knowledge (FRANK) [13] biomechanical model of the head and neck has been implemented in the ArtiSynth toolkit. Table 2.1 summarizes the different components involved in the FRANK model and the sources from which they are obtained. Figure 2.18 illustrates the hard and soft components of the model, and the airway that would be used for the acoustic simulation.

2.5.2 Flow/Acoustic Models

In this subsection, we shall look at some exemplar flow/acoustic models that have been suggested in literature for articulatory synthesis and other upper-



(a) mid-sagittal cross-section (b) bones and cartilages (c) soft-tissues

Figure 2.18: FRANK: a Functional Reference ANatomical Knowledge [13] a) midsaggital view of the components b) hard components of the model c) soft components of the model

airway related functions. In general there are a few major ways to simulate sound/flow propagation in a tube given below.

- **Digital wave-guide filters** These models are based on the classic Kelly-Lochbaum reflection-type [69] line model that models a non-uniform tube as a set of concatenated tube segments with varying values of impedance. This model is computationally very efficient and has been used in a variety of classic vocal tract models [85][74][107]. However, the model is not particularly versatile and cannot include specific turbulence and flow separation effects of the flow.

- **Transmission Line Circuit (TLM) models** This model implements the electrical circuit equivalent of the acoustic tube in question. Thus the model is converted to the form of resistors, capacitors and inductors with values that capture the impedances of the individual tube sections [64][81]. Birkholz et al illustrated how this model could be used as part of a complete articulatory synthesizer [24].

- **Hybrid time-frequency simulation** These models attempt to take advantage of both classic approaches for a fast versatile simulation [11][101]. The impulse responses calculated for the vocal tract are combined with the glottal input signal to find the radiated output pressure.

- **Direct numerical simulation of flow** This method involves solving the characteristic acoustics equations (Eg. Webster equation in 1D [126], Wave equation in 2D [132], Navier-Stokes in 3D[68]) over the vocal tract domain. This method is the most physiologically relevant but can be computationally expensive especially at higher dimensions.

Of these approaches, the direct numerical simulation of flow is of particular interest to us in terms of the overall goals of articulatory synthesis. In addition, these models can also guide us in choosing better flow models for the vocal fold simulations as well. Apart from speech synthesis, there were models put forward in literature to study fluid-structure interaction in other related biomedical and phonation problems of the upper airway. In general, a stable solution of the upper airway with colliding deformable bodies and closure of the airway is extremely hard to achieve apart from being computationally prohibitive. One of the major challenges models need to overcome is being able to predict the recovery of pressure after a constriction in the airway, and recover stably when the airway opens up again. A lot of these challenges are similar to those faced by the flow models of the vocal folds.

Many different models have been suggested in literature ranging from lumped parameter models [21], 1D [28][65], 2D [66][76][80] and 3D models [55][56][57][82] of airflow. Of these models, two models stand out for their approaches to the modelling of the flow in the upper airway. Firstly, one of the flow models that demonstrated the ability to predict the pressure recovery, and stably handle closure and even collapse of the upper airway, was the 1D model suggested by Anderson et al [12]. Handling collapsing or closure in tubes is an non-trivial tasks for most flow models, and can often lead to an unstable simulation. The model was experimentally validated, and the bulk pressure predictions were shown to agree with other comparable 3D models. The model also required significantly lower computational costs and was used for modelling Obstructive Sleep Apnea (OSA) simulations. Secondly in terms of our goals of achieving faster, higher quality speech synthesis, Zappi et al [132] provided the first real-time solution of the 2D wave equation for interactive speech synthesis applications. This model enables us to move beyond simplistic 1D representations of the vocal tract, to achieve higher quality synthesis by capturing the vocal tract geometry better without the associated increase in computation time.

2.6 Data Acquisition and Measurement

In this section we focus on the methods used to acquire data and measure phenomena associated with phonation. We focus on the vocal fold model, to highlight the experimental data and measurements that are available as a benchmark to compare the performance of computational models with. Readers are refered to the review of the field by Mittal et al [87], for a more comprehensive tabulation of the different studies on phonation. We also look at the medical imaging techniques most commonly used to capture the vocal tract structure in subsection 2.6.4.

Subsection 2.6.1, talks about *in-vivo* studies of the vocal folds to collect data on phonation, and for standard measures of the vocal folds. However, the inaccessibility of the vocal folds has meant that *in-vivo* studies of the vocal folds have had strong ethical and logistical issues, limiting their scope. This has meant that speech researchers, have attempted to find alternative methods to study and measure vocal fold characteristics. Subsection 2.6.2, explores the use of excised laryngeal from both humans and animals (eg. canine) in lieu of the actual vocal folds from a live subject. Subsection 2.6.3, looks into synthetic fabricated models of the vocal folds that were employed to measure information. Finally, subsection 2.6.4 talks about the methods used to obtain the structure of the vocal tract.

2.6.1 In-Vivo Studies

There are three main methods used are given below:

- **Laryngoscopy:** The laryngoscope is an endoscope that is particularly designed for observing the laryngeal complex from the supraglottal duct. A catheter is inserted into the throat, past the velum to just above the larynx. A camera records the oscillation of the vocal folds from the top [34]. Figure 2.19 contains sample images from a laryngoscope in top for a cycle of phonation. The opening and closing of the glottis is clearly visible in the different stages of the cycle; however, this data needs to be buttressed with other data to gain a quantitative understanding. Since there is only a top down view, it's tough to make out the exact point of opening/closure, and thus this data would ideally be combined with electroglottography (EGG) data.

- **Electroglottography (EGG):** The EGG is a non-invasive tool that is enables speech researchers to get a high-temporal resolution reading of the degree of closure in the glottis. Electrodes are attached on

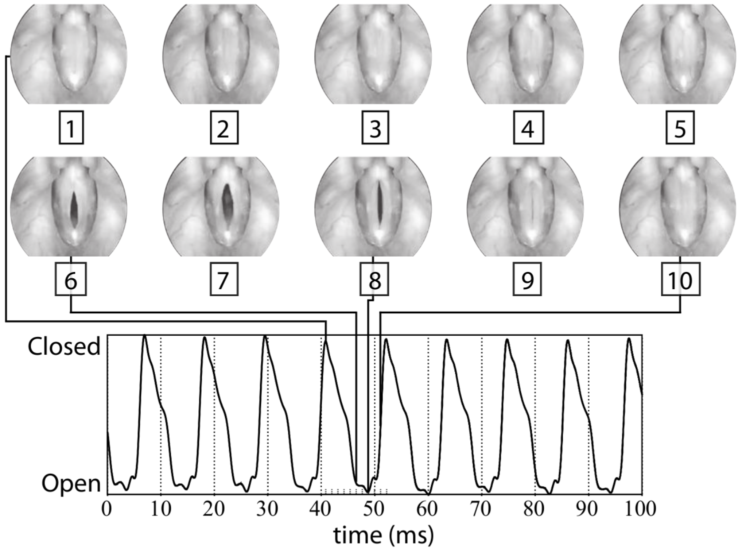


Figure 2.19: High-speed laryngoscopy images (above) synced with EGG data (below) for one cycle of vocal fold vibration, ©Wiley Publishing, Gick et al [51]

either side of the thyroid notch; the electrical resistance between the electrodes is a function of the degree of opening and closure of the glottis [47][73]. A zero value or minimum value of the electrode readout would correspond to an open glottis and vice-versa. Figure 2.19 shows the EGG signal in the bottom related to a specific laryngoscopy signal; this is often called a laryngograph.

- **Pneumotachography and Audio Recordings:** The pneumotachograph (or airflow meter) is a tool used to measure airflow and air pressure during speech. It constitutes the famous *Rothenberg* mask [96], that covers the mouth and nose, calculating the oral and nasal airflow. When used in conjuction with a microphone recording the output sound pressure, we can gain an understanding of the glottal waveform [89][61].

2.6.2 Excised Laryngeal Studies

The focus of excised laryngeal studies were usually twofold: firstly, to find out the structural properties of the various vocal fold layers, and secondly, carry out experimental measurements using these excised larynxes placed in physiologically viable conditions. This topic is vast with many different models in literature; readers are referred to the paper by Miri, [86] that provides an in-depth review of the various mechanical testing methods and the constitute materials that are consequently used for modelling. The paper explains the different methods that can be used for measurement (traction testings, shear rheometry, linear skin rheometry and indentation testing) and the different values reported by technique. Both human and canine larynxes have been used to calculate the material properties [5][2][71][70]. This wealth of data provides excellent starting points to build and compare our models; however, these studies have some drawbacks. In particular, the researchers have to work quickly since there are limited run times, difficulty in restoring the proper tensioning and specific environmental conditions which need to be maintained.

2.6.3 Synthetic Laryngeal Studies

Fabrication of synthetic larynxes for to study vibratory characteristics of vocal folds has been a widely used tool, especially over past few years. The following sections looks at the different models in literature, and the data they provide in validating our potential system.

Static Models

Early synthetic models of the vocal folds were static models, that were generally extruded 2D models progressing to more complex 3D models. While these models did not oscillate, they gave insight into the relationship between the pressure distribution inside the glottis and the glottal flow parameter. Scherer et al [97], looked at the difference in intra-glottal pressure distributions for a symmetric and tilted static laryngeal model. Fulcher et al [49] did a similar study using a symmetric system with two vocal folds, establishing the link between transglottal pressure, the geometry and the output flow rate. This was later performed with a hemilarynx model as well [48].

Driven Models

In some models an external control was given over the structure of the vocal folds; this was achieved using linear actuator to make the vocal folds periodically change their shapes. This enables us to see the synthetic vocal fold take up shapes that are seen during actual phonation with the caveat that the motion is decouple from the airflow. There were both simple linear oscillatory models [40] as well as more complex waveform-based models in literature [124].

Self-Oscillating Models

Self-oscillating models achieve coupling between the structure as well as the airflow, enabling osciallations to naturally entrain. An important result for the field that was achieved early-on when Titze et al [120], had been able to establish the threshold pressure for phonation assumping a model that was made up of a stationary body layer, and fluid-filled oscillating cover layer. Similar models diving the idealized structure into two layers was also put forward by Chan et al [30][29]. Thomson et al [115] put forward a variation of the canonical M5 vocal fold model introduced by Scherer et al [97] that was mentioned previously. Later modifications of this model were also put forward; an important model was fabricated by Becker et al [18], who studied the fluid-structure interaction inside the glottis.

2.6.4 Vocal Tract Measurements

The measurement of the various articulators of the upper-airway is a complete research field by itself. In general, for speech we shall assume that airway is directly derived from MRI data or obtained from the biomechanical model as in the case of FRANK (section 2.5.1). As mentioned in the vocal-tract coupling section 2.4.3, Story et al [108] developed a set of area functions from Magnetic Resonance Imaging (MRI) data. Story [106] later revisited this data-set by carrying out new measurements of area functions from the same patient, to better understand inter-speaker variability. Another famous MRI data-set is the *'ATR MRI database for Japanese Vowel Production'* which contains male MRI vocal tracts data [109], obtained with the phonation-synchronization method. Takemoto et al [110], provided a method for extraction of area functions, commonly used in 1D vocal tract models, from MRI data. These data-sets allow us to build vocal tract domains over which we can solve the wave equations.

2.7 Discussion and Conclusion

There has been a cornucopia of vocal fold models proposed over the last half-century covering a range of applications from articulatory speech synthesis, vocal fold pathologies and phonation dynamics. Early models made significant simplifying assumptions to both the structural and flow components of the model; the vocal fold structure was approximated as lumped rectangular masses, and a Bernoulli-based flow model was applied. As computational capabilities grew, the first continuum models of the vocal folds emerged with the constitutive equations being solved using numerical methods such as FEM. This has led to the development of very complex and comprehensive models that have given speech researchers a better insight into phonation dynamics, jet dynamics and coanda effect as well as pathology treatment. This has been mirrored by the increase sophistication of data acquisition techniques used by scientists on the vocal folds. Finally, tremendous work has been put into building vocal tract models for articulatory synthesis. All the factors combined, make the prospects for the field seem extremely ripe. To take the research forward into practical applications, there is a need for usable models of the vocal fold and vocal tract to be coupled; this include considerations of speed, robustness and usability. Equally, models are currently disparate, implemented across different proprietary and commercial tool-kits, thus pointing towards the need for integrated open-source models.

This chapter has presented an overview of vocal fold modelling in the context of articulatory speech synthesis. Exemplar models of the field have been highlighted, and an attempt has been made to see if there exists certain lacunaes in the field, despite the wealth of models in literature. We have identified areas of research that require further investigation. In particular, there is a need to move beyond lumped-element models to include continuum models of the vocal folds in articulatory synthesizers. However, there are two major issues with regards to continuum models: there is no appropriate flow model that strikes a balance between the computational and conceptual simplicity of Bernoulli, and the significantly greater computing cost of using 2D Navier-Stokes models. Secondly, the vast number of vocal fold models exist in a vacuum with static boundary conditions and no coupling to either the vocal tract or trachea. Thus, a continuum vocal fold that combined a 2D structural model with an unsteady 1D flow model would move the field towards a practical, use able vocal fold model. Running this model as part of a stable coupling articulatory synthesizer, would enable speech researchers to better study the acoustic outcomes of model decisions. In the following chapters we describe our contributions to these open research problems.

Chapter 3

2D Continuum Model with 1D Flow Model

In this chapter, we introduce our 2D continuum model of the vocal folds, driven by an 1D unsteady flow model. We first start with defining the characteristics that we look for when choosing structural and flow models of the vocal folds in Section 3.1. In the following subsections of the chapter we describe the model formulation, including the numerical implementation that we use in solving the system. Finally, the coupling of the structural and flow models and the treatment of contact is covered.

3.1 Introduction

The modelling of the vocal-fold phonation is an extremely complex fluid-structure interaction problem. The vocal fold structure is made of multiple layers, the surface of which is acted on by aero-acoustic forces that are predicted by the fluid model. In previous work, many assumptions have been made for the modelling of the airflow through the vocal folds. One of the classic assumptions, the quasi-steady assumption made for applying the Bernoulli-equation, is demonstrably false as seen in previous work. However, compromises need to be made in model formulation based on current computational capabilities. Some of the major considerations when choosing an appropriate fluid model for the vocal folds are listed below:

1. Instead of a fully-coupled approach, it would be suitable to have the structural and flow equation solved separately from each other. This would enable us to switch out the structural model for a more complex 3D model later, without affecting the entire system, thus achiving modularity. In this case, the area function of the glottal tube $A(x, t)$ in question, can be treated as a known quantity.

2. Comparison studies between Bernoulli-based solvers and 2D Navier-Stokes solvers for lumped-element and continuum-models, showed that,

while Bernoulli models performs acceptably in predicting bulk intraglottal pressures, they are unsatisfactory in computing the location of separation point and the glottal flow rate. The separation-point has been shown to be critical to stable self-sustained oscillations as it plays an important role in the force loading of the vocal folds. Thus, the Bernoulli-model is not sufficient for our formulation despite its many computational advantages.

3. While 2D/quasi-3D and 3D models of flow have shown themselves capable of reproducing many of the intricacies of phonation fluid dynamics, the computational cost involved in solving both the solid and fluid equations in 2D (or higher) dimensions continues to be prohibitive. This rings especially true in the case of articulatory synthesis, where the role of the *Coanda effect* for example, is not important to achieving modal phonation which is our primary goal [75]. Thus, we choose to focus on those features that are critical to generating better acoustic outcomes.

4. Viscous losses along with turbulences play an important role in the phonation dynamics. The model that we choose should be able to handle this automatically.

5. Many 2D models and 3D flow models, do not allow the vocal folds to collide in the structural domain as they need to have a minimum area function diameter. This is done to ensure that there are no zero/negative-volume inverted elements in the flow mesh that will lead to instability. Thus, our flow model should handle closure and reopening of the glottal tube in a numerically stable and physically realistic manner.

6. The vocal fold channel has sudden variations in area and an irregular geometry. Our choice of spatial discretization should account for this.

The natural conclusion that can be drawn from the review of literature in Chapter 2 is that our ideal flow model will be an 1D unsteady fluid model. This model should be able to predict the bulk intraglottal pressure, flow separation point, viscous losses and most importantly, glottal flow values. The model should be able to handle tube closure and reopening, apart from irregular geometries. Finally, a loosely coupled framework would be helpful in extending our structure model in the future without too much added effort. As mentioned before in section 2.5.2, the unsteady 1D fluid model suggested by Anderson et al for modelling collapsing tubes for Obstructive Sleep Apnea (OSA) [12] contains a lot of the characteristics required in our system and showed comparable performance to 3D simulations. This model will be used as a starting point around which our fluid model will be designed.

We first start with the formulation of the structural component of the vocal fold, followed by the fluid model and its numerical implementation. Information about the fluid-structure coupling is then provided in the following subsection. Section 3.4 first contains an analytical case to verify our numerical solution framework. This is followed by experiments using the coupled model to examine the fluid model's abilities to predict pressure and velocity distributions that are in-line with experimentally-measured and simulation results in literature.

3.2 Structural Model

We use a 2D structural model based on the model by Alipour et al [3] with a focus on lower computational load. Of the models that are available in literature this model made sense for a few major reasons: it is the simplest 2D continuum model available and the logical step-up from using lumped-element models. While there are more complex structural models both in terms of dimensionality as well as material properties, this model has been shown to reproduce the major phonation characteristics in previous studies when coupled with a 2D Navier-Stokes model [7]. Potentially in future iterations, the possibility of using a model-reduction technique to optimize the system further can be looked at. We do not reproduce the mathematics of deriving the matrix equations here for the sake of brevity. Readers are referred to the original paper for the same [3].

The structural mesh that we use for the finite-element method is shown in figure 3.1. There are three main regions of the mesh, each with different material properties as suggested by the *body-cover* model [59]. We choose to assume symmetry across the midline plane and simulate a hemi-laryngeal model instead of both vocal folds. The model uses a linear-elasticity assumption which is computationally cheaper to solve, and has been validated in previous studies. A linear shape function is used for the finite-element formulation and the material properties are taken from a recently updated version of the model [9] (Table 3.1). The vocal fold mesh was the same mesh used by [3] and the vocal fold structure was divided into three ma-

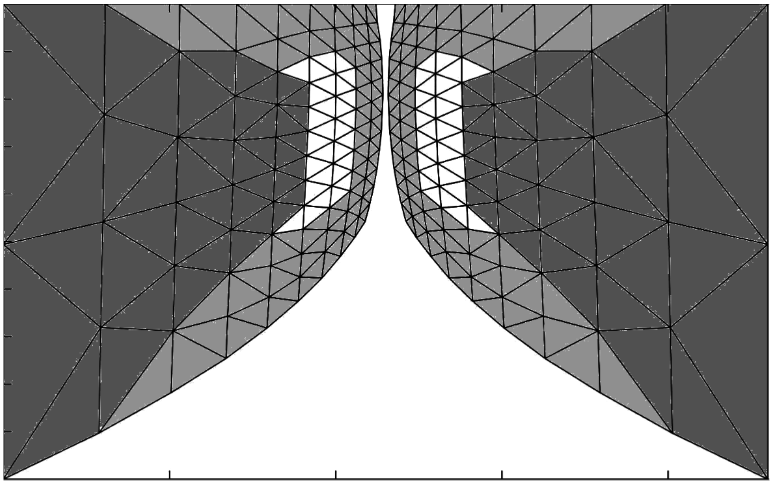


Figure 3.1: Finite element mesh shown in the coronal plane. This includes the body (dark grey), the ligament (white) and the cover (light grey) regions

terial regions: body, cover and ligament. The FEM solution of the spatial problem yields a second-order matrix differential equation (3.1) in the time domain:

$$M\ddot{\Psi} + D\dot{\Psi} + K\Psi = F \tag{3.1}$$

The equation is discretized using the second-order central scheme centred at the n$^{\text{th}}$ time-step for stability as shown in equations 3.2 3.3:

$$\dot{\psi} = \frac{\psi_{n+1} - \psi_{n-1}}{2.\Delta t} + O(\Delta t)^2 \tag{3.2}$$

$$\ddot{\psi} = \frac{\psi_{n+1} - 2\psi_n - \psi_{n-1}}{(\Delta t)^2} + O(\Delta t)^2 \tag{3.3}$$

The aerodynamic force and string forces are used to calculate forcing vector F, in equation (3.1). Since contact between the symmetric vocal folds would happen at the glottal midline, a rigid plane is assumed to be present there. When the vocal fold reaches the midline, an impact force is applied to prevent interpenetration. The contact force is normalized over a collision region defined by the A_{closed} value associated with the fluid model, that explained in subsection 3.3.3.

Name	**Value(s)**
Body Longitudinal Young's modulus	20 kPa
Cover Longitudinal Young's modulus	15 kPa
Ligament Longitudinal Young's modulus	30 kPa
Body Transverse Young's modulus	2 kPa
Cover Transverse Young's modulus	1.5 kPa
Ligament Transverse Young's modulus	3 kPa
Body Longitudinal Shear modulus	12 kPa
Cover Longitudinal Shear modulus	11 kPa
Ligament Longitudinal Shear modulus	20 kPa
Body Viscosity	6 poise
Cover Viscosity	3 poise
Ligament Viscosity	5 poise
Longitudinal Poisson's ratio (All layers)	0.4
Transverse Poisson's ratio (All layers)	0.9
Lung pressure	1.0 kPa
Fluid density	1.14 kg/m^3
Fluid dynamic viscosity	1.86e-5 Pa$\cdot$ s
χ_{min}	0.2

Table 3.1: A list of the different material properties used for the vocal fold model. Parameters derived from [3][9]

3.2.1 Numerical Implementation Procedure

The solution of the tissue mechanics is outline as follows:

1. The simulation is started with the hemilaryngeal mesh at the prephonatory position. The tissue properties of each layer are defined based on Table 3.1.

2. The mass, stiffness and damping matrices $(6X6)$ are calculated for each element. The mass matrix is a function of the density ρ and area A. The density and stiffness matrices are calculated using the finite-element interpolation shape function inside each element to integrate the strain energy and the viscoelastic properties of the different layers of the vocal folds.

3. These element masses are assembled into the global matrices that are shown in Equation 3.1.

4. The area function $A(x,t)$ is extracted from the structural model and given to the fluid model. The pressure distribution is obtained through solving the fluid model.

5. The nodal forces for each surface element in calculated and applied, along with the string forces across all elements. This gives the global force vector F.

6. The matrix differential equations are solved to give us the displacement vector ψ for the current time step.

7. The new nodal coordinates are calculated by adding the dynamic displacement vector to the previous coordinates.

8. The updated geometry is used to calculate the area function $A(x,t)$ for the next time step of the system. We repeat steps 2 to 7 for our time period of simulation.

3.3 1D Fluid Model

3.3.1 Model Formulation

Here we attempt to apply the 1D version of the Navier-Stokes mass and momentum equations for incompressible flow to our problem. Based on the ideas of Cancelli and Pedley [28], the flow continuity and the momentum equations for the equation are written as:

$$\frac{\partial}{\partial t}A + \frac{\partial}{\partial x}Au = 0 \tag{3.4}$$

$$\rho u\frac{\partial}{\partial x}u + \rho\frac{\partial}{\partial t}u + \frac{\partial p}{\partial x} - \tau\frac{s}{A} = 0 \tag{3.5}$$

$$\tau - \tau_{fric} - \tau_{\chi} = 0 \tag{3.6}$$

where s is the perimeter around the cross-section area A, u is average velocity, p is pressure and ρ is density. The term τ models the viscous losses with two major components: τ_{fric} describing the laminar losses and τ_{χ} describing the losses due to flow separation. The equations are given below:

$$\tau_{fric} = -2\mu(s/A)u \tag{3.7}$$

$$\tau_{\chi} = (A/s)(1-\chi)\rho u(\frac{\partial}{\partial x}u) \tag{3.8}$$

While τ is written separately for clarity, it can be seen that the three major solution variables are u, p and τ. The flow separation term is of particular importance in our case; we can define flow separation at the point beyond which the pressure is effectively constant i.e. $\partial p/\partial x = 0$. The flow separation point also plays another critical role: after the drop in pressure at the constriction, a pressure recovery takes place downstream in the channel. The flow separation point limits the pressure-recovery to match with the boundary conditions downstream. The χ term is defined as:

$$\chi = \begin{cases} 1, & \text{for } u\frac{\partial p}{\partial x} < 0 \\ \chi_{min}, & \text{for } u\frac{\partial p}{\partial x} \geq 0 \end{cases}$$

where bi-directional flow is accounted for and χ_{min} is defined by the user based on the problem under consideration. As suggested by Anderson et al [12], we use the inviscid approximation to calculate χ.

$$\frac{\partial p}{\partial x} \approx -\rho u(\frac{\partial u}{\partial x}) - \rho(\frac{\partial u}{\partial t}) \tag{3.9}$$

This approximation is advantageous as the τ term is no longer dependent on p, and because the τ-term is generally small when $\partial p/\partial x = 0$.

Name	Value(s)
Lung pressure	1.0 kPa
Fluid density	1.14 kg/m^3
Fluid dynamic viscosity	1.86e-5 Pa· s
χ_{min}	0.2

Table 3.2: A list of the fluid properties used for the vocal fold model. Parameters derived from [3][9][12]

3.3.2 Numerical Solution Framework

Anderson et al [12] suggested using Newton's method to solve equations (3.4), (3.5) and (3.6) since $A(x,t)$ is a known quantity. The method would be:

$$J(X) \cdot \Delta X = -F(X) \tag{3.10}$$

where X is the solution vector, J is the Jacobian matrix and F is the residual vector. The suggested solution framework is iterative, recalculating $X^{(n+1)}$ until the residual F is below a certain tolerance. However, the requirement to compute the Jacobian makes this coupled solution more complicated. While the system is still quite fast, the possibility of a faster solver of the equations can be considered.

In the case of a velocity-driven flow, we can see that the flow equations can be solved sequentially. The velocity boundary condition can be applied to (3.4), to calculate the velocity distribution $u(x)$. This can be used to compute $\tau(x)$ using (3.6). Finally, equation (3.5) is solved using the calculated u and τ values. This procedure is extremely fast, and computationally efficient. On the other hand, for pressure-driven flows we require a coupled-solution of equations (3.4)-(3.6), which can be solved using Newton's method as mentioned above. However, glottal flow is generally driven through pressure-pressure boundary conditions though there exists cases in which velocity-driven flow can be defined instead.

Novel Decoupled Solution Framework:

Thus, we suggest a method to convert pressure-pressure boundary conditions to the equivalent velocity-pressure boundary conditions, followed by a decoupled solve. The solution involves a bounded search, where we iterate the system till we find uInlet-pInlet boundary conditions equivalent to the specified pInlet-pOutlet boundary conditions. This solution would enable us to have an unsteady 1D numerical implementation that is significantly faster than the coupled solver. The model parameters are given in table 3.2 Our solution procedure is as follows:

1. Create two initial guesses for input velocity, $u1_{i=0}, u2_{i=0}$ for a given $pInlet^{n+1} - pOutlet^{n+1}$ boundary condition. The superscript refers to the simulation time, and the subscript refers to the iteration number. The previous time step's input velocity $uInlet^n$ and $1.1 * uInlet^n$ are good guesses to speed up convergence.

2. Use the decoupled solver to find the outlet pressures $p1_{i=0}, p2_{i=0}$ for the $u1_{i=0} - pInlet$ and $u2_{i=0} - pInlet$ systems respectively.

3. Calculate the change in pressure with velocity $dpdu_i = (p2-p1)/(u2-u1)$

4. Update $u1_i = u2_{i-1}$ and $p1_i = p2_{i-1}$ from the previous iteration, and create a new guess for $u2_i$ using $dpdu_i$.

5. Solve the decoupled equations for the $u2_i - pInlet$ boundary conditions

6. Calculate the difference between your target outlet pressure and the new computed outlet pressure as *diff*$=pOutlet - p2_i$

7. Iterate steps 3-6 until *diff* is below a certain tolerance value

3.3.3 Fluid-Structure Coupling

The structural and fluid models are coupled through the area-function that is the input to the fluid simulation, and the aerodynamic pressure that is used to compute the forcing vector for the FEM solid mechanics. We choose to loosely couple the solid and fluid models rather than have a combined formulation; this enables us to treat the area function as a pre-computed quantity for solving the 1D fluid model sequentially. The structural model is discretized in the coronal plane, with the fluid model computed along the centre-line/midline in Figure 3.2. Velocity components perpendicular to the midline are considered to be zero as we purely focus on 1D flow. A fourth-order asymmetric scheme is used for spatial discretization of the fluid model. At every discrete point, a cross-sectional area is extracted from the

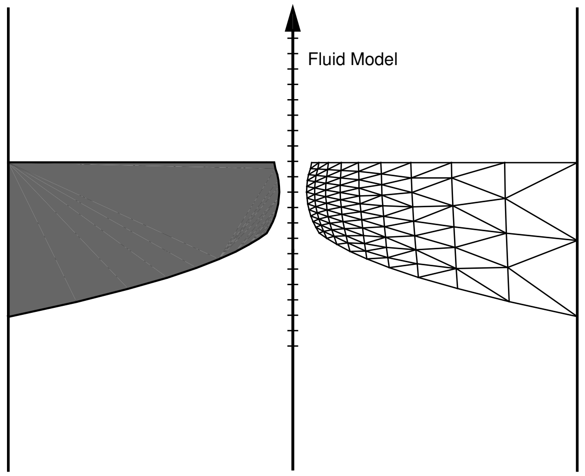


Figure 3.2: 2D continuum vocal fold model in the coronal plane. The vocal folds are assumed symmetric with the shaded out vocal fold (left) not simulated. The fluid model is calculated along the glottal centre-line

structural model, which is the medial-lateral opening between the two vocal folds, multiplied by the distance in the dorsal-ventral plane. The triangular nature of the glottal opening is taken into account when computing the cross-sectional area A and the associated perimeter s for the fluid model. Another important case, is that of collisions during the self-oscillation process. At every time-step, the updated area function is extracted from the structural model, and passed to the fluid model. Since equation 3.5 has terms divided by A, we choose to handle the fluid model area through warping as suggested in [12] with a 'safe' Area function:

$$A_{safe}(x,t) = A(x,t) + A_{closed} * w(A(x,t)) \tag{3.11}$$

The transition function is defined as:

$$w(A(x)) = \begin{cases} 0, & \text{for } A(x) > A_{small} \\ \frac{A(x)-A_{small}}{A_{closed}-A_{small}}, & \text{for} A_{closed} \leq A(x) \leq A_{small} \\ 1, & \text{for} A(x) < A_{closed} \end{cases} \tag{3.12}$$

where A_{closed} is the smallest numerically stable area that is empirically

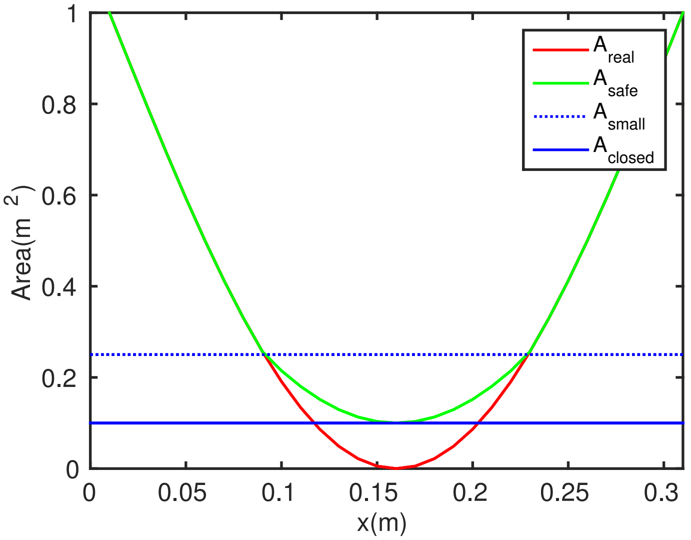


Figure 3.3: Sample area warping for fluid model. A_{safe} is used as the area function for the 1D fluid simulation to ensure stability.

determined, and A_{small} is the area at which transition begins (shown in Figure 3.3). Anderson et al [12] suggested a value of $A_{small} = 2.5{*}A_{closed}$.

The solid model is allowed to collide with the mid-line in the structural simulation (as shown in Figure 3.2) and a collision force is applied to the surface nodes. Thus, we enable the model to have realistic behaviour in both the structural and fluid domains. To ensure stability, the nodes are assumed to be in a state of collision, while the minimum glottal area is lower than A_{closed}. The pressures at the surface nodes of the FEM mesh are assumed to be equal to the concurrent computed pressures at the glottal mid-line. A linear interpolation is used to find the pressures at each surface node of the FEM body. The Force Vector on the node is then computed as follows:

$$\vec{F}_{node} = p_{node} * A_{node} * \vec{n} \tag{3.13}$$

where p_{node} is the pressure at the node, A_{node} is the effective nodal area shared between the elements the node belongs to, and $\vec{n}$ is the unit normal vector to the nodal surface. Both the fluid and structural models are temporally discretized using a central scheme with same time-stepping.

3.4 Results

Our model is implemented in MATLAB [83], a high-level matrix-based computing environment and programming language. This is done for two main reasons: firstly, MATLAB provides us with an unified environment to build our vocal fold models, and vocal tract implementations on. Since MATLAB uses a Java-based API, it can also interface with ArtiSynth [77], enabling us to build complete articulatory synthesizers including biomechanical models. The second reason is to provide speech researchers with a toolkit that can be used to test the different vocal fold and vocal tract models. Most speech researchers do not come from an engineering background and thus MATLAB is the easiest starting point in terms of language complexity compared to other lower-level languages such as C++ and Java.

The 1D fluid model is applied to three test cases. First, the accuracy of our decoupled implementation is verified for a standard problem in the field. Then it is applied to driving a vocal fold model for a static boundary-condition problem. Here, we look at it's prediction of the flow-separation point as well as the pressure-velocity distribution that it predicts.

3.4.1 Fluid Model Validation

We choose the same problem defined by Anderson et al [12], which is an example of the 2D starling resistor class of problems. This class of problem involves flow through a flexible tube connected to two rigid ends; readers are referred to a review for the different types of these models [22]. We define the following area function for the model:

$$A(x,t) = A_0 - A_m sin(\pi x) sin(\pi t) \tag{3.14}$$

where A_0 is the initial area and A_m is the magnitude of the collapse of the tube (or more colloquially, the extent of constriction of the tube), which are both constants. We define $\alpha_{min} = min(A(x,t))/A(0,0)$ which implies that:

$$A_m = A_0 \cdot (1 - \alpha_{min}) \tag{3.15}$$

Thus lower values of α_{min}, implies a more constricted tube, with values close to zero implying almost complete closure. We assumed mixed $u - p$ inlet boundary conditions for the problem, and no viscous losses to enable an analytical solution $\tau(x,t) = 0$. For these constraints we get the analytical solution for the mass (3.4) and momentum equations (3.5) as:

$$u(x,t) = \frac{1}{A}(u_0 A_0 + A_m cos(\pi t)(1 - cos(\pi x))) \tag{3.16}$$

$$p(x,t) = \pi \rho A_m \int \frac{sin(\pi t)}{A}(1 + (u^2 - 1)cos(\pi x))dx - \rho u^2 + c_p \tag{3.17}$$

Readers are referred to [14] for a derivation of the solution. Equation 3.17 is integrated numerically, and the constant c_p is defined numerically. While this model undeniably does not include viscous losses that play a significant role in vocal fold simulation, it serves as an useful tool to ensure that our decoupled fluid solver performs as well as the coupled solver that was originally proposed [12]. Table 3.4 gives a summary of the values and boundary conditions (BCs) used for the simulations. A non-dimensional pressure error p^* was defined comparing the analytical and numerical implementation solutions:

$$p^* = max(|p(x,t) - p_a(x,t)|)/max(\rho \cdot u(x,t)^2) \tag{3.18}$$

where p_a is the analytical solution. The error is calculated as part of a mesh-refinement study identical to Anderson et al [12]. The study is performed for $0.01 \leq \alpha_{min} \leq 0.99$, for meshes of 3 different discretizations: a coarse mesh ($\Delta x = 0.1$, $\Delta t = 0.02$), a medium mesh ($\Delta x = 0.05$, $\Delta t = 0.01$) and a fine mesh ($\Delta x = 0.025$, $\Delta t = 0.005$). The result of the experiments are shown in the figure 3.4, where the pressure error ($p*$) is plotted against α_{min}.

The main observations of the results from figure 3.4:

- As expected, the error p^* reduces with the improvement in mesh equality. We see an illustration of the 2^{nd} order accuracy of the discretization method, as we have an approximate reduction in the numerical error by a factor of 4 when the mesh is refined by a factor of 2.
- We see that the error increases rapidly for highly constricted geometries ($\alpha_{min} \approx 0$). Thus, defining a finite $A_{safe}(x,t)$ in equation 3.11 is important to achieve a stable solution.
- For the finest mesh, we get good pressure results up to $A_{closed}/A_0 = 0.05$ or 95% closure. This serves as a valuable tool in choosing both mesh equality and area warping for our vocal fold solver.

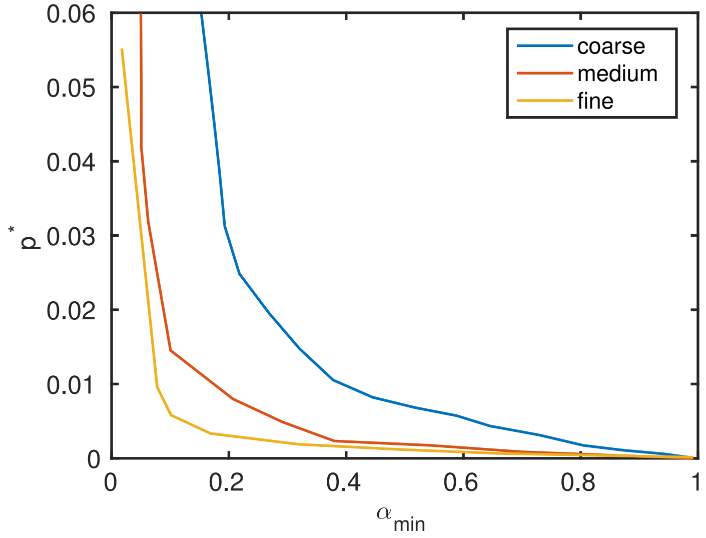


Figure 3.4: Mesh refinement study where numerical error p^* is plotted as a function of the factor α_{min} for three levels of mesh quality. Coarse mesh (blue), Medium mesh (red) and Fine mesh (yellow)

- Our results are practically identical to those of Anderson et al [12], with a general error difference (Δp^*) less than 10% between the methods. This implies that our decoupled solution is almost identical to the coupled solution despite potentially being much faster.

To confirm our model's computational advantages over the coupled solution of equations 3.4 and 3.5, we compare their time-based performances. We take the previous analytical equation 3.14, and solve it using both our decoupled solution scheme and the coupled solution scheme. We use MATLAB's internally built *tic−toc* functions to achieve comparable values. Table 3.3, shows that our model is almost 50 times faster than the coupled simulation when taking a Δx value of 0.01 and 0.005 for the analytical problem in question.

However, we are not able to achieve a relevant direct comparison to the 2D Navier-Stokes equation. This is because of two main reasons: firstly, the 2D Navier-Stokes is generally solved using dedicated finite-element toolkits such as ADINA making it an unfair comparison to our MATLAB-based solvers. Secondly, the standard 2D Navier-Stokes implementations for MATLAB are usually quite primitive, only allowing for steady-state problems

Discretization Δx (m)	**Solver**	**Time (s)**
0.01	Decoupled	2.046032
0.01	Coupled	91.330272
0.005	Decoupled	4.032497
0.005	Coupled	195.852746

Table 3.3: Performance of decoupled solver vs coupled solver. Simulation conducted for a time period of 2s with $\Delta t = 0.02$s. Length of channel is 0.6m.

or purely rectangular domains. They also suffer from instability arising from lack of accurate turbulence estimations, non-convergence of the Newton method for solvers for non-linear problems and lack of numerical stabilization [99][93][134]. These are often built into commercially available toolkits at the downside of computational cost. However, in general we can see that our model even for a rectangular domain will be much faster than a 2D Navier-Stokes equation. Taking a standard finite-difference discretization ($100X1 = 100$ cells for the 1D model and $100X100 = 10000$ cells for the 2D model) , we can see that for a refinement of the mesh by a factor of 2, would imply a 4 times increases in total number of cells for a 2D model ($200X200 = 40000$ cells). On the other hand, this would only equate to a 2 times increase for 1D model ($200X1 = 200$ cells). This can quickly add up, as we attempt to find a compromise between the mesh quality and the pressure error as seen from figure 3.4. This is an even greater issue when using the FVM or FEM for solving the fluid; achieving requisite mesh quality for FSI simulations is an extremely challenging problem. Thus, our model implementation represents an extremely fast solution of the flow equations, in comparison to other 1D model implementations and the 2D Navier-Stokes models.

3.4.2 Flow-Separation Experiment

A critical feature for a robust vocal fold model is its ability to estimate flow separation points. Pelorson et al [91], showed the importance of flow separation on the self-oscillation of the vocal folds and created a theoretical model to augment the Bernoulli's equation for estimating the flow-separation point. However, most 1D models still make an assumption based on area ratio's, instead of estimating the actual flow separation location. (Eg. flow

Name	**Value(s)**
Inlet velocity BCs u_{inlet}	1.0 m/s
Inlet pressure BCs p_{inlet}	0.0 Pa
Density (ρ)	1.2 kg/m^3
A_0	0.2 m^2
Length of Domain (x)	1.0 m

Table 3.4: A summary of the parameters used for the solution of the analytical problem. Parameters derived from [12]

separates when cross-sectional area is 1.2*a_{min}, where a_{min} is the minimum glottal area). Our fluid model directly accounts for the flow separation through the χ and τ_χ terms.

Figure 3.5 shows an illustration of the model's flow-estimation capacity. We use a modified vocal fold where the inferior edge of the glottis is fixed and the superior edge is allowed to oscillate. We look closer at the upper surface to understand the results better. In general, the flow separation point is between **1.2** to **1.4** times of the minimum glottal area which is within the range given in literature [39]. The results are qualitatively similar to those obtained by Alipour and Scherer [8], when a computational flow model was used to study flow separation further. Thus, we can now attempt to use the model as part of a coupled structural fluid simulation.

3.4.3 Coupled Vocal Fold Simulation

We now run a coupled vocal fold model simulation to validate the model's phonatory response. To achieve a fair comparison, we create a set-up as similar as possible to both experimental data as well as previously published models. The current gold-standard in the field is the quasi-3D finite element model by Alipour et al [9], which is one of the few continuum models that has been simulated in a full coupled acoustic simulation. Similar to the paper, we implemented a trachea and vocal-tract acoustics model using the 1D Kelly-Lochbaum digital filter method [69]. This enables a coupled light-weight acoustic simulation to validate our vocal fold model. Figure 3.6 shows the vocal fold shapes over one cycle of phonation. It can immediately be seen that our model reproduces one of the major features of phonation, which is the *vertical travelling wave*. However, it can also be seen that the lower edge of the vocal folds, which leads the collision phase of the vocal

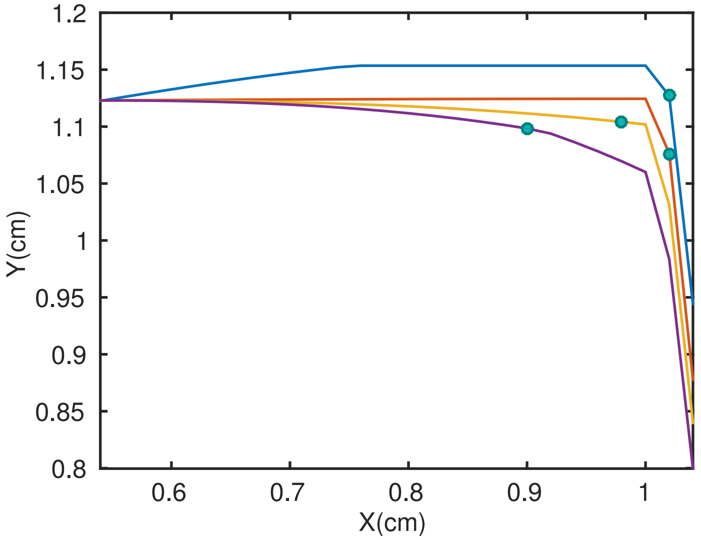


Figure 3.5: Flow-separation prediction of the flow model when inferior end of the vocal folds is fixed and superior is allowed to oscillated. The filled dots represent the point of flow separation, assuming flow from left to right

fold oscillation, does not have a significant collision period.

To truly validate the vocal fold model, we need to compare its time-dependent pressure and flow behaviour at various stages of the glottal cycle. In figure 3.8, the poses taken by the vocal fold structure during different time-steps of the phonation cycle are shown. Figure 3.9 gives the centreline velocity values along the axial distance for selected frames. Finally, figure 3.10, gives the respective pressure values for the selected frames.

As suggested by Alipour et al [9], we can see some clear patterns in the pressure-velocity distributions of the system. The selected frames display important points in the overall phonation cycle, .i.e. a convergent, divergent and neutral glottis. There are specific characteristics we expect to see in the pressure and velocity distributions of each these frames; for example, we expect negative pressures during vocal fold closing as this enables the vocal folds to be pulled back together. This is followed by a pressure recovery within the glottis itself. These characteristics are clearly seen in the figure 3.10, and will be discussed in further detail in section 3.5.

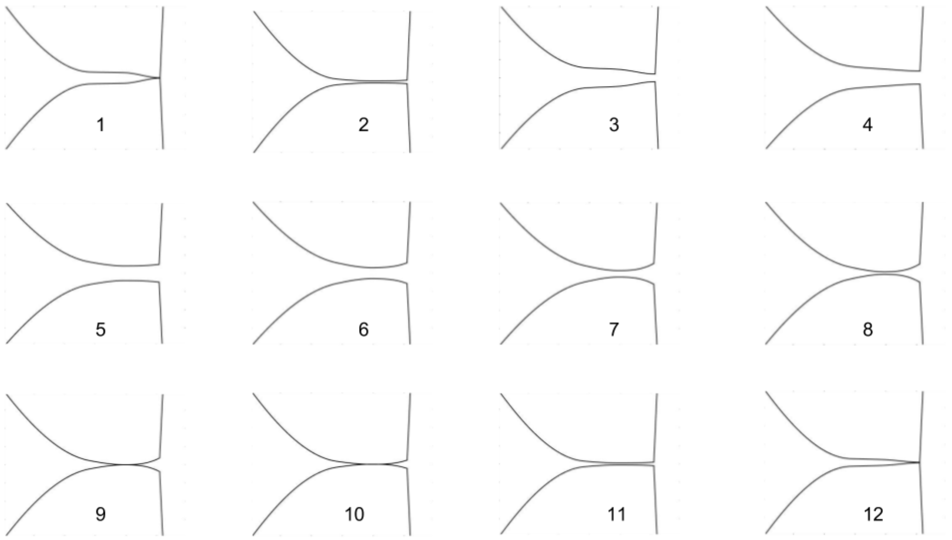


Figure 3.6: Vocal fold model shapes in one cycle of vibration. Fundamental frequency of oscillation is 146 Hz. Time difference between each time step is approximately 0.57 ms

A particularly important quantity for speech synthesis is the glottal flow (U_g). The glottal flow is given as the output from the vocal folds into the vocal tract, and acts as the excitation source to the vocal tract filter. Figure 3.7 shows the glottal flow waveform (above) and the time-varying sub-glottal pressure waveform (below). The vocal fold model is coupled to the trachea through the subglottal pressure; one of the challenges is to ensure model stability even when the subglottal pressure varies rapidly in time. As we can see, there are significant high frequency variations in the subglottal pressure seen in figure 3.7; this value is the boundary condition to the input of our vocal fold fluid simulation. These variations are also visible on the left extremity of the graph 3.10.

3.5 Discussion

In this section we look at the results in context of observations noticed in experimental studies and other canonical papers of the field. The extreme

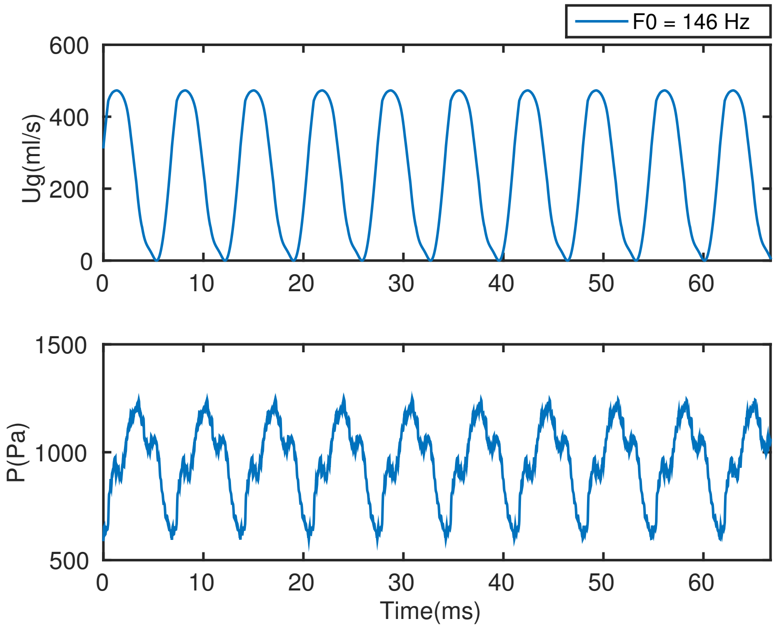


Figure 3.7: Typical glottal waveforms including glottal volume flow (top) and subglottal pressure (bottom).

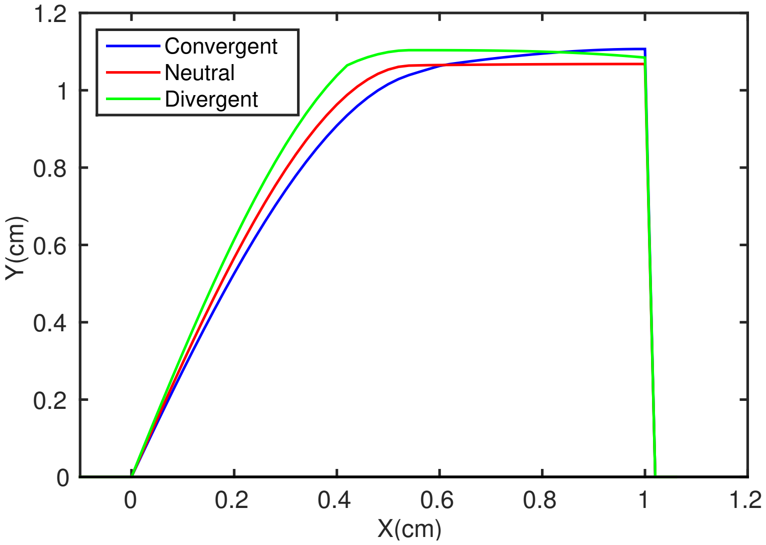


Figure 3.8: Selected vocal fold frames for the convergent, neutral and divergent glottal shapes

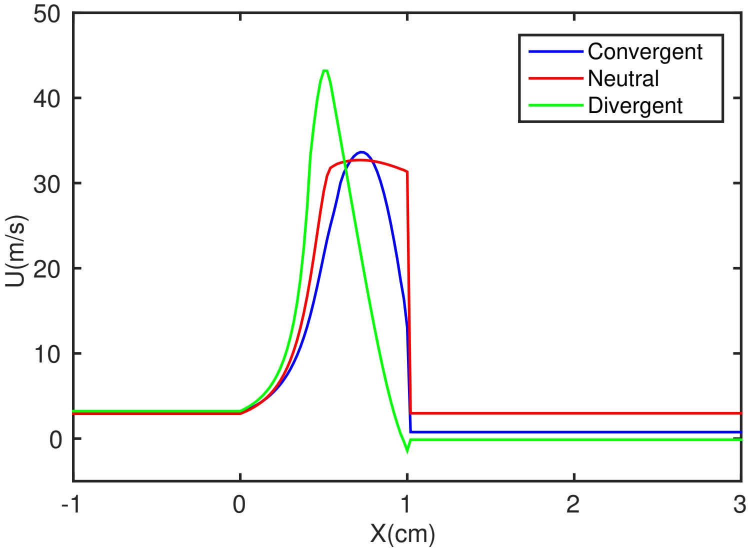


Figure 3.9: Centerline velocity predictions for the convergent, neutral and divergent glottal shapes

difficulty in accessing the vocal folds has meant that apart from the glottal flow data, directly comparable data is extremely sparse in the field. Thus, most of the comparisons to other models in literature and experimental data are qualitative. It should be noted that the paper by Alipour et al [9], included the false vocal-folds as well in their formulation; this would cause some natural deviation in the distributions especially towards the outlet of the glottis. However, the earlier models by the group [7][3] are direct comparisons to our model.

Vocal Fold Motion

Figure 3.6 showed the different shapes taken by the vocal fold model. The glottis' behaviour was in line with previous values from literature. There is a convergent shapes during glottal opening and divergent shape during glottal closing. The glottal angles ranged from **60** degrees divergent to **48** degrees convergent. As noted previously [9], this implies the presence of a robust mucosal wave. One of the the drawbacks mentioned previously, was with regards to the short collision time that was observed in the model. This is most likely a consequence of the area warping that is carried out for the fluid simulation; flow in the glottis still exists but at significantly

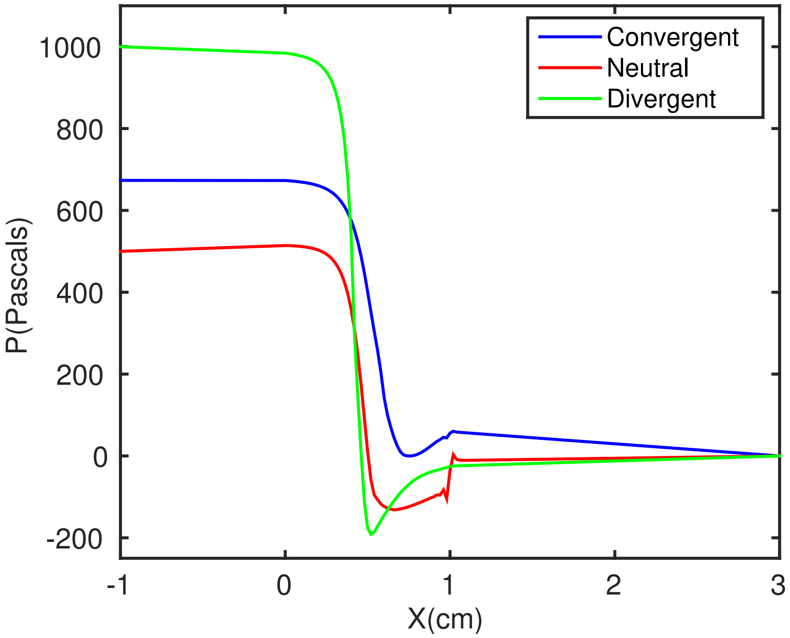


Figure 3.10: Centerline pressure predictions for the convergent, neutral and divergent glottal shapes

Glottal Configuration	**Peak V_p**	**Peak V_a**	**Peak Ps_p**	**Peak Ps_a**
Convergent	34 m/s	33 m/s	670 Pa	690 Pa
Neutral	33 m/s	32 m/s	520 Pa	515 Pa
Divergent	43 m/s	45 m/s	990 Pa	950 Pa

Table 3.5: A comparison of peak centerline velocities and subglottal pressures of the model presented in the paper (V_p, Ps_p) and from Alipour et al [9] (V_a, Ps_a). Note that the values from literature are estimates from published graph data

diminished levels. This causes the vocal folds to rebound faster than they otherwise might when the flow is zero. A potential improvement to the fluid model would be the addition of a τ_{small} term that artificially damps down the pressures down even further when the area of the fluid model is warped. This will ensure that the main pressure that acts on the vocal folds is only the impact pressure, P_{im}.

Velocity Distribution

The velocity distributions generally increased through convergent shapes and decreased through a divergent glottal shape. This is reflected in the higher peak that the divergent glottis has, with a much quicker fall off. This also points to flow-separation taking place earlier in the cycle. One of the interesting observations we can make is a slight negative velocity that is predicted by the model at the outlet of a strongly divergent glottis as shown in figure 3.9. This would correspond to the location of the downstream vortex that has been shown to play a significant role in the 3D glottis. This is both a positive and a drawback for the fluid model: it is positive that the model is able to predict the likely rise of turbulence at the outlet of the divergent glottis, however, a negative velocity is not a physically realistic characterization of turbulence in a 1D fluid model, especially when it is used to calculate the velocity flow. However, this does not affect the overall flow of the model, which is heavily predicated on the velocity at the point of minimum glottal area. Table 3.5 shows that there is strong agreement in peak values predicted by our model in comparison to vocal fold models driven by a 2D Navier-Stokes flow model in literature [9].

Pressure Distribution

As mentioned by Alipour et al [9], the convergent glottal shapes have high pressures within the glottis and decreasing pressures towards the end. The divergent glottis on the other hand tends to show a dip in pressure near the minimum constriction area. This is seen as a negative pressure in the figure, that acts on the vocal folds to pull them together for collision. It is also clearly seen that the most open glottal shape in figure 3.8 has the lowest pressure distribution in figure 3.10 (neutral glottis: colour red). There is a noticeable pressure recovery that takes place at the outlet of the vocal fold for the divergent glottal shape; this can again be attributed to the flow separation in the divergent glottis and associated pressure recovery. Again as seen previously, table 3.5 shows that there is again agreement in peak values predicted by our model in comparison to vocal fold models driven by a 2D Navier-Stokes flow model in literature [9]. Thus, the 1D model does an impressive job of replicating a lot of features of higher-order models in the results.

Glottal Waveform

The glottal waveform is arguably the most important data point to evaluate the vocal fold model as it acts as the output to the vocal tract simulation. There are a few characteristics that are readily apparent from 3.7:

- The glottal flow is slightly skewed to the right but is generally quite symmetric. This is similar to other continuum models in literature [3][113], but dissimilar to experimentally observed data. We expected the opening phase to be significantly more gradual than the closing phase of the model. This is since the negative pressures inside the glottis act quickly with the myoelastic muscle forces, to pull the vocal folds back together. A possible reason for this disparity is the broader and less concentrated negative glottal pressure for the divergent glottis shown in figure 3.10.

- The average value of the flow is **241 mL/s** and the peak flow is about **473 mL/s**. This is in good agreement with published data; Alipour in his canonical paper [9] reported an average flow value of 191 mL/s and a peak flow of just over 400 mL/s.

- The ratio of the closed cycle to open cycle or open quotient is about **0.9**. This is significantly different from lumped-element models in literature that usually report cycles of about 0.6-0.7 [26]. However, this is again similar to other continuum models in literature. We also see that complete closure is achieved for an extremely small part of the vocal fold phonation cycle. An obvious explanation, that was also mentioned previously, is the role of the area-warping in reducing the closed cycle to open cycle ratio. As an extension of the model, we can look at including a τ_{small} term to add extra viscous losses to simulate complete closure.

- The fundamental frequency of the flow is **146 Hz**. This is identical to the value reported by Alipour et al [3] in the paper containing identical geometry to our vocal folds (146 Hz).

- The subglottal pressure (P_{sg}) predicted by our model shows very strong resemblance to the inferior glottis pressure recorded by Alipour et al [6] during experimental studies of dynamic glottal pressures on excised larynges. This shows that our model does an excellent job of replicating the pressure conditions inside the glottis during phonation.

3.6 Summary

In this chapter, we have presented a vocal fold model to meet the requirements of articulatory speech synthesis. We started by enumerating the considerations that are there when choosing appropriate constituent models for our simulation in section 3.1. Section 3.2, introduced our structural model of choice with the numerical implementation procedure. The model is solved in the structural domain using a 2D finite-element method procedure and was discretized using a 2^{nd} order central scheme. It was also identified in section 3.1 that a major lacunae in the field, is the fluid models used to predict the aerodynamic pressures and velocities in the glottis. In particular, prediction of intraglottal pressures, separation point and overall glottal flow rate were shown to be critical factors that determined the performance of the flow model. In addition, only higher dimensional models of the flow were currently capable of estimating the viscous losses that play a critical role in the pressure recovery, and as a consequence the bulk intraglottal pressure distribution. However, 2D and 3D Navier-Stokes solvers are computationally prohibitive, forcing speech researchers to resort to low-dimensional models of the flow to build articulatory synthesizers.

Based on these criteria, we introduced and validated a novel 2D finite-element continuum model loosely coupled with a 1D unsteady fluid model. This model is unique in a number of ways. It is the first unsteady 1D fluid model that has been used for vocal fold vibration dynamics and first such combination of structural and flow models in literature. The model is loosely coupled to the solid model, making it possible to use extremely sophisticated solvers in the future seamlessly. The model is presented and formulated in section 3.3. The flow model is capable of handling irregular geometries, different boundary conditions and closure of the glottis. We propose a method for a fast decoupled solution of the flow equations that does not require the computation of the Jacobian matrix. We create an implementation where the model is discretized with a 2^{nd} order temporal accuracy and 4^{th} order spatial accuracy scheme.

The numerical implementation of the fluid model is validated for a standard problem with an analytical solution, and then combined with the structural model. A coupled vocal fold simulation is performed with a trachea and vocal tract model designed using the Kelly-Lochbaum wave-digital filter method. The simulation results are compared with data from literature and shown to be in good agreement. However, significant further work is still required in improving vocal fold simulations. This is discussed in detail in the final chapter.

Chapter 4

Coupled Articulatory Synthesizer

The main goal of our vocal fold model is articulatory speech synthesis. In this chapter we shall build an articulatory synthesizer using our vocal fold model as one of the components. In particular, we hope to illustrate the vocal fold model's utility in the context of articulatory speech synthesis and the effect of coupling on the system output. Section 4.1 describes the model formulation and the different components that are involved in building the model. In Section 4.2, preliminary results are given from the system in question. Section 4.3 discusses the significance of the results in the context of the overall field of articulatory speech synthesis.

4.1 Model Formulation

Building an articulatory synthesizer includes a range of components: a vocal fold model (which includes inside it a structural and flow model), a biomechanical model of the vocal tract and its articulators, and a vocal tract acoustics model that simulates the acoustic pressure field and gives us the final radiated pressure from the mouth. In Chapter 3, we explored a new 2D structural model of the vocal folds coupled with an 1D Navier-Stokes based unsteady flow model. Figure 4.1 gives a look at the different components that are part of our articulatory speech synthesizer. In further subsections 4.1.1, 4.1.2 and 4.1.3 we explore the individual components in greater detail.

4.1.1 Vocal Tract Biomechanics

The vocal tract biomechanics is primarily involved with defining a domain for the acoustic simulation. The vocal tract biomechanics can be defined in two major ways: meshes of the vocal tract can be directly derived from published data or from a connected biomechanical model. Most models in literature used vocal tracts that are either directly derived from meshes

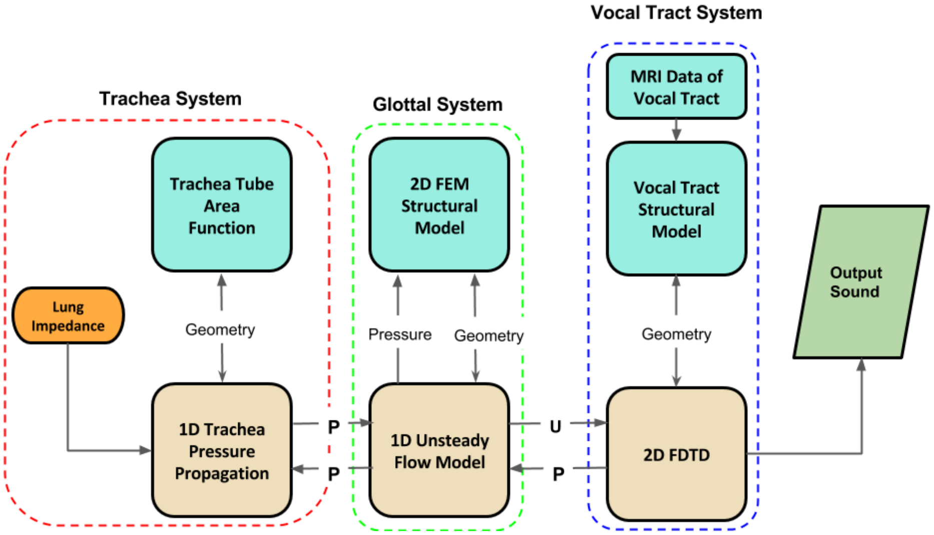


Figure 4.1: Coupled articulatory synthesizer flowchart

or converted to area function notation as shown in figure 4.2. Fant [42] and Story's [106] data sets, still remain extremely important in the context of building 1D and 2D vocal tract models. There is a natural correlation between the area function description and 1D-tube representation of the vocal tract. The vocal tract is divided into concatenated tubular sections of different diameters as shown in figure 4.3. A point to note is that while figure 4.3 shows a nasal section, this is not required for our simulation as we are only modelling vowel shapes. However, the 1D representation loses a lot of the asymmetric structural information that plays a critical role in the perceptual quality of generated speech; thus, 2D and higher order models have become more popular in literature. The representation of 3D data in 2D form is still an interesting open problem in the field that is now being explored. Arnela et al [15] for example, have proposed methods to represent 3D mesh data in 2D while simultaneously capturing as much of the modelling intricacies as possible. The focus of these systems is to preserve the 3^{rd} dimensional spatial information, while reducing computational load that comes with the additional dimensionality.

The vocal tract biomechanics can be seen as an additional step that is built on top of this process. First, the airway mesh is extracted for the vocal tract pose at the time instant under consideration. This is then converted

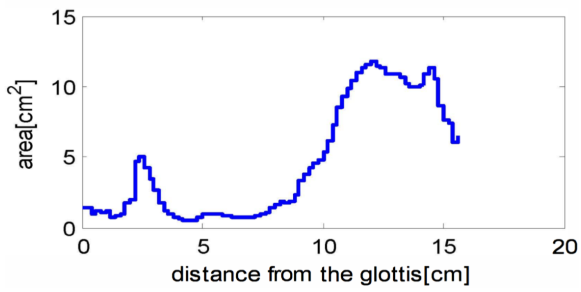


Figure 4.2: Area function representation of vocal tract

to the dimensionality of the fluid solver (1D, 2D or even full 3D) and used to build a structural domain. Finally, the acoustic equation is solved over this domain. Recently, Anderson et al [13] introduced the latest edition of the FRANK model referenced in chapter 2. This model includes a comprehensive list of hard and soft articulators that can be controlled for speech synthesis using muscle activations in a physics simulation environment. A similar model was also created by Dabbaghchian et al [35]. The 3D airway mesh can be extracted, and used as the domain for the pressure wave simulation of the vocal tract as shown in figure 4.4. This is later rasterized to transform it to a 2D domain. As the shape of the airway changes from time-step to time-step, we extract the new airway mesh to define our structural domain. Thus, this enables the simulation of the vocal tract biomechanics. However, in our current simulation, we focus on preliminary results using simple symmetric tubes. This is because the conversion of the 3D mesh to a 2D rasterized domain is a process that has not been completely validated in previous papers [15][132]; therefore we focus on standard data in literature for our simulations instead of adding greater uncertainity to the process.

4.1.2 Vocal Tract Acoustics

As mentioned in Chapter 2, there have been a range of models suggested to simulate the acoustics of the vocal tract. They can generally be divided into 1D, 2D and 3D models in terms of dimensionality or into different categories based on the type of simulation (eg. digital wave-guide filter, direct numerical flow simulation). For articulatory synthesis, the direct numerical simulation model is particularly attractive. While the other models are

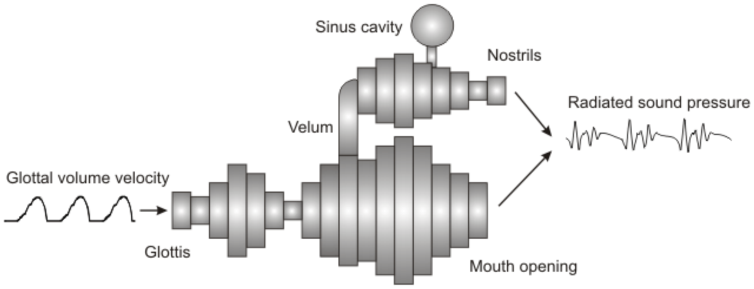


Figure 4.3: Concatenated tube model used for 1D speech synthesis

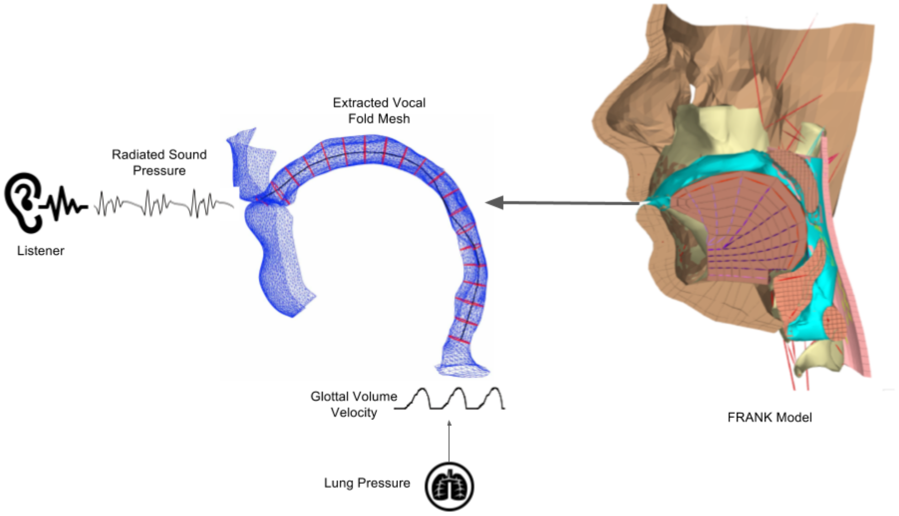


Figure 4.4: Framework for biomechanically driven articulatory speech synthesis. The vocal tract airway mesh is extracted from the FRANK model over time and used as the domain for the acoustic simulation. It is coupled with the trachea model and vocal fold model to create a complete articulatory synthesizer.

potentially faster or easier to implement, we can gain a fundamental understanding of the pressure propagation and velocity of airflow through this method. It also ties in logically with the rest of the components of the system, i.e. vocal fold models and biomechanical model.

Faster simulations can be achieved by using a simplified representation of the vocal tract, consisting of a straight concatenation of cylindrical segments [104], and by bounding propagation in one dimension only. This approach eventually allows to reach real-time simulation rates and good results particularly for the first formant of the tube [126][24][107]. However, this method has some drawbacks: the model doesn't include any higher order modes due to its straight symmetric geometries and struggles to naturally simulate forks/cavities in the vocal tract. While there have been some models that have suggested remedies [88][60], these model produce loose simulations in the upper end of the spectrum which affects the naturalness of the sound.

On the other hand, 3D models, while very accurate take a prohibitive time to simulate very short pieces of speech (e.g., 60 minutes [111], 44 minutes [15] for 5ms of audio).

2D Finite-Difference Time Domain

We base our vocal tract solver on the model suggested by Zappi et al [132]. They suggested a GPU-based 2D Finite-Difference Time Domain (FDTD) simulation as a compromise for high-quality synthesis. The model achieved real-time run rates and small positional errors in the first formants. However, the model calls for an extremely complex implementation scheme that enables the parallization of the solving of individual cells to achieve a real-time solution. For our 2D continuum vocal fold model, a solution on the shader would be difficult to achieve in conjunction with the 2D FDTD solver. Thus, we create an alternative implementation of the model that runs on the CPU instead, where computational speed is sacrificed for system inter-operability and robustness. There exists the future potential for a complete continuum model implementation on the GPU in conjunction with the 2D FDTD simulation; this would be a natural step forward after validation of the 2D VF model. We also create shader implementations of canonical lumped-element models to act as the glottal excitation for the GPU-based 2D FDTD system. Details of these implementations are given in subsection 4.1.4. The equations to be solved is an augmented version of the two-dimensional wave equations written as:

$$\frac{\partial p}{\partial t} + (1 - \beta)p = -\rho c^2 \Delta.v \tag{4.1}$$

$$\beta\frac{\partial v}{\partial t} + (1-\beta)v = -\beta^2\frac{\Delta p}{p} + (1-\beta)v_b \tag{4.2}$$

where p is the pressure at a discretized cell, and v is the velocity. This equation allows each point in space to transition between fully solid (wall) and fully open (air) states, via a scalar parameter field $0 \leq \beta \leq 1$. The parameter is changed smoothly at time scales larger than the main system's vibrational time-periods. Thus the equation amounts to linear interpolation between the standard wave equation when $\beta = 1$ (air), and enforcing some prescribed velocity boundary conditions of $v = v_b$ when $\beta = 0$ (boundary). For intermediate values of β, the affected region acts partially reflective and partially transmissive.

The above equations are discretized and solved numerically similar to standard FDTD solvers, using second-order accurate spatial and temporal derivatives with velocities sampled on a staggered grid. The β field is sampled at cell centres and 6 Perfectly-Matched Layers (PML) are employed at the edges of the doman to absorb outgoing radiation. Since we use an explicit scheme for time integration, the time step Δt and spatial cell size Δx must obey the related Courant-Friedrichs-Levy (CFL) stability condition in two dimensions: $\Delta t < \Delta x/(c)^2$, where c is the speed of sound in the medium. The original paper [132] also validated the vocal tract system's performance by comparing it's impulse response to formants published in literature.

The choice of this model is logical in the context of our vocal fold simulation; we expect that the synergy of dimensionality between the vocal folds (2D) and vocal tract (2D) models will enable more complex geometries to be represented and their acoustic effects captured. This falls in the continuum between fast, but simplified 1D vocal tract models and complex, but computationally prohibitive 3D models. Equally, we avoid the dimensionality mismatch of Alipour et al [9], where a 3D vocal fold model was combined with a 1D Kelly-Lochbaum vocal tract. It is unclear if any of the potential benefits of the accurate 3D vocal fold simulation, can actually be gleaned through a 1D simplistic wave-reflection analog system.

4.1.3 Trachea Model

We use a simple 1D representation of the trachea in building this articulatory synthesizer; it is an implementation of the 1D digital wave-guide proposed by Kelly-Lochbaum [69]. As we saw in the previous chapter, the trachea

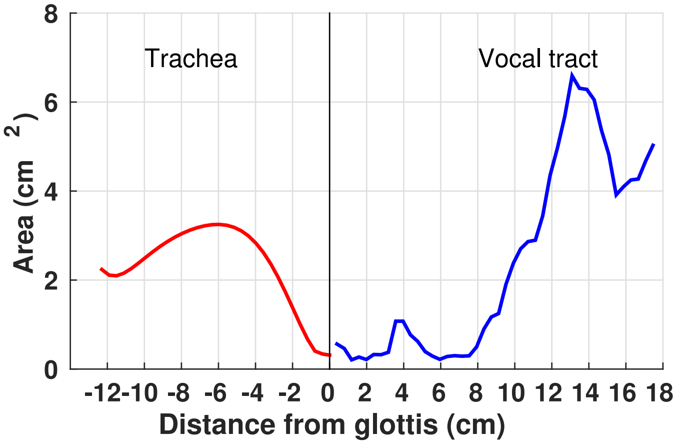


Figure 4.5: Area function representation of trachea with concurrent vocal tract model

model plays a critical role in the speech synthesis. A time-varying subglottal pressure (P_{sg}) signal drives the overall vocal fold phonation in addition to the epilaryngeal pressure (P_e). A standard lung impedance pressure of 800 Pa is assumed in this case, and the trachea geometry suggested by Story et al [107] is used. The area-function of the trachea that is used is shown in figure 4.5. Unlike the vocal tract, the trachea is a much more symmetric, cylindrical tube; this makes it an ideal candidate for an 1D representation without losing out on accuracy.

Name	**Value(s)**
Speed of Sound	350 m/s
Fluid density	1.14 kg/m^3
Sampling Rate	220500 Hz
Δt	4.535e-06 s
Δs	0.002244783432338

Table 4.1: A list of the model parameters used for the 2D FDTD simulation

4.1.4 Alternate Vocal Fold Models

To compare the performance of our vocal fold model, we create implementations of two canonical models in literature. Both the two-mass model [64] and the body-cover model [107] are implemented as part of the overall system. Thanks to the modular nature of the system, these can be switched in for the 2D FEM vocal fold model with no extra changes required to the overall functioning of the system. Both the models have shown that they are capable of reproducing the glottal waveform to a reasonable extent. They can also be coupled to the vocal tract simulation easily without the stability issues associated with a CFD simulation.

4.2 Results

We attempt to illustrate our vocal fold model's ability to drive an entire articulatory synthesis simulation in a stable and physically realistic manner. The model is coupled together as shown in figure 4.1: this includes the 1D trachea, the vocal fold models (two-mass, body cover and continuum) and the 2D FDTD vocal tract. Table 4.1 gives the values used to drive the 2D FDTD simulation. We used the area functions published by Story et al [106], where vocal tract area functions were derived from MRI scans of the human airway. This data is a standard for the field, and provides an excellent starting point for our synthesizer. Potentially, meshes can be directly extracted from ArtiSynth models [13], as well. Figures 4.6 and 4.7 show the vocal fold meshes that we use in the 2D FDTD simulation.

The discretization in the spatial and time domains are related by the Courant—Friedrichs—Lewy (CFL) condition for convergence of finite-difference numerical implementation: $\Delta s = \Delta t * c * \sqrt{2.0}$, where Δs is the spatial discretization, Δt is the temporal discretization and c is the speed of sound in the domain.

Coupled vs Uncoupled Vocal Folds

One of the major goals of our system is to drive a complete coupled articulatory synthesizer. We take the 2D domain shown in figure 4.6 and use our 2D continuum vocal fold model as the glottal input. The computed pressure waveform at the listener is recorded over time; the simulation is run for a time period of $1s$. The same experiment is carried out with an uncoupled version of our 2D vocal fold model instead. This would imply that instead of

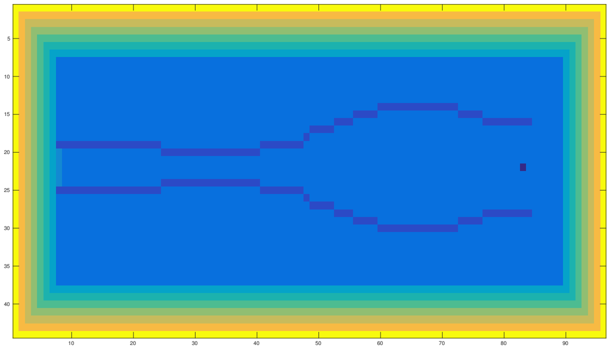


Figure 4.6: 2D vocal tract domain for vowel /a/. The boundary includes 6 Perfectly Matched Layers (PML) for absorption. The black dot represents the listener and the left end of the symmetric tube contains the glottal inputs and feedback pressure cells.

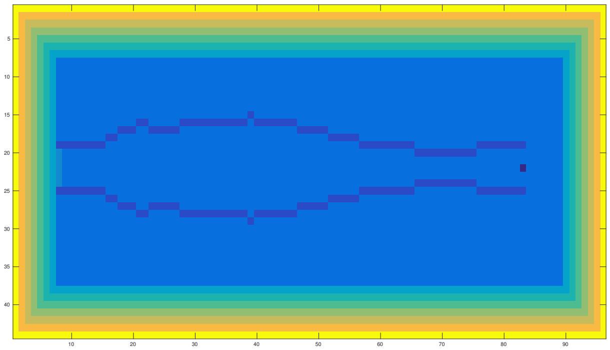


Figure 4.7: 2D vocal tract domain for vowel /i/. The boundary includes 6 Perfectly Matched Layers (PML) for absorption. The black dot represents the listener and the left end of the symmetric tube contains the glottal inputs and feedback pressure cells.

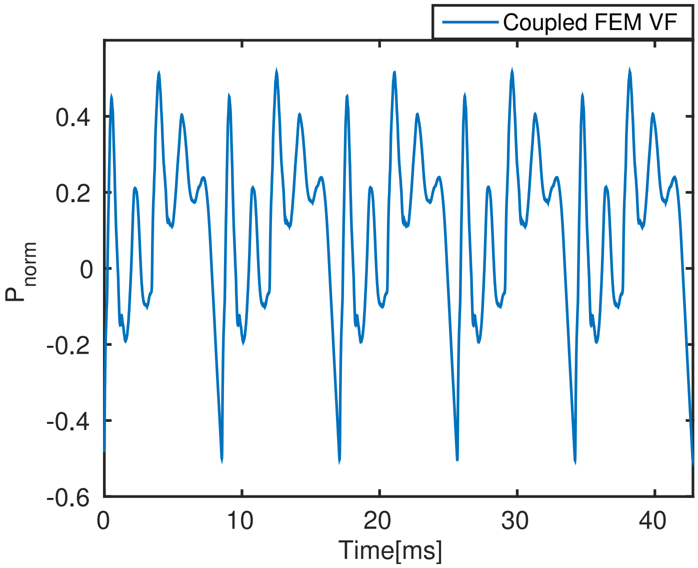


Figure 4.8: Normalized output pressure for a coupled 2D continuum vocal fold model used as glottal source to 2D FDTD simulation for vowel shape /a/.

having time-varying boundary conditions thanks to trachea and vocal tract coupling, this model will have static boundaries conditions instead. This can be seen akin to a vocal fold model driven by the entire $800Pa$ lung pressure, acting at the subglottal duct. The wave-forms for both models are shown below in figure 4.8 and 4.9.

We can immediately notice from the results the presence of a significant non-linearity in the coupled simulation that is not present in the uncoupled simulation. The coupled simulation has significant negative pressure troughs that are not visible in the uncoupled model. To understand if this is a function of the coupling or a simulation artifact, we also couple two lower-order models, the *two-mass* and the *body-cover* model to the simulation as a comparison. Since these models are quite primitive with simpler symmetric glottal flow (U_g) wave-forms, it is unlikely that they would be able to generate such a pressure non-linearity by themselves, thus suggesting that it would be a function of the non-linear coupling instead. Figures 4.10 and 4.11 present the normalized pressure outputs at the listener for the articulatory synthesizers driven by the two-mass model and body-cover

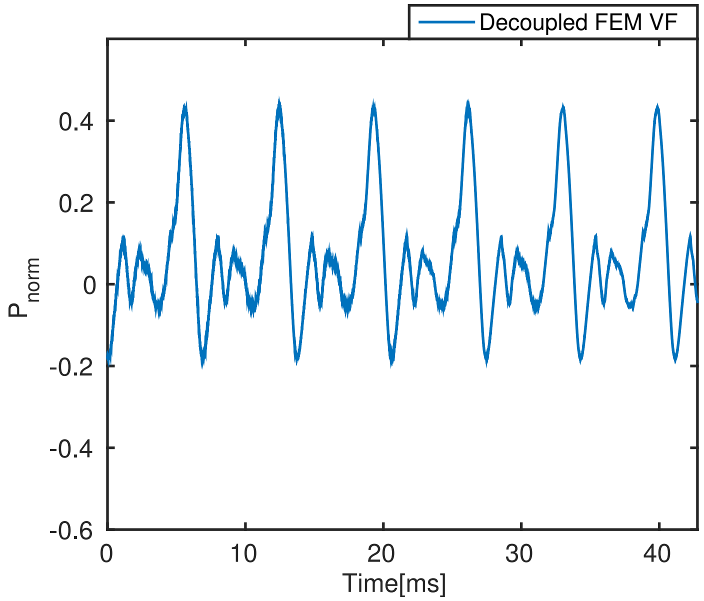


Figure 4.9: Normalized output pressure for a uncoupled 2D continuum vocal fold model used as glottal source to 2D FDTD simulation for vowel shape /a/.

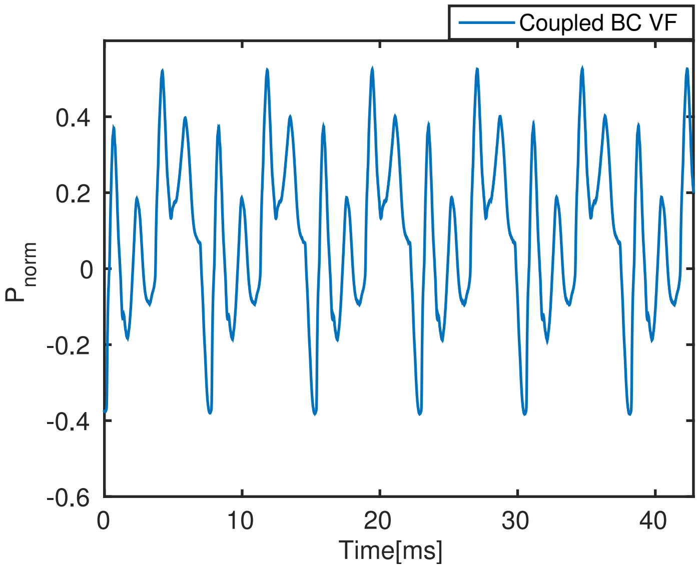


Figure 4.10: Normalized output pressure for a coupled body cover model used as glottal source to 2D FDTD simulation for vowel shape /a/.

model respectively.

While the individual wave-forms have different frequencies as expected, both the coupled two-mass and body-cover model also contain the identical non-linearity seen in the pressure waveform of our coupled 2D continuum model. This validates the significant role played by coupling in an articulatory synthesizer, and reaffirms the non-linear relationship that exists between the source and filter models as suggested by Titze [119]. Table 4.2 summarizes the maximum and minimum of the normalized output pressure in the different vocal fold models; we can clearly see that the uncoupled vocal fold model has significantly different results than even simple lumped-element models. It is important to note that the ratio shown in table 4.2 has no physical relevance; it's just a tool to understand the pretty significant differences between the waveform values. This is an interesting result moving forward; it is important to design continuum models that can also be coupled stably to ensure the maximum non-linearity of the source-filter relationship is captured.

Finally, as a reference we include a sample graph by Arnela et al [32] of the output pressure at the listener for a 3D vocal tract simulation. There

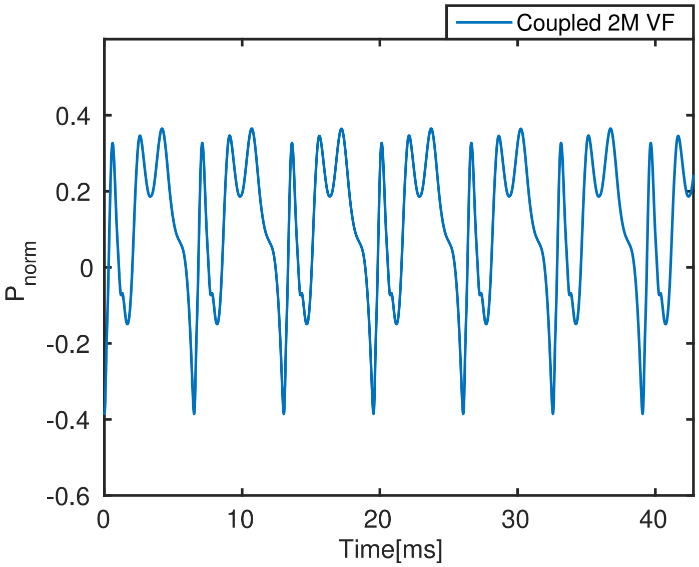


Figure 4.11: Normalized output pressure for a coupled two-mass model used as glottal source to 2D FDTD simulation for vowel shape /a/.

VF Model	**Max** P_{norm}	**Min** P_{norm}	**Max/Min(**P_{norm}**)**
Coupled Continuum VF	0.5163	-0.5044	1.0236
Uncoupled Continuum VF	0.4333	-0.1833	2.3639
Coupled BC VF	0.5260	-0.3829	1.3737
Coupled 2M VF	0.3648	-0.3850	0.9475

Table 4.2: A comparison of the maximum and minimum output normalized pressures predicted by different vocal fold models

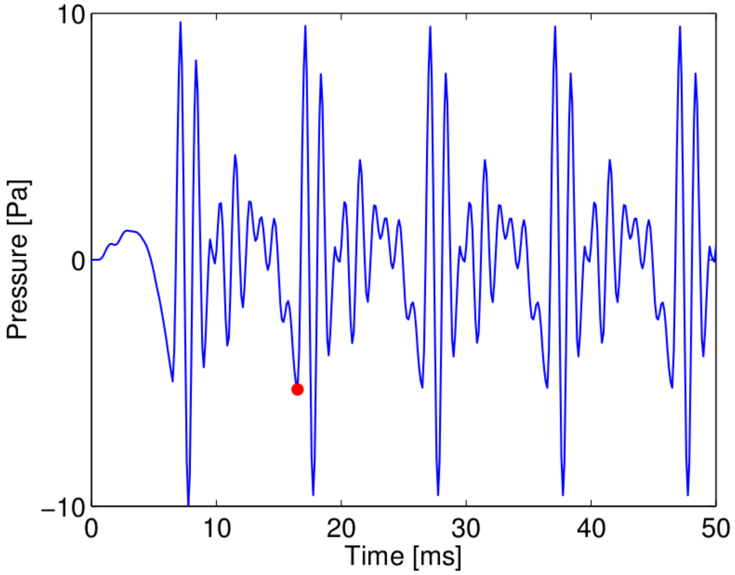


Figure 4.12: Output pressure for a 3D FEM acoustic simulation for vowel shape /a/. Reproduced from [32]

are a few significant differences which means that this is only an instructive graph for reference and is not a complete comparison. Firstly, the model uses a parametric vocal fold model suggested by Fant et al [43], with a fundamental frequency (F_0) of 110 Hz. This model is not coupled in the true sense; this is unlike self-oscillating models such as ours. Secondly, the implementation contains a head model of radiation that can significantly affect the final pressure waveform. Figure 4.12 shows the pressure waveforms. We can see that our results look reasonably similar in terms of overall shape. Since the results from the /i/ simulation were qualitatively similar with the presence of an identical negative pressure trough for coupled models, we choose not to reproduce them for the sake of brevity.

4.3 Discussion and Conclusions

Building an articulatory synthesizer is a non-trivial task. In this chapter, we have put forward one possible articulatory synthesizer utilizing our novel vocal fold model. While the individual components that constitute an articulatory synthesizer have been designed many times over in literature, making these disparate components work together is both conceptually and computationally challenging. The difficulties in achieving coupled stable articulatory synthesis has made speech researchers forgo the possibility of using more comprehensive models, for simplicity. In particular, the vocal fold model is often ignored as researchers prefer to focus their efforts on modelling the vocal tract acoustics; this is driven by the belief that the payoff is greater for the computational cost in vocal tract modelling.

In this chapter, we have attempted to lay down a modular articulatory synthesis framework, that uses a stable, coupled finite-element vocal fold model. Initial results are promising; the output pressures at the mouth seem in line with waveforms seen in literature. However, significant work needs to be done in testing and validating the entire articulatory synthesizer. In general the main contributions of this illustrative case-study are the following:

- **Coupling**: As mentioned in Chapter 2, the vast majority of vocal fold models remain uncoupled in literature. This is especially true in the case of continuum models, where coupling the model means time-varying boundary conditions for the flow simulation. Many flow models struggle to handle these rapidly changing boundary conditions in a stable, physically realistic manner. Our choice of 1D flow model makes it easier to achieve a stable coupled simulation as shown in section 4.2. Coupling also adds non-linearities that can significantly change the final sound generated.

- **2D Vocal Tract Model**: Apart from Alipour et al [9], this is the only complete articulatory synthesizer using a continuum vocal fold model in conjunction with models of the trachea and vocal tract. In particular, our model is the only model not using a 1D wave-reflection vocal tract, but solving the entire 2D wave-equation for the vocal tract as well. Both the 2D FEM vocal fold model and the 2D FDTD model lie in the unexplored space between 1D and 3D models, potentially acting as a bridge between the computational and conceptual advantages of the models respectively.

- **Accessibility and Modularity**: Most continuum models of the vocal folds are either solved using commercial finite-element tool-kits such as ANSYS[113] or ADINA[39] or implemented using proprietary code that belongs to labs [3][135] or even a combination of both [9]. This has served as a major hindrance in two major ways: speech researchers have struggled to access and reproduce results [4], and more importantly, combining these models as part of a larger articulatory synthesizer has remained an unachievable goal. In our model, the entire system is completely implemented in MATLAB (1D trachea + 2D FEM vocal folds + 2D FDTD vocal tract) and interfaced naturally. A link is also provided to the biomechanical toolkit ArtiSynth to drive the structural vocal tract models using biomechanics.

However, there are some significant shortcomings to this study. The results are very preliminary; the coupled system needs to go through vast testing to validate each component apart from just the simulation output. The vocal-tract model is extremely coarse by itself; it is still not clear how a 2D vocal tract should be represented unlike the simple 1D area function model and the full 3D mesh models. Finally, while the possibility for user-specified meshes exists, it hasn't been displayed yet in this system because of the above reasons.

Chapter 5

Conclusions

This thesis presents the definition and validation of a novel vocal fold model targeting the application of articulatory speech synthesis. Driven by the need for vocal fold models that balance the competing priorities of computational cost and complexity, we enable speech researchers to potentially move beyond simple lumped-element models for building articulatory synthesizers. To do so, we propose a model comprising of a 2D structural model loosely coupled with a 1D unsteady flow model; this enables us to combine the completeness of higher dimensional structural models with the computational advantages of 1D fluid models.

In Chapter 3, we presented a formulation and implementation of our vocal fold model. We start by taking a canonical 2D structural model that uses a linear-elasticity based constitutive equation. A hemilaryngeal continuum model is considered, with symmetry assumed across the glottal mid-line. We start by writing our 1D fluid equations based on the implementations of Cancelli et al [28] and Anderson et al [12]. The coupled solution of these equations however, requires the computation of the Jacobian that could potentially reduce many of the computational advantages gleaned by using a 1D model over a 2D Navier-Stokes based solver. We thus present a method for a fast decoupled solution of the flow equations that does not require the computation of the Jacobian matrix; this is achieved through an iterative bounded-search procedure that estimates the equivalent velocity-driven boundary conditions We finally couple these models together, using an aerodynamic force interpolation scheme.

First, we demonstrate the fluid model's performance for a standard analytical problem in literature; the model shows low non-dimensional pressure errors with the refinement of the mesh. These results are shown to be similar to previously published results using the coupled Jacobian approach, at a computational cost that is more than an order of magnitude faster. In addition, the model can be shown to have an even greater computational advantage over 2D Navier-Stokes equations. Secondly, the model is used to predict the flow separation point for a forced-oscillating vocal fold surface. The model predicts that the flow separates at cross-sectional areas between

1.2 to 1.4 times the minimum glottal area (a_{min}); this is in line with results in literature and shows that our 1D fluid model does not suffer from the major issues that plagues other 1D models based on Bernoulli's equation. Finally, we validate our model's performance for a full coupled simulation. This included validation of the vertical mucosal wave, velocity and pressure distributions over time and values of the glottal flow (U_g).

In Chapter 4, we looked at the feasibility of using our vocal fold model to build a complete articulatory speech synthesizer. We implemented an 1D wave-reflection analog of the trachea driven by a constant lung pressure. A 2D Finite-Difference Time-Domain (FDTD) solver based on the work of Zappi et al [132] was implemented in MATLAB to solve the 2D wave equation over the acoustic vocal tract domain. Apart from our vocal fold model, two canonical models of the field, the two-mass model and the body-cover model, are also implemented to provide a comparison. We showcase a framework to add user-specified meshes to the vocal tract solver; this can either be through a biomechanical toolkit such as ArtiSynth or through data from literature. We choose to use the area functions from Story et al [106], to create our symmetric vocal tract. The output wave was shown to be physically realistic for our vocal fold model, and in agreement with other exemplar vocal fold models.

5.1 Discussion

Our results demonstrate that the unique approach of coupling the 1D fluid model with a 2D structural model is appropriate for vocal fold phonation. The fluid-structure interaction of the vocal folds is dominated by the solid model, and is mainly driven by the bulk fluid pressure rather than minute variations in the flow. The significant differences in density and viscosity of the two bodies also implies that we do not need a tightly-coupled regime to achieve stability in the FSI. Equally, our method of using string energy to approximate the third dimensional behaviour works to our advantage; while the range of motion of the system is reduced, it is significant enough to capture the standard cycle of vibration.

In particular, we avoid the standard eigen/modal analysis that accompanies many papers in the field, based on two major reasons. Firstly, published work in literature has shown that despite many of the model's ability to entrain at expected ranges of primary modes, they fail to produce a glottal waveform that is physically reasonable. This leads to our second reason: with our goal being articulatory speech synthesis, we focus on the flow model that plays the most critical role in deciding the output of the vocal folds and as a consequence the output of the overall system.

In terms of the glottal flow (U_g), there is no standard method in literature to compute the quantity for 2D and 3D vocal fold models. In general, the glottal flow is understood to be the product of the glottal velocity and the glottal area; usually the minimum glottal area or glottal area at the point of flow separation are considered for the latter. In our simulation, we multiply the minimum 2D glottal cross-sectional area with the flow velocity at that point; this gives a clean glottal waveform without too many random variations. However, it would be interesting to see if other formulations of glottal flow give radically different glottal waveforms and, as a consequence, different inputs to the vocal tract simulation.

A general comment is the lack of watertight validation of our model with respect to the field; this arguably remains the biggest drawback of this study. This however remains a major issue for the entire field at large; data on the vocal fold is sparse at best, and misleading at worst. Due to the inaccessibility of the vocal folds, speech scientists have resorted to studies on excised larynges and fabricated models apart from computational simulations. Thus, we are forced to compare our models mainly to existing graphs or to static laryngeal pressure distributions that do not have a strong resemblance to a dynamic simulation. Therefore, the validation of the model is a patchwork of observations that are combined together to understand the model performance. The pressing need in the field is an open computational data-set that presents a canonical validation to compare models to. At this point in time, the closest comparison remains the model of Alipour et al, recently validated in 2015 [9]. Thus we choose to compare our pressure and velocity distributions to that model, attempting to correlate our system's predictions to those seen in the overall model. It is pertinent to remember that we will have qualitatively similar and not quanitatively similar results to this model; a combination of the model's slightly different geometry, and its significantly different computational set-up (quasi-3D structural model with 2D Navier-Stokes flow model), means that such a comparison would be disingenuous. However, we show that we can achieve similar results, at a significantly lower computational cost, enabling FEM models to be seen as a feasible tool for articulatory synthesis applications.

Finally, one of the major achievements of the model is the ability to run coupled simulations. It is important to keep in mind that the manner of this coupling is open to debate. In an 1D flow model for the trachea/vocal tract, there is only a single value through which coupled is achieved making it a natural coupling with our 1D vocal fold fluid model. However, when

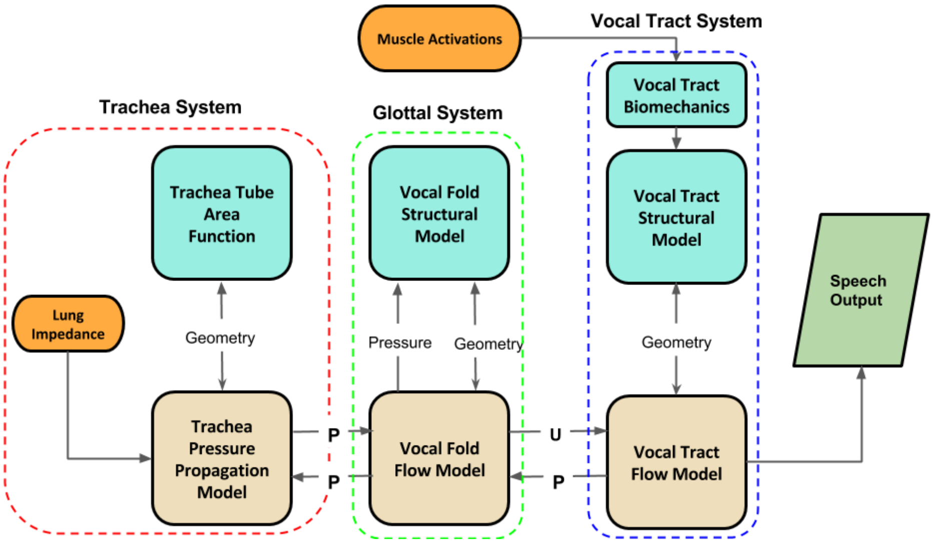


Figure 5.1: Components of an Articulatory Synthesizer

using the 2D vocal tract solver, the question arises about which cells would be appropriate for pressure feedback and how should the glottal output be fed into the vocal tract solver. This is an open question that now arises with the possibility of having higher order vocal-tract solvers coupled with vocal fold models, an issue which previously did not exist in the field. This is a consequence of pushing the envelope of the state-of-the-art in the field.

5.2 Future Work

Potential improvements and future work have been noted at the end of each chapter; here, a few directions from the perspective of articulatory synthesis at large are highlighted. We refer back to figure 5.1 to explore future work, component by component.

Trachea System

The exploration of the role of the trachea in speech synthesis remains quite nascent. Currently, the accepted wisdom in the field is that the role of trachea is minimal apart from reflecting the time-varying lung pressure at the subglottal duct. This is a reasonable assumption to make considering that many models have altogether done away with the trachea and still managed to produce reasonable vocal fold vibrations and vocal tract propagation values. The lungs, as they expand or contract, behave like bellows suggesting velocity or volumetric flow-rate BCs. However, since the lungs are also limited in strength from person to person, they behave like a pressure-driven system in the limiting simulation case. Potentially, our 1D fluid model could also be applied to seamlessly model the lungs as a velocity or pressure BCs, as well as calculating the flow through the trachea. While this might not have any perceptible difference in modal phonation, as we aim for better voice qualities, it might help in augmenting the subglottal signal with gender-based (different lung capacities) and register-based (different fundamental frequence of subglottal signal) information.

Glottal System

With regards to the glottal system, we shall divide potential future work into two parts: improvements to our model and general improvements for the field.

One possibility to improve the overall performance of the system is to go for a fully-coupled, implicit FSI solver to enhance simulation accuracy and stability. Overall, there remains the need to validate the system further; potentially we can look at a sensitivity analysis of our model to understand its performance over a range of values. Alternatively, we can look at using a 2D Navier-Stokes model coupled with a 2D structural model and comparing its performance to our system. This will isolate the performance of our 1D fluid model and help us better understand its shortcomings. One of the improvements mentioned previously was the inclusion of a τ_{small} term that can ensure a more realistic collision time for the model. Equally, it would be useful to understand the role that coupling plays in the overall performance of the system; this can potentially be understood by running thorough comparisons of coupled and non-coupled vocal fold models keeping all other parts of the system identical. Finally, the focus of this model was articulatory speech synthesis. To make this model usable for that purpose, it needs to be simulated in quasi real-time simulation rates. One potential solution is to follow the path suggested by Zappi et al [132] for the vocal tract; by parallelizing and optimizing the solver on the GPU we can ensure a complete acoustic simulation at extremely fast simulation rates. This would be a massive step-ahead for the entire field at large.

In general, one of the main issues holding back vocal fold models is the lack of validation data in literature. One of the major achievements for the field would be the generation of an open-source data store that provides comparison data for models to be validated against. Equally, there is a need to have models that are easily accessible by speech researchers; most continuum models are implemented in commercial or proprietary systems making it practically impossible for speech researchers to reproduce these systems. This has severely hampered progress in the field, to the extent that only a handful of research groups are in any position to make contributions in this area. We hope that by making our model available freely, we can help create systems to effectively compare vocal fold models.

Vocal Tract Systems

There are constant improvements to the state-of-the-art in vocal tract modelling. They can be mainly divided into two parts: 1) higher complexity models and 2) faster models. In terms of the latter, the use of real-time 2D models can enable speech synthesis at usable simulation times, and with the rise of high-quality GPU's, even move towards interactive simulations. A potential extension of our work would be to generate speech using biomechanically-driven vocal tract shapes. This would require an improvement in methods of extracting the centreline and 2D contours from 3D meshes, as well as stable dynamic vocal tract simulations.

5.3 Concluding Remarks

To conclude, this thesis has presented a new vocal fold model for articulatory speech synthesis. The model is significantly faster than existing continuum models while predicting similar glottal waveforms as higher order vocal fold models. The model was used as part of a complete articulatory speech synthesis simulation, where it produced physically realistic values. This work provides a starting point for developing an accurate, efficient and interactive articulatory synthesis toolkit which will eventually enable building better speech models and potentially lead to a clearer understanding of speech itself.

Bibliography

[1] Seiji Adachi and Jason Yu. Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing. *The Journal of the Acoustical Society of America*, 117(5):3213–3224, 2005.

[2] F Alipour, RC Scherer, and VC Patel. An experimental study of pulsatile flow in canine larynges. *Journal of fluids engineering*, 117(4):577–581, 1995.

[3] Fariborz Alipour, David A Berry, and Ingo R Titze. A finite-element model of vocal-fold vibration. *The Journal of the Acoustical Society of America*, 108(6):3003–3012, 2000.

[4] Fariborz Alipour, Christoph Brucker, Douglas D Cook, Andreas Gommel, Manfred Kaltenbacher, Willy Mattheus, Luc Mongeau, Eric Nauman, Rudiger Schwarze, Isao Tokuda, et al. Mathematical models and numerical schemes for the simulation of human phonation. *Current Bioinformatics*, 6(3):323–343, 2011.

[5] Fariborz Alipour and Ronald C Scherer. Pulsatile airflow during phonation: an excised larynx model. *The Journal of the Acoustical Society of America*, 97(2):1241–1248, 1995.

[6] Fariborz Alipour and Ronald C Scherer. Dynamic glottal pressures in an excised hemilarynx model. *Journal of Voice*, 14(4):443–454, 2000.

[7] Fariborz Alipour and Ronald C Scherer. Vocal fold bulging effects on phonation using a biophysical computer model. *Journal of Voice*, 14(4):470–483, 2000.

[8] Fariborz Alipour and Ronald C Scherer. Flow separation in a computational oscillating vocal fold model. *The Journal of the Acoustical Society of America*, 116(3):1710–1719, 2004.

[9] Fariborz Alipour and Ronald C Scherer. Time-dependent pressure and flow behavior of a self-oscillating laryngeal model with ventricular folds. *Journal of Voice*, 29(6):649–659, 2015.

[10] Fariborz Alipour and I Titze. Combined simulation of two dimensional airflow and vocal fold vibration. *Status and Progress Report, National Center for Voice and Speech*, 8:9–14, 1995.

[11] Donald R Allen and William J Strong. A model for the synthesis of natural sounding vowels. *The Journal of the Acoustical Society of America*, 78(1):58–69, 1985.

[12] Peter Anderson, Sidney Fels, and Sheldon Green. Implementation and validation of a 1d fluid model for collapsible channels. *Journal of biomechanical engineering*, 135(11):111006, 2013.

[13] Peter Anderson, Negar M Harandi, Scott R Moisik, Ian Stavness, and Sidney Fels. A comprehensive 3d biomechanically-driven vocal tract model including inverse dynamics for speech research. In *Interspeech 2015: 16th Annual Conference of the International Speech Communication Association*, pages 2395–2399, 2015.

[14] Peter J. Anderson. *Modeling the fluid-structure interaction of the upper airway: towards simulation of obstructive sleep apnea*. PhD thesis, University of British Columbia, 2014.

[15] Marc Arnela and Oriol Guasch. Two-dimensional vocal tracts with three-dimensional behavior in the numerical generation of vowels. *The Journal of the Acoustical Society of America*, 135(1):369–379, 2014.

[16] Pierre Badin, Gerard Bailly, Lionel Reveret, Monica Baciu, Christoph Segebarth, and Christophe Savariaux. Three-dimensional linear articulatory modeling of tongue, lips and face, based on mri and video images. *Journal of Phonetics*, 30(3):533–553, 2002.

[17] Lucie Bailly, Xavier Pelorson, Nathalie Henrich, and Nicolas Ruty. Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation. *The Journal of the Acoustical Society of America*, 124(5):3296–3308, 2008.

[18] Stefan Becker, Stefan Kniesburges, Stefan Müller, Antonio Delgado, Gerhard Link, Manfred Kaltenbacher, and Michael Döllinger. Flow-structure-acoustic interaction in a human voice model. *The Journal of the Acoustical Society of America*, 125(3):1351–1361, 2009.

[19] David A Berry, Hanspeter Herzel, Ingo R Titze, and Katharina Krischer. Interpretation of biomechanical simulations of normal and

chaotic vocal fold oscillations with empirical eigenfunctions. *The Journal of the Acoustical Society of America*, 95(6):3595–3604, 1994.

[20] David A Berry and Ingo R Titze. Normal modes in a continuum model of vocal fold tissues. *The Journal of the Acoustical Society of America*, 100(5):3345–3354, 1996.

[21] CD Bertram and TJ Pedley. A mathematical model of unsteady collapsible tube behaviour. *Journal of Biomechanics*, 15(1):39–50, 1982.

[22] Christopher D Bertram. Flow-induced oscillation of collapsed tubes and airway structures. *Respiratory physiology & neurobiology*, 163(1):256–265, 2008.

[23] Jonas Beskow. *Talking heads-Models and applications for multimodal speech synthesis*. PhD thesis, Institutionen för talöverföring och musikakustik, 2003.

[24] Peter Birkholz, Dietmar Jackèl, and Bernd J Kroger. Simulation of losses due to turbulence in the time-varying vocal system. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4):1218–1226, 2007.

[25] Peter Birkholz, Bernd J Kröger, and Christiane Neuschaefer-Rube. Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis.

[26] Peter Birkholz, BJ Kröger, and P Birkholz. A survey of self-oscillating lumped-element models of the vocal folds. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung*, pages 47–58, 2011.

[27] Stéphanie Buchaillard, Pascal Perrier, and Yohan Payan. A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *The Journal of the Acoustical Society of America*, 126(4):2033–2051, 2009.

[28] Claudio Cancelli and TJ Pedley. A separated-flow model for collapsible-tube oscillations. *Journal of Fluid Mechanics*, 157:375–404, 1985.

[29] Roger W Chan and Ingo R Titze. Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics. *The Journal of the Acoustical Society of America*, 119(4):2351–2362, 2006.

[30] Roger W Chan, Ingo R Titze, and Michael R Titze. Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa. *The Journal of the Acoustical Society of America*, 101(6):3722–3727, 1997.

[31] Siyuan Chang, Fang-Bao Tian, Haoxiang Luo, James F Doyle, and Bernard Rousseau. The role of finite displacements in vocal fold modeling. *Journal of biomechanical engineering*, 135(11):111008, 2013.

[32] Marc Arnela Coll. Numerical production of vowels and diphthongs using finite element methods. 2015.

[33] Douglas Cook, Pradeep George, and Margaret Julias. 2d/3d hybrid structural model of vocal folds. *Journal of biomechanics*, 45(2):269–274, 2012.

[34] Bert Cranen and Louis Boves. On the measurement of glottal flow. *The Journal of the Acoustical Society of America*, 84(3):888–900, 1988.

[35] Saeed Dabbaghchian, Marc Arnela, Olov Engwall, Oriol Guasch, Ian Stavness, and Pierre Badin. Using a biomechanical model and articulatory data for the numerical production of vowels. In *Interspeech, 8-12 Sep 2016, San Francisco*, pages 3569–3573, 2016.

[36] Jianwu Dang and Kiyoshi Honda. Construction and control of a physiological articulatory model. *The Journal of the Acoustical Society of America*, 115(2):853–870, 2004.

[37] Marcelo de Oliveira Rosa, José Carlos Pereira, Marcos Grellet, and Abeer Alwan. A contribution to simulating a three-dimensional larynx model using the finite element method. *The Journal of the Acoustical Society of America*, 114(5):2893–2905, 2003.

[38] MP De Vries, HK Schutte, and GJ Verkerke. Determination of parameters for lumped parameter models of the vocal folds using a finite-element method approach. *The Journal of the Acoustical Society of America*, 106(6):3620–3628, 1999.

[39] Gifford Z Decker and Scott L Thomson. Computational simulations of vocal fold vibration: Bernoulli versus navier–stokes. *Journal of Voice*, 21(3):273–284, 2007.

[40] Mickael Deverge, Xavier Pelorson, Coriandre Vilain, P-Y Lagrée, F Chentouf, J Willems, and A Hirschberg. Influence of collision on the flow through in-vitro rigid models of the vocal folds. *The Journal of the Acoustical Society of America*, 114(6):3354–3362, 2003.

[41] Comer Duncan, Guangnian Zhai, and Ronald Scherer. Modeling coupled aerodynamics and vocal fold dynamics using immersed boundary methods. *The Journal of the Acoustical Society of America*, 120(5):2859–2871, 2006.

[42] Gunnar Fant. *Acoustic Theory of Speech Production.* Mouton, The Hague, 1960.

[43] Gunnar Fant, Johan Liljencrants, and Qi-guang Lin. A four-parameter model of glottal flow. 1985.

[44] Mehrdad H Farahani and Zhaoyan Zhang. Experimental validation of a three-dimensional reduced-order continuum model of phonation. *The Journal of the Acoustical Society of America*, 140(2):EL172–EL177, 2016.

[45] J Flanagan and Lois Landgraf. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, 16(1):57–64, 1968.

[46] JL Flanagan and K Ishizaka. Computer model to characterize the air volume displaced by the vibrating vocal cords. *The Journal of the Acoustical Society of America*, 63(5):1559–1565, 1978.

[47] Adrian J Fourcin and Evelyn Abberton. First applications of a new laryngograph. *Medical & biological illustration*, 21(3):172–182, 1971.

[48] Lewis P Fulcher, Ronald C Scherer, Kenneth J De Witt, Pushkal Thapa, Yang Bo, and Bogdan R Kucinschi. Pressure distributions in a static physical model of the hemilarynx: measurements and computations. *Journal of Voice*, 24(1):2–20, 2010.

[49] Lewis P Fulcher, Ronald C Scherer, Guangnian Zhai, and Zipeng Zhu. Analytic representation of volume flow as a function of geometry and pressure in a static physical model of the glottis. *Journal of Voice*, 20(4):489–512, 2006.

[50] Jean-Michel Gérard, Pascal Perrier, and Yohan Payan. 3d biomechanical tongue modeling to study speech production, 2006.

[51] Bryan Gick, Ian Wilson, and Donald Derrick. *Articulatory phonetics*. John Wiley & Sons, 2012.

[52] Heather E Gunter. Modeling mechanical stresses as a factor in the etiology of benign vocal fold lesions. *Journal of biomechanics*, 37(7):1119–1124, 2004.

[53] Petr Hájek, Pavel Švancara, Jaromír Horáček, and Jan G Švec. Numerical simulation of the self-oscillating vocal folds in interaction with vocal tract shaped for particular czech vowels. In *Recent Global Research and Education: Technological Challenges*, pages 317–323. Springer, Cham, 2017.

[54] Negar M Harandi, J Woo, MR Farazi, L Stavness, M Stone, Sidney Fels, and Rafeef Abugharbieh. Subject-specific biomechanical modelling of the oropharynx with application to speech production. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 1389–1392. IEEE, 2015.

[55] Matthias Heil. Stokes flow in collapsible tubes: computation and experiment. *Journal of Fluid Mechanics*, 353:285–312, 1997.

[56] Matthias Heil, Andrew L Hazel, and Jaclyn A Smith. The mechanics of airway closure. *Respiratory physiology & neurobiology*, 163(1):214–221, 2008.

[57] John Henry Heinbockel. *Introduction to tensor calculus and continuum mechanics*, volume 52.

[58] Minoru Hirano. Morphological structure of the vocal cord as a vibrator and its variations. *Folia Phoniatrica et Logopaedica*, 26(2):89–94, 1974.

[59] Minoru Hirano and Yuki Kakita. Cover-body theory of vocal fold vibration. *Speech science*, pages 1–46, 1985.

[60] Julio C Ho, Matías Zañartu, and George R Wodicka. An anatomically based, time-domain acoustic model of the subglottal system for speech production. *The Journal of the Acoustical Society of America*, 129(3):1531–1547, 2011.

[61] Eva B Holmberg, Robert E Hillman, and Joseph S Perkell. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America*, 84(2):511–529, 1988.

[62] J Horáček, P Šidlof, and JG Švec. Numerical simulation of self-oscillations of human vocal folds with hertz model of impact forces. *Journal of fluids and structures*, 20(6):853–869, 2005.

[63] K Ishizaka and JL Flanagan. Acoustic properties of longitudinal displacement in vocal cord vibration. *Bell System Technical Journal*, 56(6):889–918, 1977.

[64] Kenzo Ishizaka and James L Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell system technical journal*, 51(6):1233–1268, 1972.

[65] OE Jensen. Instabilities of flow in a collapsed tube. *Journal of Fluid Mechanics*, 220:623–659, 1990.

[66] Oliver E Jensen and Matthias Heil. High-frequency self-excited oscillations in a collapsible-channel flow. *Journal of Fluid Mechanics*, 481:235–268, 2003.

[67] Jack Jiang, Emily Lin, and David G Hanson. Vocal fold physiology. *Otolaryngologic Clinics of North America*, 33(4):699–718, 2000.

[68] Weili Jiang, Xudong Zheng, and Qian Xue. computational modeling of fluid–structure–acoustics interaction during voice production. *Frontiers in bioengineering and biotechnology*, 5, 2017.

[69] John L Kelly and Carol C Lochbaum. Speech synthesis. 1962.

[70] Sid Khosla, Shanmugam Murugappan, Raghavaraju Lakhamraju, and Ephraim Gutmark. Using particle imaging velocimetry to measure anterior-posterior velocity gradients in the excised canine larynx model. *The Annals of otology, rhinology, and laryngology*, 117(2):134, 2008.

[71] Sid Khosla, Shanmugam Muruguppan, Ephraim Gutmark, and Ronald Scherer. Vortical flow field during phonation in an excised canine larynx model. *Annals of Otology, Rhinology & Laryngology*, 116(3):217–228, 2007.

[72] M Drew LaMar, Yingyong Qi, and Jack Xin. Modeling vocal fold motion with a hydrodynamic semicontinuum model. *The Journal of the Acoustical Society of America*, 114(1):455–464, 2003.

[73] FLE Lecluse, MP Brocaar, and J Verschuure. The electroglottography and its relation to glottal activity. *Folia Phoniatrica et Logopaedica*, 27(3):215–224, 1975.

[74] Johan Liljencrants. A translating and rotating mass model of the vocal folds. 1991.

[75] Gerhard Link, M Kaltenbacher, Michael Breuer, and M Döllinger. A 2d finite-element scheme for fluid–solid–acoustic interactions and its application to human phonation. *Computer Methods in Applied Mechanics and Engineering*, 198(41):3321–3334, 2009.

[76] HF Liu, XY Luo, ZX Cai, and TJ Pedley. Sensitivity of unsteady collapsible channel flows to modelling assumptions. *International Journal for Numerical Methods in Biomedical Engineering*, 25(5):483–504, 2009.

[77] John E Lloyd, Ian Stavness, and Sidney Fels. Artisynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. In Yohan Payan, editor, *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, pages 355–394. Springer, New York, 2012.

[78] Jorge C Lucero and Kevin G Munhall. A model of facial biomechanics for speech production. *The Journal of the Acoustical Society of America*, 106(5):2834–2842, 1999.

[79] Haoxiang Luo, Rajat Mittal, Xudong Zheng, Steven A Bielamowicz, Raymond J Walsh, and James K Hahn. An immersed-boundary method for flow–structure interaction in biological systems with application to phonation. *Journal of computational physics*, 227(22):9303–9332, 2008.

[80] XY Luo and TJ Pedley. The effects of wall inertia on flow in a two-dimensional collapsible channel. *Journal of Fluid Mechanics*, 363:253–280, 1998.

[81] Shinji Maeda. A digital simulation method of the vocal-tract system. *Speech communication*, 1(3-4):199–229, 1982.

[82] A Marzo, XY Luo, and CD Bertram. Three-dimensional collapse and steady flow in thick-walled flexible tubes. *Journal of Fluids and Structures*, 20(6):817–835, 2005.

[83] The Mathworks, Inc., Natick, Massachusetts. *MATLAB version 8.5.0.197613 (R2015a)*, 2015.

[84] Paul Mermelstein. Articulatory model for the study of speech production. *The Journal of the Acoustical Society of America*, 53(4):1070–1082, 1973.

[85] Peter Meyer, Reiner Wilhelms, and Hans Werner Strube. A quasiarticulatory speech synthesizer for german language running in real time. *The Journal of the Acoustical Society of America*, 86(2):523–539, 1989.

[86] Amir K Miri. Mechanical characterization of vocal fold tissue: a review study. *Journal of Voice*, 28(6):657–667, 2014.

[87] Rajat Mittal, Byron D Erath, and Michael W Plesniak. Fluid dynamics of human phonation and speech. *Annual Review of Fluid Mechanics*, 45:437–467, 2013.

[88] Parham Mokhtari, Hironori Takemoto, and Tatsuya Kitamura. Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches. *Speech Communication*, 50(3):179–190, 2008.

[89] Randall B Monsen and A Maynard Engebretson. Study of variations in the male and female glottal wave. *The Journal of the Acoustical Society of America*, 62(4):981–993, 1977.

[90] Pertti Palo. *A review of articulatory speech synthesis*. PhD thesis, Citeseer, 2006.

[91] Xavier Pelorson, Avraham Hirschberg, RR Van Hassel, APJ Wijnands, and Yves Auregan. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model. *The Journal of the Acoustical Society of America*, 96(6):3416–3431, 1994.

[92] Pascal Perrier, Yohan Payan, Stéphanie Buchaillard, Mohammad Ali Nazari, and Matthieu Chabanas. Biomechanical models to study speech. *Faits de langues*, 37:155–171, 2011.

[93] Per-Olof Persson. Implementation of finite element-based navier-stokes solver 2.094-project. 2002.

[94] Petra Pořízková, Karel Kozel, and Jaromír Horáček. Numerical solution of compressible and incompressible unsteady flows in channel

inspired by vocal tract. *Journal of Computational and Applied Mathematics*, 270:323–329, 2014.

[95] Pedro Paulo Leite Do Prado. a target-based articulatory synthesizer. 1991.

[96] Martin Rothenberg. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *The Journal of the Acoustical Society of America*, 53(6):1632–1645, 1973.

[97] Ronald C Scherer, Daoud Shinwari, Kenneth J De Witt, Chao Zhang, Bogdan R Kucinschi, and Abdollah A Afjeh. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *The Journal of the Acoustical Society of America*, 109(4):1616–1630, 2001.

[98] Raphael Schwarz, Ulrich Hoppe, Maria Schuster, Tobias Wurzbacher, Ulrich Eysholdt, and Jörg Lohscheller. Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model. *IEEE transactions on biomedical engineering*, 53(6):1099–1108, 2006.

[99] Benjamin Seibold. A compact and fast matlab code solving the incompressible navier-stokes equations on rectangular domains mit18086 navierstokes. m. 2008.

[100] Eftychios Sifakis, Andrew Selle, Avram Robinson-Mosher, and Ronald Fedkiw. Simulating speech with a physics-based facial muscle model. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 261–270. Eurographics Association, 2006.

[101] Man Sondhi and Juergen Schroeter. A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(7):955–967, 1987.

[102] Susan Standring. *Gray's anatomy: the anatomical basis of clinical practice*. Elsevier Health Sciences, 2015.

[103] Ina Steinecke and Hanspeter Herzel. Bifurcations in an asymmetric vocal-fold model. *The Journal of the Acoustical Society of America*, 97(3):1874–1884, 1995.

[104] Kenneth N Stevens. *Acoustic phonetics*, volume 30. 2000.

[105] Brad H Story. An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology*, 23(4):195–206, 2002.

[106] Brad H Story. Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *The Journal of the Acoustical Society of America*, 123(1):327–335, 2008.

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

[108] Brad H Story, Ingo R Titze, and Eric A Hoffman. Vocal tract area functions from magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 100(1):537–554, 1996.

[109] Sayoko Takano, Kiyoshi Honda, and Keisuke Kinoshita. Measurement of cricothyroid articulation using high-resolution mri and 3d pattern matching. *Acta acustica united with acustica*, 92(5):725–730, 2006.

[110] Hironori Takemoto, Kiyoshi Honda, Shinobu Masaki, Yasuhiro Shimada, and Ichiro Fujimoto. Measurement of temporal changes in vocal tract area function from 3d cine-mri data. *The Journal of the Acoustical Society of America*, 119(2):1037–1049, 2006.

[111] Hironori Takemoto, Parham Mokhtari, and Tatsuya Kitamura. Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method. *The Journal of the Acoustical Society of America*, 128(6):3724–3738, 2010.

[112] Keyi Tang. *A feasibility study of template-based subject-specific modelling and simulation of upper-airway complex.* PhD thesis, University of British Columbia, 2017.

[113] Chao Tao, Jack J Jiang, and Yu Zhang. Simulation of vocal fold impact pressures with a self-oscillating finite-element model. *The Journal of the Acoustical Society of America*, 119(6):3987–3994, 2006.

[114] António JS Teixeira, Roberto Martinez, Luís Nuno Silva, Luis MT Jesus, Jose C Príncipe, and Francisco AC Vaz. Simulation of human

speech production applied to the study and synthesis of european portuguese. *EURASIP Journal on Applied Signal Processing*, 2005:1435–1448, 2005.

[115] Scott L Thomson, Luc Mongeau, and Steven H Frankel. Aerodynamic transfer of energy to the vocal folds. *The Journal of the Acoustical Society of America*, 118(3):1689–1700, 2005.

[116] Ingo R Titze. Mechanical stress in phonation. *Journal of Voice*, 8(2):99–105, 1994.

[117] Ingo R Titze. Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model. *The Journal of the Acoustical Society of America*, 111(1):367–376, 2002.

[118] Ingo R Titze. *The Myoelastic Aerodynamic Theory of Phonation.* National Center for Voice and Speech, 2006.

[119] Ingo R Titze. Nonlinear source–filter coupling in phonation: Theory a. *The Journal of the Acoustical Society of America*, 123(4):1902–1915, 2008.

[120] Ingo R Titze, Sheila S Schmidt, and Michael R Titze. Phonation threshold pressure in a physical model of the vocal fold mucosa. *The Journal of the Acoustical Society of America*, 97(5):3080–3084, 1995.

[121] Ingo R Titze and Brad H Story. Rules for controlling low-dimensional vocal fold models with muscle activation. *The Journal of the Acoustical Society of America*, 112(3):1064–1076, 2002.

[122] Isao T Tokuda, Jaromir Horáček, Jan G Švec, and Hanspeter Herzel. Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments. *The Journal of the Acoustical Society of America*, 122(1):519–531, 2007.

[123] Isao T Tokuda, Marco Zemke, Malte Kob, and Hanspeter Herzel. Biomechanical modeling of register transitions and the role of vocal tract resonators a. *The Journal of the Acoustical Society of America*, 127(3):1528–1536, 2010.

[124] M Triep, Ch Brücker, and W Schröder. High-speed piv measurements of the flow downstream of a dynamic mechanical model of the human vocal folds. *Experiments in Fluids*, 39(2):232–245, 2005.

[125] Janwillem Van den Berg. Myoelastic-aerodynamic theory of voice production. *Journal of Speech, Language, and Hearing Research*, 1(3):227–244, 1958.

[126] Kees van den Doel and Uri M Ascher. Real-time numerical solution of webster's equation on a nonuniform grid. *IEEE transactions on audio, speech, and language processing*, 16(6):1163–1172, 2008.

[127] Reiner Wilhelms-Tricarico. Physiological modeling of speech production: Methods for modeling soft-tissue articulators. *The Journal of the Acoustical Society of America*, 97(5):3085–3098, 1995.

[128] Q Xue, R Mittal, X Zheng, and S Bielamowicz. Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx. *The Journal of the Acoustical Society of America*, 132(3):1602–1613, 2012.

[129] Qian Xue and Xudong Zheng. The effect of false vocal folds on laryngeal flow resistance in a tubular three-dimensional computational laryngeal model. *Journal of Voice*, 2016.

[130] Anxiong Yang, Jörg Lohscheller, David A Berry, Stefan Becker, Ulrich Eysholdt, Daniel Voigt, and Michael Döllinger. Biomechanical modeling of the three-dimensional aspects of human vocal fold dynamics. *The Journal of the Acoustical Society of America*, 127(2):1014–1031, 2010.

[131] Matías Zañartu, Luc Mongeau, and George R Wodicka. Influence of acoustic loading on an effective single mass model of the vocal folds. *The Journal of the Acoustical Society of America*, 121(2):1119–1129, 2007.

[132] Victor Zappi, Arvind Vasudevan, Andrew Allen, Nikunj Raghuvanshi, and Sidney Fels. Towards real-time two-dimensional wave propagation for articulatory speech synthesis. In *Proceedings of Meetings on Acoustics 171ASA*, volume 26. ASA, 2016.

[133] SM Zeitels. Phonosurgery: Past present and future. *OPERATIVE TECHNIQUES IN OTOLARYNGOLOGY HEAD AND NECK SURGERY*, 10:1–1, 1999.

[134] Tao Zhang, Amgad Salama, Shuyu Sun, and Hua Zhong. A compact numerical implementation for solving stokes equations using matrix-vector operations. *Procedia Computer Science*, 51:1208–1218, 2015.

[135] X Zheng, R Mittal, Q Xue, and S Bielamowicz. Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model. *The Journal of the Acoustical Society of America*, 130(1):404–415, 2011.

[136] Xudong Zheng, Steve Bielamowicz, Haoxiang Luo, and Rajat Mittal. A computational study of the effect of false vocal folds on glottal flow and vocal fold vibration during phonation. *Annals of biomedical engineering*, 37(3):625–642, 2009.