Evaluation of High-Speed Videoendoscopy for Bayesian Inference on Reduced Order Vocal Fold Models

by

Jonathan Deng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The ability to use our voice occurs through a complex bio-mechanical process known as phonation. The study of this process is interesting, not only because of the complex physical phenomena involved, but also because of the presence of phonation disorders that can make the everyday task of using ones voice difficult. Clinical studies of phonation aim to help diagnose such disorders using various measurement techniques, such as microphone recordings, video of the vocal folds, and perceptual sound quality measures. In contrast, scientific investigations of phonation have focused on understanding the physical phenomena behind phonation using simplified physical and numerical models constructed using representative population based parameters. A particularly useful type of model, reduced-order numerical models, are simplified representations of the vocal folds with low computational complexity that allow broad parameter changes to be investigated.

To bring the physical understanding of phonation from these models into clinical usage, it is necessary to have patient specific parameters. Due to the difficulty of measuring vocal fold parameters and other structures in phonation directly, inverse analysis techniques must be employed. These techniques estimate the parameters of a model, by finding model parameters that lead to outputs of the model which compare well with measured outputs. With the measured outputs being patient specific measurements, these techniques can produce patient specific model parameters. However, this is complicated by the fact that measurements are uncertain, which leads to uncertainty in inferred parameters. The uncertainty in the parameters provides a way to judge how confident clinicians should be in using them. Large measurements errors could result in high uncertainties (and vice versa), which should guide clinicians on whether or not to believe the estimated parameters. Bayesian inference is an inverse analysis technique, that can take into account the inherent uncertainty in measurements in a probabilistic framework. Applying Bayesian inference to reduced-order models and clinical measurements allows patient specific model parameters with associated uncertainties to be inferred.

A promising clinical measurement for use in Bayesian inference is high-speed videoendoscopy, in which high-speed video is taken of the vocal folds in motion. This captures the time varying motion of the vocal folds, which allows many quantitative measurements to be derived from the resulting video, for example the glottal width (distance between the vocal folds) or glottal area (area between the vocal folds). High-speed videoendoscopy is subject to variable imaging parameters, in particular the frame rate, spatial resolution, and tilted views of the camera can all modify the resulting video of the vocal folds, changing the uncertainty in the derived measurements. To investigate the effect of these three imaging parameters on Bayesian inference applied to high-speed video endoscopy, a simulated

high-speed videoendoscopy experiment was conducted.

Using a reduced order model, with known parameters, a set of enlarged, artificial vocal folds were driven in slow motion. These were imaged by a consumer DSLR camera, where the slow motion increased the effective frame rate, and the enlarged vocal folds increased the effective spatial resolution, to a fidelity much greater than typical high-speed videos of the vocal folds. This allowed investigation of the three parameters; titled views of the camera were investigated by physically tilting the camera, while variable frame rates and spatial resolutions were investigated by numerical downsampling of the original recording. Bayesian inference was conducted on these simulated high-speed videos, by measuring the distance between the vocal folds (the glottal width), in order to determine the parameters of the same reduced-order model driving the artificial vocal folds. This provided a reference to compare the estimated parameters with. The changes in estimated parameters from Bayesian inference were then investigated as the angle of view, frame rate, and spatial resolution were modified.

From the experiment, the effects of frame rate, spatial resolution, and angle of view in high-speed videoendoscopy were found relative to changes from a reference video. Specifically, uncertainty in estimates increased linearly with respect to downsampling factor of frame rate. A frame rate that is half that of the reference video will have an uncertainty on estimated parameters that is twice as large. Spatial resolution affects the level of uncertainty based on the edge detection techniques that are used to extract quantitative data (*i.e.*,, the glottal width in this study). As the spatial resolution was downsampled, the level of error from the edge detection algorithm increased linearly with respect to the downsampling factor, which subsequently led to the same linear increase in the level of uncertainty in the estimate. However, different edge detection algorithms will likely have different accuracies as the resolution of the image decreases. While in this study it is preferable to decrease spatial resolution instead of frame rate, more general conclusions would be dependent on the specific edge detection technique used.

The angle of view was found to bias estimates as a result of projecting the vocal folds (glottis) onto an offset image plane (like viewing a coin from an angle, results in increasingly narrow ellipses until a single line is formed, rather than a circle). This decreased the glottal width measured, which biased the estimated parameters. To account for this bias, it is suggested that the angle of view can be treated as an uncertain parameter, which leads to increased uncertainty in the quantitative measures from high-speed video. Alternatively, the angle of view can be estimated as an additional parameter.

Acknowledgements

To my advisor, Prof. Sean Peterson, thank you for your patience, pushing me to be better, and helpful discussions! Thank you to Dr. Paul Hadwin, who's guidance taught me a great deal on Bayesian inference. I would also like to thank the committee members, Prof. Kyle Daun, and Prof. Serhiy Yarusevych.

To my family, thanks for always supporting me! To my mom and dad, Lin Qiao and Shaotang Deng, for the unconditional love, and to my sisters Jane and Sarah for the lifelong sibling friendship.

I am grateful to all my colleagues, Amit Dutta, JiaCheng (Winston) Hu, John Kurelek, Erik Marble, Jeff McClure, Supun Pieris, Ben Pocock, Yash Shah, and Xueqing (Caddie) Zhang, for the fun times and interesting discussions.

I would also like to thank all the staff members at Waterloo I've had the pleasure of working with, and my friends who I have shared many great times with.

Dedication

This is dedicated to my family, who have always supported me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BCM body-cover model

CFD computational fluid dynamics

DOF degree of freedom

DSLR digital single lens reflex camera

FEM finite element methods

HSV high-speed videoendoscopy

MAP maximum aposteriori

MLE maximum likelihood estimate

VF vocal fold

Nomenclature

Q	Volume flow rate through the glottis
P_{sub}	Subglottal pressure
P_{sup}	Supraglottal pressure
P_{sep}	Intraglottal pressure at the separation location
A_{sub}	Transverse cross-sectional area of the glottis at the beginning of the subglottal tract
A_{sup}	Transverse cross-sectional area of the glottis at the beginning of the vocal tract
A_{sep}	Transverse cross-sectional area of the glottis at the separation location
a_{ct}	Cricothyroid activity
a_{ta}	Thyroarytenoid activity
a_{lc}	Lateral cricoarytenoid activity
x_{u}	Displacement of the upper mass (superior) in the body-cover model
x_{l}	Displacement of the lower mass (inferior) in the body-cover model
x_{b}	Displacement of the bottom mass in the body-cover model
v_{u}	Velocity of the upper mass (superior) in the body-cover model
v_{l}	Velocity of the lower mass (inferior) in the body-cover model
v_{b}	Velocity of the bottom mass in the body-cover model
m_{u}	Upper mass (superior) in the body-cover model

m_{l} Lower mass (inferior) in the body-cover model

m_{b} Bottom mass in the body-cover model

F_{u} Force due to pressure on the upper mass (superior) in the body-cover model

F_{l} Force due to pressure on the lower mass (inferior) in the body-cover model

d_{spatial} Spatial downsampling factor of a video

d_{temporal} Temporal downsampling factor of a video

α Angle of view of simulated high-speed video

Chapter 1

Introduction

The use of our voice is second nature to nearly everyone. Ask someone how their day was, spend a night out with friends at karaoke, or talk to your relatives over the phone; all these activities involve our voice. Our voice is generated through a process known as phonation, which occurs when the lungs contract, forcing air between a pair of specialized structures called the vocal folds (VFs) housed in the larynx, located in the throat. Normally the VFs are separated to enable breathing, but during phonation they are brought together and air is forced between them, exerting pressure on the VFs. This transfers energy from the air to the VFs, which when balanced with the elastic properties of the VFs, leads to self-oscillation (vibration). The oscillation of the VFs leads to a pulsatile flow behind the glottis, which creates a buzz like tone. The tone subsequently resonates through the airways, oral, and nasal cavities where the unique shape of these cavities modifies it into the complex sounds that are what other people ultimately hear.

The study of phonation has long been of interest, not only for the interesting question of how and why phonation works, but also due to the fact that phonation disorders can interrupt this delicate process. As a result, the study of phonation can be roughly split into two classes: clinical studies where medical professionals try to treat phonation related difficulties, and model based studies where engineers and scientists look into the mechanisms behind phonation. In the clinic, medical professionals (such as speech language pathologists or otolaryngologists) take various measurements of a patient's voice to treat and diagnose any discomforts (or disorders) the patient may have in speaking. Some of these measurements include the volume flow rate of air out of the nose and mouth, a recording of the patient's voice, or video of the VFs through an endoscope. Using these measurements as well as perceptual information, medical professionals try to help patients. In contrast to clinical studies, model based studies aim to study phonation from a physical point of

view using representative models based on population average estimates of physiological properties. A wide range of such studies have been conducted. For example, synthetic rubber VFs with a generic geometry have been used to study how the air interacts with the VFs to create vibration [69, 17, 52, 51, 49, 48]. Rigid VF models have been used to investigate the airflow between the VFs and how it is affected by the VF geometry [55, 80, 78, 35, 40, 24, 20]. And numerical studies have also been conducted using both high fidelity computational fluid dynamics (CFD) and finite element methods (FEM) [66, 96, 64], as well as simplified reduced order models (also called lumped mass models) [23, 41, 82, 63, 77, 87]. Reduced order models are a particularly attractive technique to investigate phonation due to their simplicity. A reduced order model consists of a collection of masses connected by springs and dampers that represent the VFs, which is coupled with simplified models of the air flow between the VFs, and sometimes additional acoustic models for the tracts above and below the glottis. The computational simplicity allows for broad studies over large parameter changes.

Model based studies of the VFs have greatly improved our understanding of the physics of phonation, however, since these studies are based on average population properties of the VFs, the resulting predictions are likewise restricted to be representative of the average population. However, phonation is a highly variable process between individuals, for example, despite the fact that we all use VFs to speak, everyone sounds different. While clinical studies obviously reflect patient specific differences, they overlook the physics of phonation. Both approaches are valuable. Model based studies give insight on why and how phonation occurs, and clinical studies measure what phonation results in on a patient specific basis.

To reconcile model based studies with clinical studies, and to make this powerful framework useful to clinicians, efforts have been made to produce patient specific models. Patient specific models require patient specific model parameters, however the viscoelastic properties of the VFs, and complex organic geometries inside the throat make patient measurements nearly impossible. Even if measurements could be made, the small size of the VFs and the invasive nature of measuring inside a patient's throat makes this approach unattractive. This has led to the use of indirect approaches through inverse analysis techniques. Inverse analysis techniques aim to estimate the parameters of a model that produce given measurements. Thus by applying inverse analysis techniques on patient specific measurements, intuitively one should obtain patient specific model parameters. A wealth of clinical, non-invasive measurements already exist. The volume flow rate of air at the nose and mouth, or high-speed recordings of the VFs are just a few [4]. These measurements can technically be applied to any VF model. For example, inverse analysis could be used to infer the material properties in a finite-element model of the VFs. However, inverse

techniques incur an additional layer of computational complexity on top of the the complexity of the VF model itself, thus simplified reduced order models are generally used [47, 16, 90, 59, 84, 7, 58, 83, 88].

1.1 Inverse Analysis for Patient Specific Modelling

Early approaches to inverse analysis used optimization techniques to estimate patient specific parameters [16, 58, 88, 87] on reduced order models. This approach generally produces a single set of parameters which is considered to be the 'best'. However, all measurements are subject to uncertainty, which can make the estimates of optimization approaches misleading; the fact that measurements are uncertain means that multiple combinations of model parameters could be responsible for a given measurement. In contrast, Bayesian inverse analysis techniques provide a natural setting to consider this uncertainty. Instead of a single parameter set, Bayesian techniques characterize the probability of a range of parameter sets. This approach has already been used to infer parameters from a patient's voice, as well as validated from a simulated VF motion [6, 28].

To gain clinical relevance, Bayesian techniques should be applicable to a wide range of clinical measures. A particularly important measurement is that of high-speed video of the VFs obtained through high-speed videoendoscopy (HSV). High-speed video is information rich, as it captures the time varying 2-D contours of the VFs. This allows the extraction of simple measurements, such as the area between the VFs (the glottal area), or more complex ones such as trajectories of the glottal contours of the VFs. The conversion of high-speed video into these quantitative measurements, however, is subject to many uncertainties. For example the video may not necessarily be centred on the axis of the VFs due to misplacement of the endoscope or motion during imaging [11]. Furthermore, for quantitative measures, pixel distances must be calibrated to physical distances, and edges of the VFs must be detected, which can be made uncertain by improperly focused videos. These are just a subset and many more factors can all contribute to uncertainty in measurements derived from high-speed video. To use high-speed video as a measurement for Bayesian inference techniques, these uncertainties must be characterized, which, to the author's knowledge has yet to be done in the context of VFs.

1.2 Objectives

HSV is subject to a plethora of factors that affect the quantitative measurements that are derived from it. To characterize all of these would be a gargantuan task. In this work we characterize what we believe to be three of the most important parameters of HSV on Bayesian inference applied to reduced order models, namely: the frame rate, resolution, and angle of view of the video. A reduction in frame rate directly reduces the number of measurements that can be used in Bayesian inference (where each frame is considered a measurement), while reduced resolution affects the accuracy of measuring the glottal contours and offset angles create parallax errors in converting pixel measurements to physical measurements. Furthermore, frame rate and resolution are two parameters that can be controlled in high-speed cameras, where higher resolutions can be achieved by reducing frame rate, and vice-versa. While these parameters are only a subset of the factors that affect HSV, they are nevertheless an important step in characterizing the feasibility of using HSV for Bayesian inference.

To characterize the effect of these three parameters on Bayesian inference we aim to answer the following questions:

1. How do offset-angled views change the estimated parameters?
2. Is there a trade-off between frame rate and resolution?
3. How can one expect uncertainty to vary under multiple sources of error?

1.3 Organization

The remainder of the thesis is organized into four chapters. In Chapter 2, a background on phonation is given, including the basic anatomy and a more detailed description of how phonation works. Then clinical measurements and the details of HSV are outlined. This is followed by details of the general equations behind reduced order models and the application of these models in the context of inverse analysis techniques. In Chapter 3, the experiment used in the study is described. Furthermore the methodology for conducting Bayesian inference is described. In Chapter 4, the results of the inference are shown and interpreted. In Chapter 5, conclusions and recommendations for future studies are discussed.

Chapter 2

Background

The study of phonation is an interdisciplinary field, drawing on a wide range of subjects from anatomy and biology to fluid mechanics and mathematics. In this section the pertinent background knowledge needed to follow the thesis is presented. Section 2.1 discusses the anatomy of phonation, including the structures involved, the structure of the VFs and common anatomical terminology. Section 2.2 details the mechanisms of phonation. In Section 2.3, a discussion of clinical approaches to the study of the phonation process is provided to familiarize the reader with the different types of clinical measurements. Next in Section 2.4, approaches to the numerical modelling of phonation are discussed. Finally, a discussion of inverse analysis in Section 2.5 provides the background knowledge necessary to understand the application of Bayesian inference to reduced order models of the VFs.

2.1 Anatomy

In this section the major anatomical structures, and specialized terminology are introduced; for more information the reader is referred to [60], on which much of this section is based.

2.1.1 Anatomical Directions, Planes, and Motions

The study of anatomy has a special set of terms for specification of locations, orientations, and directions. An overview of the anatomical directions and planes is shown in Figure 2.1. There are three major planes in the body [64]: the coronal plane splits the body into a front half and back half, the sagittal plane splits the body into left and right sides, and the

transverse plane splits the body into top and bottom halves. In addition to these three major planes of the body, three main directions are used to describe orientation. The anterior and posterior directions point towards the front and back of the body, respectively; the superior and inferior directions point towards the head and feet, respectively; and the medial and lateral directions point towards the middle and left/right sides of the body, respectively.

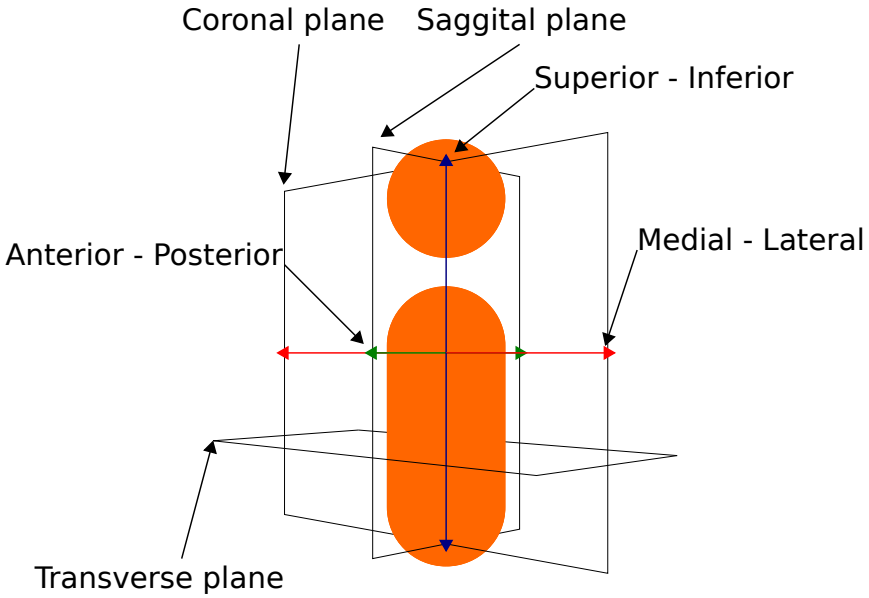


Figure 2.1: Three sets of anatomical directions create a coordinate system referenced to the body as well as three corresponding planes. These terms are frequently used to describe the positions and motions of anatomical structures.

These directional terms are often used to locate structures relative to some reference point. For example, saying that the VFs are inferior to the nasal and oral cavities, indicates that they are located below the nasal and oral cavities. Another example is in specifying a cross-sectional area. One may refer to the cross-sectional area of the vocal tract in the transverse plane.

In addition to directions and planes, there are also terms used to describe motions of the body. Of particular importance in phonation are adduction and abduction. Adduction and abduction refer to the bringing of structures of the body towards and away from the midline, respectively. This is often used to describe the motion of the VFs. For example prior to when phonation initiates, the VFs are adducted.

2.1.2 Anatomical Structures of Phonation

The general anatomical structures of phonation are shown in Figure 2.2 below. The larynx, which is the primary sound generating structure during many vocal gestures (for example, vowel sounds), is a collection of cartilages and muscles that house the VFs. Inferior to the larynx is the subglottal tract (also called the trachea), which is the airway to the lungs. Superior to the larynx is the supraglottal tract, which consists of the vocal tract, as well as the nasal and oral cavities.

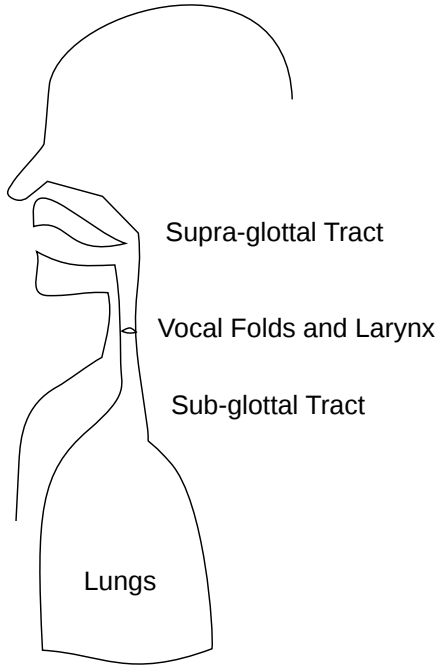


Figure 2.2: The key structures in the phonation process include: the lungs, the larynx, and VFs, as well as the subglottal and supraglottal tracts.

Larynx and Musculature

The larynx consists of 5 pieces of cartilage: the cricoid, epiglottis, thyroid, and the corniculate, and arytenoid cartilages, both of which are paired; these are shown in Figure 2.3.

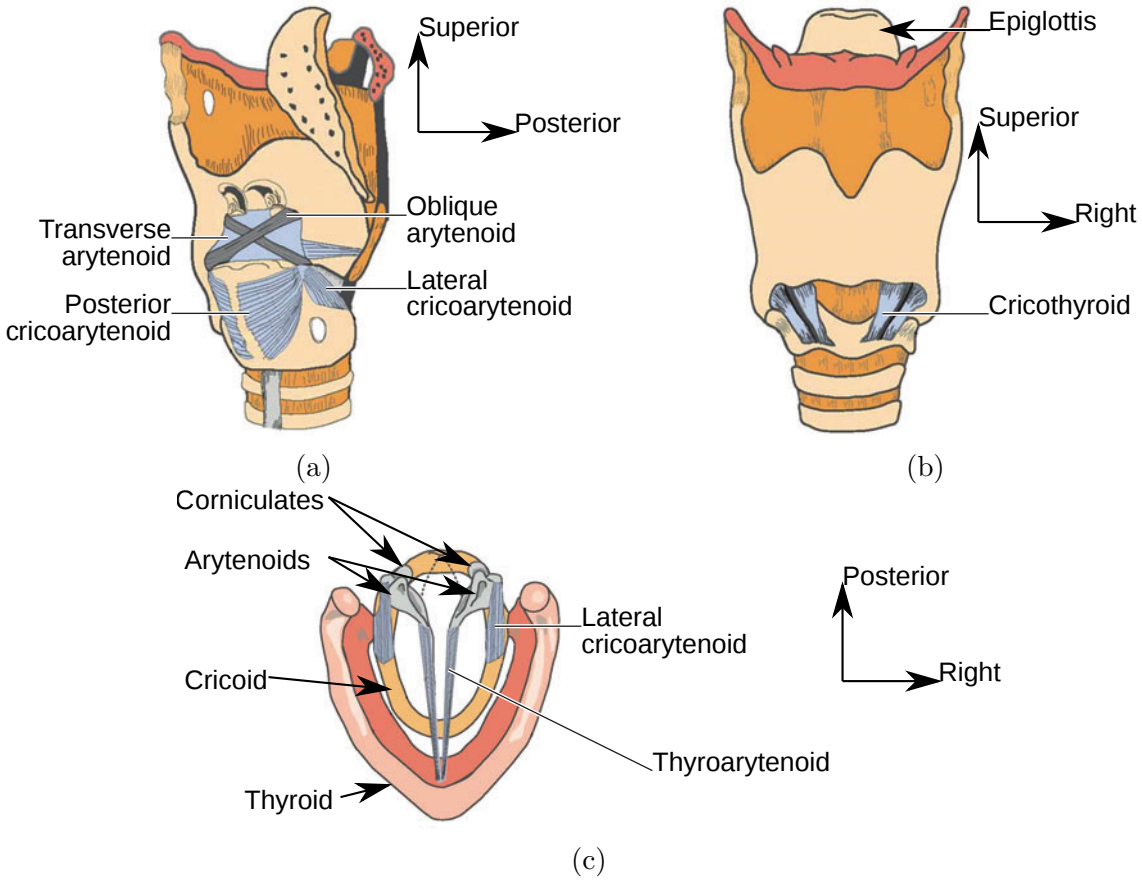


Figure 2.3: The larynx consists of 5 cartilages (signalled by arrows): cricoid, epiglottis, thyroid, corniculate and arytenoid (the last two are paired). These are articulated relative to each other through the action of intrinsic muscles (signalled by straight lines). (a) Coronal cross-sectional view of the larynx. (b) Anterior view of the larynx. (c) Superior view of the larynx. Adapted from [48].

These cartilages articulate through the action of various muscles. Muscles attached to the larynx can be classified into two types: intrinsic and extrinsic musculature. Extrinsic muscles connect the larynx to structures outside of the larynx, and so move the larynx relative to the body. Intrinsic muscles connect cartilages of the larynx to each other, and so are responsible for movement of cartilages within the larynx. Intrinsic muscles are particularly important as the action of many of these intrinsic muscles modify the geometry and material properties of the VFs, which directly influences phonation. These intrinsic muscles are named after the cartilages that they attach to. Of primary importance

are three intrinsic muscles: the cricothyroid (attached between the cricoid and thyroid cartilages), the lateral-cricoarytenoid, and the thyroarytenoid. In particular, the VFs are attached between the arytenoid and thyroid cartilages (the thyroarytenoid forms the innermost layer of the VFs, as seen in Figure 2.3(c)), so articulation of these cartilages modifies the VF geometry. When the arytenoid cartilages are articulated, the VFs are either abducted or adducted, that is, pressed together or pulled apart. When the thyroid is articulated, it is pulled away form the cricoid, and the VFs are stretched. The cricothyroid is primarily responsible for tensing (stretching) the VFs by articulating the thyroid. The lateral-cricoarytenoid is responsible for adduction (bringing the VFs together) by rotating the arytenoids. The thyroarytenoid is responsible for stiffening and relaxing the VFs, since it forms the body of the VFs [60, Chapter 5]. These three muscles have been used in rules relating muscle activation to material properties of the folds, which are used in this work (see Chapter 3 for details). These are only a subset of the muscles in the larynx, refer to [60, Chapter 5] for a comprehensive review.

Vocal Folds

The VFs are a set of paired whitish coloured layered structures. A sagittal cross-sectional image of a VF, and its internal layered structure is shown in Figure 2.4. The outermost layer of the VFs consists of the epithelium which serves to protect them and retain moisture [60, pg.168]. Underneath the epithelium is the lamina propria. This layer is further split into three sub-layers: the superficial, intermediate and deep layers [60, pg.168]. The body of the VF consists of the thyroarytenoid muscle. The combination of the epithelium and superficial layer of the lamina propria is known as the mucosa, or mucosal lining [60, pg.168]. The intermediate and deep layers of the lamina propria are known together as the vocal ligament [60, pg.168].

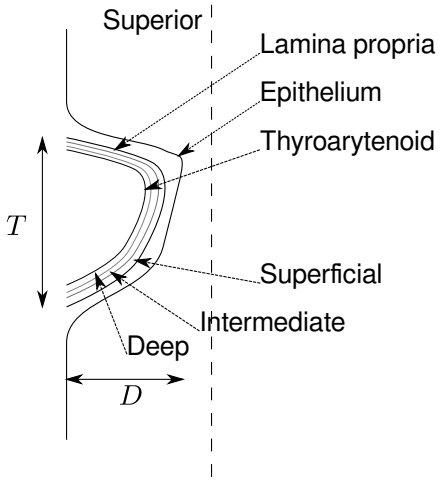


Figure 2.4: Coronal cross-sectional view of a VF showing its layered structure. At rest, the depth, D, of the VFs ranges from 0.6 cm for women to 0.8 cm for men [32]; the thickness of the VFs, T, ranges from 0.20 cm for women to 0.30 cm for men; the length of the VFs (in the anterior-posterior direction, or into the page in the figure) ranges from 1.0 cm for women to 1.6 cm for men [77].

The VFs have viscoelastic properties, meaning that they have both elastic properties and damping properties [25]. These viscoelastic material properties vary across each of the VF layers. While the inner thyroarytenoid muscle and deep layer of the lamina propria are relatively stiff, the intermediate and superficial layer of the lamina propria, along with the superficial mucosa are loose and elastic. For this reason, a functional classification into a stiff 'body' layer and loose 'cover' layer is often made [60, pg.169]. Under this classification, the body layer consists of the thyroarytenoid muscle and the deep layer of the lamina propria. The cover layer consists of all the remaining superficial tissues. Lastly, the area between the VFs is known as the glottis.

2.2 Mechanics of Phonation

The phonation process can roughly be broken down into two mechanisms: how interaction of the VFs with the glottal flow leads to self-oscillation, which produces a source tone, and how that source tone is changed by the supraglottal tract to produce the complex sounds that emanate out of our mouths. The widely accepted myo-elastic aerodynamic theory of phonation, formalized by Van den Berg [79], proposed that VF self-oscillation is due to

a balance of the elastic properties of the VFs and the aerodynamic pressures exerted by the glottal flow. Beginning from an adducted configuration of the VFs, pressure from the lungs increases, increasing the pressure in the glottis, and eventually exceeding the elastic forces of the VFs. This forces the VFs apart, opening the glottis. Initially the glottal entrance (the inferior end) is larger than the glottal exit (the superior end). This is due to the glottal pressure building from the lungs, and therefore acting on the entrance of the glottis first. This configuration, commonly referred to as a converging configuration, acts like a nozzle, causing higher pressures on the inferior surface of the VFs relative to the superior surface through the Bernoulli effect. This forces the inferior surfaces further apart, and helps to transfer energy from the glottal flow to the VFs.

As the VFs compress, elastic forces increase and eventually exceed the glottal pressures. This forces the VFs into a diverging configuration (when the glottal entrance is smaller than the glottal exit). Such a configuration leads to separation of the flow which reduces the glottal pressures. This hampers the transfer of energy from the glottal flow to the VFs, allowing the elastic forces within the VFs to return them to a closed position, thus starting the oscillation cycle over again. This alternating converging-diverging motion (also referred to as an inferior-superior phase shift) is called the mucosal wave and is illustrated in Figure 2.5.

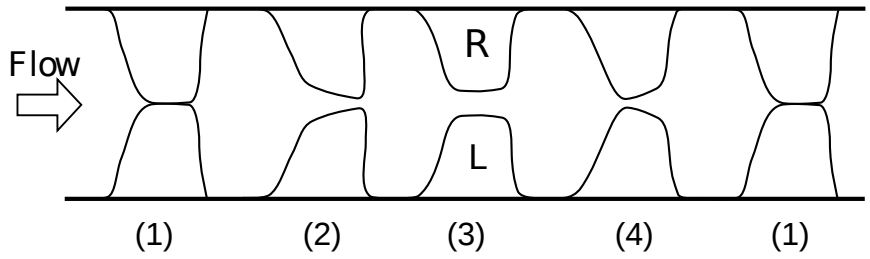


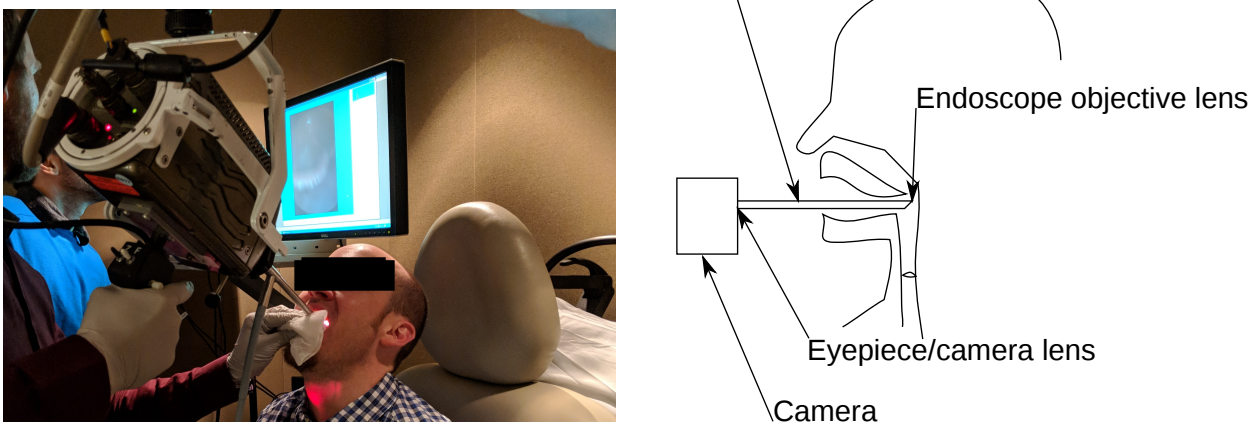
Figure 2.5: Illustration of the key phases in VF vibration. In (1) the VFs are initially closed. In (2) they are forced apart into a converging configuration by the glottal flow. In (3) the elastic forces balance the aerodynamic glottal forces. In (4) the elastic forces exceed aerodynamic forces and bring the VFs back to a close in a diverging configuration. Adapted from [19].

Since the initial work of Van den Berg, many refinements have been made to the theory of how the VFs self-oscillate. For example, some studies have shown that acoustic pressures generated in the subglottal and supraglottal tracts can help to reinforce self-oscillation [72, 71]. Others have elucidated the effect of the false VFs downstream on VF vibration [94].

The oscillation of the VFs leads to release of the glottal flow in bursts. These periodic bursts of air form the primary sound source for phonation. However, taken by itself, this source tone sounds like a simple buzz, which is far different from the output sound at the mouth. This difference, is due to the filtering effect of the supraglottal tract, detailed by Fant [22], who viewed the supraglottal tract as a type of filter that selectively strengthens and weakens certain frequency bands of the source tone. Specifically, the source tone has a nominal fundamental frequency in the range of 100 Hz to 200 Hz (for comfortable vowel intonation for men and women, respectively), with harmonics at varying strengths. The shape of the supraglottal tract strengthens and weakens different harmonics, producing local frequency bands with high intensities, known as formants. While the fundamental frequency of the source sound characterizes the pitch of our voice, the formants of the sound characterize its qualitative properties.

2.3 Clinical Measurements

There are a wealth of clinical measurements used for classifying patient voices. Some example measurements include qualitative voice perception by clinicians from audio recordings (or by ear), endoscopic video of VFs, the flow rate of air from the mouth and nose [34], electroglottography [31], and many more [4]. Of particular interest are video based measurements, since they capture the time-varying contours of the VFs. These time varying contours are directly related to the motion of the VFs, which makes them a valuable measurement for VF model parameters. Types of video-based measurements include videokymography, videostroboscopy, and HSV [12]; video techniques through endoscopy are collectively known as videoendoscopy. To conduct video based measurements, patients are seated, the tongue is retracted by the clinician, and an endoscope is inserted into the open mouth. After the camera has been focused, the patient is made to vocalize using the sound 'ee' (or other vocal gestures) [50]. An illustration of a videoendoscopy setup is shown in Figure 2.6.



(a) (b)

Figure 2.6: (a) A patient undergoing a HSV procedure in the clinic. The glow in the patient's neck is due to the strength of the light source. (b) A schematic of a typical videoendoscopy recording, using a rigid endoscope. The rigid endoscope (flexible endoscopes may also be used) usually contains two fibre optic bundles; one bundle carries a light source to illuminate the VFs, and another carries the image back to the camera (either for HSV, videostroboscopy, or videokymography).

Videostroboscopy records the VFs over multiple vibrations. By recording at a slightly higher frequency than the vibration of the VFs (this is done with a stroboscope that flashes at the higher frequency, synchronized with microphone readings to detect oscillations of the glottis [50]), the resulting videos appear to show a smooth vibration of the VFs through a single oscillation. Each frame however, actually captures the glottis across different cycles of the VF vibration. Thus, resolving the motion of the glottis inherently requires the oscillations to be periodic, aliasing the measurements in time [14]. In contrast to videostroboscopy, videokymography can record the vibration of the VFs through a single oscillation, with the trade off that the video be taken only along a single line (a line scan video). This obviously neglects the 2D shape of the glottal contours; in other words, videokymography has a poor spatial resolution. With advancements in imaging technology, recently the ability to capture vibrations of the VFs, without aliasing in time or decreasing spatial resolution, have been made possible using HSV. HSV is an information rich measurement from which many quantitative measurements can be derived, such as: the time-varying glottal contours, trajectories of the glottal contours, or the glottal area [12]. This is typically conducted with a rigid endoscope setup, shown schematically

in Figure 2.6(b).

There are many variable parameters in conducting HSV. Some parameters are controlled by the clinician, these include but are not limited to, the lighting conditions, frame rate, resolution, and choice of objective lens [13]. Of particular importance are the frame rate and resolution. Existing high-speed cameras used in endoscopy can record at frame rates ranging from about 2000 FPS to 20 000 FPS [14]. Typical spatial resolutions range from about 1000 px by 1000 px down to 100 px by 100 px [14]. Naturally the spatial resolution is only relevant when the physical size of the object being imaged is present. Without detailed data available, one can estimate the effective resolution (magnification factor) of the VFs in a generic high-speed glottal video based on typical VF dimensions. Since the typical length of a VF is on the order of 10 mm, and HSV aims to record the entire length of the VFs, we can estimate the ratio of pixels to equivalent physical distances as ranging from $10\,\mathrm{px\,mm^{-1}}$ to $100\,\mathrm{px\,mm^{-1}}$. Other imaging parameters are not easily controlled by the clinician. For example, the orientation of the camera relative to the VFs. This includes factors like the distance from the VFs to the objective lens, or tilting of the axis of view, such that it is not normal to the transverse plane, along which the glottis lies. Tilting of the axis of view, is a particularly important factor since off axis views of objects decrease their apparent size in an image (for example viewing a coin from and angle makes it look like an ellipse rather than a circle).

To use HSV in inverse analysis, quantitative measurements must be derived from the video. All such measurements are subject to uncertainty, which is influenced by a variety of imaging factors, such as the ones mentioned previously. In order to utilize HSV in inverse analysis procedures, the errors involved in any derived measurements should be known, since large errors on the order of the measured signal, can make inferred parameters meaningless. These measurement errors arise from a variety of sources. For example, distances in the video in pixels, must be converted into equivalent physical measurements through a set of reference points of known dimensions located in the video. This has been done with a projected grid on the surface of the folds with known dimensions, or with two laser points separated by a known distance [56, 57]. In tracking the glottal contours, edge detection algorithms may bias the true location of the glottal edge. This could be the result of poor lighting in the video, which increases image noise, or motion artifacts during imaging. Low spatial resolution videos, also increase uncertainty in derived measurements due to quantization of the glottis into larger pixels. Low frame rates can can alias measures of glottal width in time, by insufficiently resolving the motion of the VFs within an oscillation. Furthermore, due to the 2D nature of video, viewing the camera at non-orthogonal angles to the VFs can bias the apparent size of the glottis.

2.4 Numerical Phonation Modelling

Clinical approaches to the study of phonation provide a wealth of measurements, allowing clinicians to study what is happening during phonation. However, clinical measurements generally can not explain the reason, or mechanisms, behind these different measurements. In contrast to clinical studies, modelling of phonation takes a physics based approach to phonation, by modelling the underlying mechanisms, such as the layered viscoelastic properties of the VFs, the interaction of the glottal flow and VFs, or the propagation of acoustic waves in the supraglottal tract. Thus models can be used to understand why phonation behaves the way it does. As mentioned in Chapter 1, models of phonation include experimental models such as synthetic VFs that vibrate under an applied stream of air, as well as numerical models. Numerical models of phonation have the benefit of a computational framework; no manufacturing of new models, or physical sensors are needed to investigate a numerical model. This allows them to be easily coupled with inverse analysis techniques. There are numerous approaches to numerically modelling the phonation process. As mentioned in Chapter 1 these include, high-fidelity numerical models utilizing CFD and FEM, as well as low-fidelity numerical models using reduced order models.

Numerical models of phonation typically use representative population based parameters, such as a typical VF stiffness for a male or female. This is typically done due to the difficulty of directly measuring VF properties (as well as other structures in phonation). To explain patient specific phonation, model parameters must also be patient specific, which necessitates the use of inverse analysis techniques. However, not all phonation models can easily be incorporated into these techniques. While high-fidelity numerical models generally produce more accurate simulations, this does not make high-fidelity models preferable for inverse analysis. In general, the increased fidelity comes at the expense of increased computational cost. When applied in inverse analysis, this is particularly undesirable since inverse analysis techniques incur an additional computational cost on top of the numerical model. Furthermore, high-fidelity models are generally governed by numerous parameters (for example, in a simulation of a continuous VF, the material properties varying throughout the body would all be parameters). The large number of parameters means that many different parameter combinations could produce similar phonation behaviours. In turn, this reduces the certainty with which inverse analysis techniques can infer any particular parameter set. However, this conversely does not imply that low-fidelity models are optimal. For example, a low fidelity model may be insufficient in representing observed phonation behaviour. Parameter estimates for such a model are therefore meaningless. A balance exists then, in selecting a simplified model with fidelity sufficient to capture pertinent be-

haviour, yet simple enough to maintain computational tractability, and have meaningful model parameters. We elaborate on numerical phonation models in the next sections with a focus on reduced order models, which offer a balance between computational complexity and model fidelity. This is split into three components of phonation modelling: modelling of the VFs, the glottal flow, and acoustics.

2.4.1 Vocal Fold Models

Finite Element Methods

As described in Section 2.1.2, the VFs are a viscoelastic layered structure consisting of 3 main layers: an innermost core, consisting of the thyroarytenoid muscle, the lamina propria (consisting of three sub-layers), and an outermost mucosal layer. High-fidelity models of the VFs are typically performed with FEM techniques, which can simulate the continuum behaviour of the real VFs. FEM models are based on discretization of the domain (the VFs in this case) into a computational mesh. The viscoelastic properties of the VFs can be modelled in the framework directly with viscoelastic material laws. To account for the layered structure (variation of the viscoelastic properties with depth) of the VFs, the discretized mesh in FEM studies can incorporate mesh layers with varying material properties [1, 96]. To simulate phonation, FEM models are coupled with a glottal flow model, which supplies the driving pressures on the VF surface. While the highest fidelity FEM models simulate the 3D VFs, the discretization of the VFs can also be performed in 2D to decrease computational complexity [43]. For the purpose of the large scale studies on multiple parameter variations however, even 2D simulations are computationally prohibitive. Thus, reduced order models have been used to model the VFs.

Reduced Order Vocal Fold Models

Reduced order models (also called lumped mass models) simplify the VFs into a collection of discrete masses that interact with each other through springs and dampers. In these models, the combination of springs and dampers simulate the viscoelastic properties of the VFs, while each discrete mass simulates a small portion of the VF. Varying the number of masses allows the incorporation of varying material properties. For example, in the density of the material, as well as in the springs and dampers connecting the masses. Varying the number of degrees of freedom of the reduced order model (which is tied to varying the number of masses), allows for more complex motions, or mode shapes, to be simulated. As shown in theoretical studies, the motion of the VFs can be broken down into mode

shapes with varying amounts of energy [73, 74, 5]. Of primary importance are the mode shapes correponding bulk lateral motion, and the mucosal wave. Reduced order models are capable of representing different numbers of these modes based on the number of degrees of freedom they have. A few configurations of reduced order models of the VFs are shown in Figure 2.7.

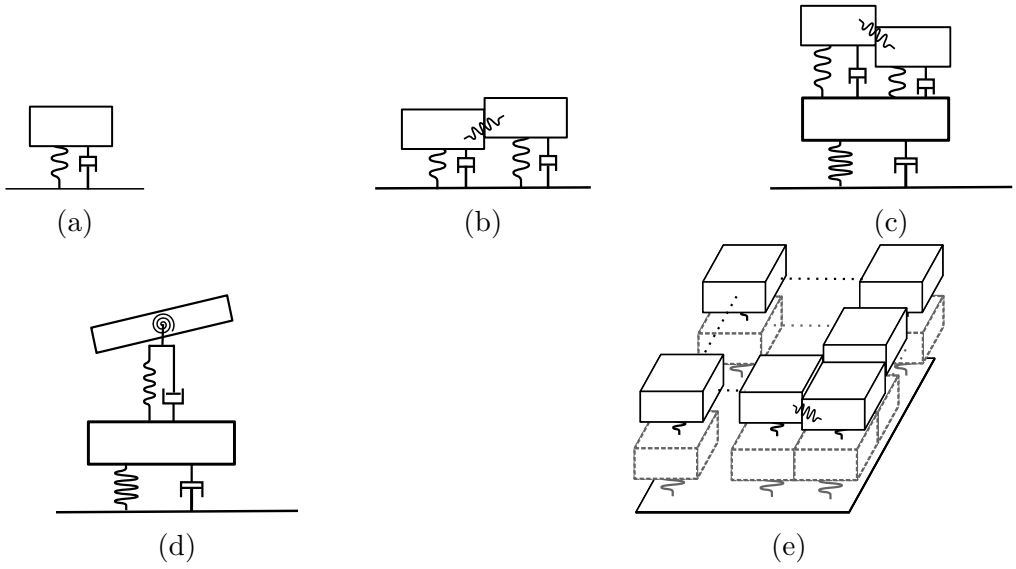


Figure 2.7: Illustration of some common lumped mass models with varying degree of freedoms (DOFs). (a) The one mass model with one DOFs; (b) two mass model with two DOFs; (c) three mass model with three DOFs; (d) two mass model with three DOFs; (e) a generic model can have many masses and many DOFs in order to represent 3D VF behaviour. Figures adapted from [21, 62].

The earliest reduced order models, modelled only a 2D slice of the VFs in the coronal plane, such as the one mass model (seen in Figure 2.7(a)) proposed by Flanagan [23]. The one mass model used one mass with one DOF to represent a VF. Such a model requires acoustic pressure feedback in order to achieve self oscillation due to the fact that it cannot model the mucosal wave; it can only capture a bulk lateral translation of the VFs. As mentioned in Section 2.2 however, the primary mechanism behind self-oscillation of the VFs is the glottal wave phenomenon in which the glottis forms alternating converging-diverging channels during the opening and closing phases respectively. This has been shown to create an asymmetry in driving pressures between the opening and closing phases of the VF oscillation, which is responsible for the net positive energy transfer from the airflow to

the VFs necessary for self-oscillation [76, 69]. To take into account this inferior-superior phase shift in the VF motion, the one mass model was extended into a two mass model by adding an additional mass along the inferior-superior direction (for a total of two DOFs) by Ishizaka and Flanagan [33] (shown in Figure 2.7(b)). Further fidelity may be achieved by taking into account the layered structure of the VFs (shown in Figure 2.4). In this case, an additional mass can be added below any masses in contact with the glottal flow, as in the 3 mass body-cover model (BCM), possessing three DOFs, of the VFs, shown in Figure 2.7(c), or its 2 mass variant that incorporates an additional rotational degree of freedom for the cover mass (shown in Figure 2.7(d)) [65]. This accounts for the different material properties between the body and cover layers (see Section 2.1.2 for details). Further complexity of the reduced order model can be achieved by extending the number of masses in the anterior-posterior directions to capture the 3D nature of the VFs as seen in Figure 2.7(e).

The governing equations for all reduced order models can be written in a similar fashion. In general, all reduced order models consist of a group of discrete masses. These masses can be spatially arranged to reflect the varying densities within the VFs. The masses are connected to each other with springs, representing the elastic properties of the VFs, and dampers, which represent the viscous behaviour of the VFs, together modelling the viscoelastic material properties. Forces from pressure exerted by the glottal flow, act on masses in contact with the flow. A generic model is illustrated in Figure 2.8.

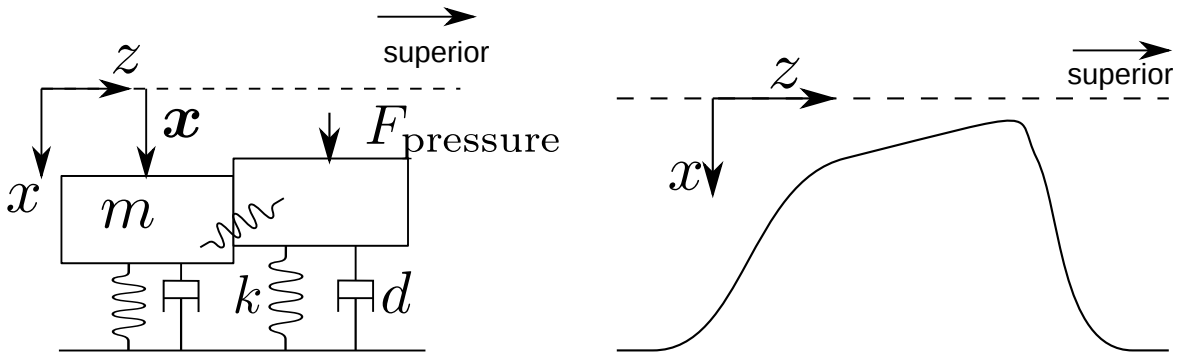


Figure 2.8: (a) A collection of masses with springs and dampers is used to represent the VFs, (b), in a reduced order model. The usage of spring and damping elements, k and d, simulate the viscoelastic damping properties of the VFs. Each mass m, represents a discrete portion of the VF.

All reduced order models are fundamentally based on Newton's second law, which

relates the forces acting on a mass to the rate at which it accelerates. For systems of discrete masses connected by various spring and damping elements, Newton's second law can be extended to represent the behaviour of the combined system's response to external forces using notation from linear algebra [46]. For a reduced order model, this is represented by equation (2.1)

$$\boldsymbol{F}_{\text{pressure}} = \boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{D}\dot{\boldsymbol{x}} + \boldsymbol{K}(\boldsymbol{x} - \boldsymbol{x}_{\text{rest}}), \tag{2.1}$$

where $\boldsymbol{M}$ is a matrix representing the masses in the system, $\boldsymbol{D}$ is a matrix representing the damping elements, and $\boldsymbol{K}$ is a matrix representing the spring elements. The terms $\boldsymbol{x}$ and $\boldsymbol{x}_{\text{rest}}$ represent the positions of the masses, and the rest lengths of the springs, respectively. Finally $\boldsymbol{F}_{\text{pressure}}$ is a vector representing the forces due to pressures from the glottal flow. Coupling equation (2.1) with an equation for $\boldsymbol{F}_{\text{pressure}}$ yields a model for VF self-oscillation.

As mentioned in Section 2.2, the self-oscillation of the VFs involves the VFs moving apart and then back together, in some cases colliding. Special consideration must be made to account for this collision in reduced order models. In reality, collision of the continuous VFs (and objects in general) occurs when their surfaces contact. This causes the remainder of the body to deform resulting in stresses generated through the continuous body. Reduced order model are based on discrete rigid masses however, and so cannot model collision through a deformation of the masses [21]. Thus, modelling of collision in reduced order models is typically done by allowing the VFs to pass through each other, whereupon additional collision springs are used to simulate the collision forces. The governing equation for such a model is given in equation (2.2)

$$\boldsymbol{F}_{\text{pressure}} = \boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{D}\dot{\boldsymbol{x}} + \boldsymbol{K}(\boldsymbol{x} - \boldsymbol{x}_{\text{rest}}) + \boldsymbol{K}_{\text{coll}}(t)(\boldsymbol{x} - \boldsymbol{x}_{\text{coll}}), \tag{2.2}$$

where $\boldsymbol{K}_{\text{coll}}(t)$ represents the collision springs, which are turned on only when collision occurs, and $\boldsymbol{x}_{\text{coll}}$ represents the rest length of the collision springs.

Body-Cover Model

Estimation of model parameters through Bayesian inference requires a model with a balance of high fidelity and low computational complexity. Since a pair of 2D driven VFs was used in this study to simulate HSV (see Chapter 3 for details), only 2D reduced order models are necessary. The BCM [65] is a well-known 2D reduced order model that captures most of the key aspects of phonation in 2D: the phase shift between the inferior and superior edges of the VFs (the mucosal wave), and the change in material properties between the stiff body layer and elastic cover layer of the VFs. Coupled with a 1D quasi-steady potential flow model (detailed equations for this model are given later in the section) allows for

self-oscillation to be simulated. Acoustics can be neglected since acoustics add additional computational complexity not pertinent to characterizing error in HSV. As a result, the BCM, with a slightly modified flow equation, was chosen to conduct Bayesian inference in this work. A schematic of the BCM is shown in Figure 2.9 below. Note that the 3 mass variant is used in this work [65].

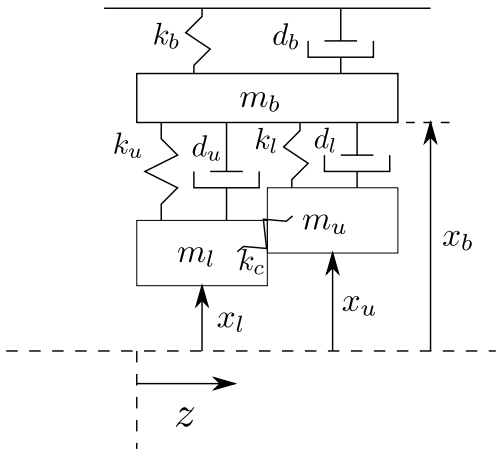


Figure 2.9: A schematic of the BCM. The z coordinate increases in the superior direction, the x coordinate is zero at the midline, and varies in the lateral direction. m represents the mass. The subscripts u, l, and b indicate properties associated with the upper, lower, and bottom masses.

The governing equations for the BCM are given in equation (2.3) [65]

$$\begin{aligned}\begin{bmatrix} F_{\mathrm{u}} \\ F_{\mathrm{l}} \\ 0 \end{bmatrix} =& \boldsymbol{M} \frac{\mathrm{d}^2}{\mathrm{d}t^2} \begin{bmatrix} x_{\mathrm{u}} \\ x_{\mathrm{l}} \\ x_{\mathrm{b}} \end{bmatrix} \\ &+ \boldsymbol{D} \frac{\mathrm{d}}{\mathrm{d}t} \left(\begin{bmatrix} x_{\mathrm{u}} - x_{\mathrm{u},0} \\ x_{\mathrm{l}} - x_{\mathrm{l},0} \\ x_{\mathrm{b}} - x_{\mathrm{b},0} \end{bmatrix} \right) + \boldsymbol{D}_{\mathrm{coll}} \frac{\mathrm{d}}{\mathrm{d}t} \left(\begin{bmatrix} x_{\mathrm{u}} - x_{\mathrm{u},0} \\ x_{\mathrm{l}} - x_{\mathrm{l},0} \\ x_{\mathrm{b}} - x_{\mathrm{b},0} \end{bmatrix} \right) \\ &+ \boldsymbol{K} \left(\left(\begin{bmatrix} x_{\mathrm{u}} - x_{\mathrm{u},0} \\ x_{\mathrm{l}} - x_{\mathrm{l},0} \\ x_{\mathrm{b}} - x_{\mathrm{b},0} \end{bmatrix} \right) + \eta \left(\begin{bmatrix} \eta_{\mathrm{u}}(x_{\mathrm{u}} - x_{\mathrm{u},0}) \\ \eta_{\mathrm{l}}(x_{\mathrm{l}} - x_{\mathrm{l},0}) \\ \eta_{\mathrm{b}}(x_{\mathrm{b}} - x_{\mathrm{b},0}) \end{bmatrix} \right)^3 \right) \\ &+ \boldsymbol{K}_{\mathrm{coll}} \left(\left(\begin{bmatrix} x_{\mathrm{u}} - x_{\mathrm{u,coll}} \\ x_{\mathrm{l}} - x_{\mathrm{l,coll}} \\ x_{\mathrm{b}} - x_{\mathrm{b,coll}} \end{bmatrix} \right) + \eta_{\mathrm{coll}} \left(\begin{bmatrix} (x_{\mathrm{u}} - x_{\mathrm{u,coll}}) \\ (x_{\mathrm{l}} - x_{\mathrm{l,coll}}) \\ (x_{\mathrm{b}} - x_{\mathrm{b,coll}}) \end{bmatrix} \right)^3 \right) \end{aligned}, \tag{2.3}$$

where $\boldsymbol{M}$, $\boldsymbol{D}$, and $\boldsymbol{K}$ represent the mass, stiffness, and damping matrices respectively. The η coefficients are coefficients of non-linearity and are given by $\eta = 100$ and $\eta_{\text{coll}} = 500$. In equation (2.3), each of the matrices are defined in equation (2.4).

$$\boldsymbol{M} = \begin{bmatrix} m_{\text{u}} & 0 & 0 \\ 0 & m_{\text{l}} & 0 \\ 0 & 0 & m_{\text{b}} \end{bmatrix}, \tag{2.4}$$

$$\boldsymbol{D} = \begin{bmatrix} d_{\text{u}} + d_c & -d_c & -d_{\text{u}} \\ -d_c & d_{\text{l}} + d_c & -d_{\text{l}} \\ -d_{\text{u}} & -d_{\text{l}} & d_{\text{b}} + d_{\text{u}} + d_c \end{bmatrix}, \tag{2.5}$$

$$\boldsymbol{K} = \begin{bmatrix} k_{\text{u}} + k_c & -k_c & -k_{\text{u}} \\ -k_c & k_{\text{l}} + k_c & -k_{\text{l}} \\ -k_{\text{u}} & -k_{\text{l}} & k_{\text{b}} + k_{\text{u}} + k_c \end{bmatrix}, \tag{2.6}$$

$$\boldsymbol{K}_{\text{coll}} = \begin{bmatrix} k_{\text{u,coll}} & 0 & 0 \\ 0 & k_{\text{l,coll}} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{2.7}$$

where each of the damping terms, d, are computed in terms of spring and mass constants given in the equations below

$$d_{\text{u}} = 2\zeta_{\text{l}}(m_{\text{u}}k_{\text{u}})^{\frac{1}{2}}, \tag{2.8}$$

$$d_{\text{l}} = 2\zeta_{\text{u}}(m_{\text{l}}k_{\text{l}})^{\frac{1}{2}}, \tag{2.9}$$

$$d_{\text{b}} = 2\zeta_{\text{b}}(m_{\text{b}}k_{\text{b}})^{\frac{1}{2}}, \tag{2.10}$$

$$d_{\text{u,coll}} = 2(\zeta_{\text{l}} + 1)(m_{\text{l}}k_{\text{l}})^{\frac{1}{2}}, \tag{2.11}$$

$$d_{\text{l,coll}} = 2(\zeta_{\text{u}} + 1)(m_{\text{u}}k_{\text{u}})^{\frac{1}{2}}. \tag{2.12}$$

The collision spring constants in equation (2.4) are given in equations (2.13) and (2.14).

$$k_{\text{u,coll}} = 3k_{\text{u}} \tag{2.13}$$

$$k_{\text{l,coll}} = 3k_{\text{l}} \tag{2.14}$$

All undefined values in the previous equations are given by a series of rules developed by Titze *et al.* [77] that relate three muscle activation parameters (a_{ct}, a_{ta}, a_{lc}) to the missing material properties above (see Section 2.1.2 for details). These muscle rules are elaborated on in the next section. In this work, only symmetric VF oscillations were considered, thus the BCM requires the specification of parameters for just one VF. Combined with the

muscle rules, the symmetric BCM is thus specified by 9 parameters (an asymmetric simulation would require 18 parameters; duplicating the set of 9 parameters for the other VF): a set of initial positions $(x_{\mathrm{u},0}, x_{\mathrm{l},0}, x_{\mathrm{b},0})$, initial velocities $(\dot{x}_{\mathrm{u},0}, \dot{x}_{\mathrm{l},0}, \dot{x}_{\mathrm{b},0})$, and three muscle activations $(a_{\mathrm{ct}}, a_{\mathrm{ta}}, a_{\mathrm{lc}})$, representing activations of the cricothyroid, thyroarytenoid, and lateral-cricoarytenoid (a negative activation exists for this muscle to represent the effect of the posterior cricoarytenoid) respectively.

Muscle Activation Rules

Titze *et al.* [77] specified three muscle activation rules for the cricothyroid, thyroarytenoid, and lateral-cricoarytenoid as control parameters for the material properties of the BCM. These muscle rules are useful because they allow relating physiological parameters (muscle activation effort of the intrinsic laryngeal muscles) to primitive parameters of a reduced order VF model, which then influences the outputs of phonation. This has physical meaning, as muscle activation is how individuals change VF parameters to change pitch, for example. Thus control of muscle activation to modify model parameters (rather than direct specification of material properties), is closer to the physiological mechanism of controlling phonation [77].

To measure the activation of these three muscles, Titze *et al.* [77] used a normalized value of 1 to represent maximum muscle activation, 0 to represent minimum muscle activation, and -1 to represent the maximum activation of an antagonistic muscle. For the cricothyroid and thyroarytenoid, the muscle activations range from 0 to 1 ($a_{\mathrm{ct}} \in [0, 1]$, $a_{\mathrm{ta}} \in [0, 1]$). For the lateral-cricoarytenoid, the activation ranges from -1 to 1, where the negative activation is used to represent activation of the posterior cricoarytenoid ($a_{\mathrm{lc}} \in [-1, 1]$). As mentioned in Section 2.1, the cricothyroid is primarily responsible for tensing (stretching) the VFs, the lateral-cricoarytenoid for adduction (bringing the VFs together), and the thyroarytenoid for relaxing the VFs [60, Chapter 5].

In terms of parameters of the BCM, these three muscle activations have different effects. The activation of the lateral-cricoarytenoid corresponds primarily to influencing the rest positions of the masses (decreasing $\boldsymbol{x}_0$), since it is related to adduction of the VFs. Both a_{ct} and a_{ta} are related to tensing of the VFs, although in different ways. Cricothyroid activation, a_{ct}, tenses the entire VFs by pulling the the thyroid cartilage that the VFs are attached to, away from the cricoid, thus lengthening (or shortening) the VFs. Thyroarytenoid activation, a_{ta}, mainly tenses the body layer of the VFs (the body layer consists of the thyroarytenoid). As a result, in the BCM, the activation of the cricothyroid corresponds mainly to a general stiffening of all the BCM springs (increasing k_u, k_l, k_b, k_c); while the activation of the thyroarytenoid corresponds mainly to a stiffening of the

body spring k_b [75, Fig.7, Fig.8]. Activation of the thyroarytenoid has the further effect of redistributing the masses in the BCM, increasing the body mass and decreasing the cover mass when the thyroarytenoid is activated [75, Fig.8]. The reader is referred to [77] for detailed equations on implementing the rules.

2.4.2 Fluid Models

Computational Fluid Dynamics

The highest fidelity models of the glottal flow use CFD to approximately solve the governing equations of the glottal flow (the Navier-Stokes equations). All CFD simulations begin with discretization of the glottis, and portions of the supraglottal and subglottal tracts, into a computational mesh, followed by specification of boundary conditions at the boundaries of the mesh. Simulation of the glottal is a coupled phenomenon, where the glottal flow forces the VFs through pressures and stresses exerted at the boundaries, while the VFs in turn modify the glottal flow by deforming, changing the boundaries of the glottal flow. There are several levels of fidelity at which the fluid portion of this coupled simulation process can be done.

One large difference in fidelity of CFD simulations is in dimensionality. As the VFs are inherently 3D structures (as seen in Figure 2.3), the glottal flow must also be 3D. High-fidelity CFD models have simulated the glottal flow in 3D [9, 95]. These simulations have shown agreement with physiologically observed values [96, 95] demonstrating the validity of 3D CFD glottal flow models (coupled with high-order numerical simulations of the solid VFs). While closer to the true dimensionality of the VFs, 3D simulations are computationally complex. As a result, 2D simulations of the glottal flow in the coronal plane are an attractive alternative [93, 2, 67, 91, 44, 43, 94]. This simplification to 2D simulations in the coronal plane, can be justified by the fact that many physiologically observed behaviours of the glottis occur primarily in the coronal plane. For example, the mucosal wave motion, or the adduction and abduction of the VFs. Furthermore, examination of the VF geometry has shown that the coronal cross section remains largely the same along the length of the VFs (in the anterior-posterior direction) [61]. While 2D simulations do reduce computational complexity, investigation of multiple parameters for phonation models (such as variable subglottal pressures) is still prohibitive. Thus further simplifications of the glottal flow have been investigated in simplified fluid models.

Simplified Fluid Models

Simplified models of the glottal flow use approximations of the Navier-Stokes equations. Some examples include 2D inviscid flows [38], inviscid flows with modelling of boundary layer phenomena during formation of the glottal jet [19], or 1D inviscid flow approximations. One of the simplest, and most common simplified fluid models is to assume a 1D inviscid flow under a quasi-steady assumption [21]. While such a model neglects much of the complex 3D effects of the glottal flow, for many flow conditions their results have been shown to agree reasonably well with CFD and experimental studies [10, 81, 92]. Under this assumption, the flow and resulting pressure on the VFs is governed by the Bernoulli equation. This assumption is only valid under specific conditions however. An example of two glottal flows configurations are illustrated in Figure 2.10, that illustrate when the inviscid approximation is valid.

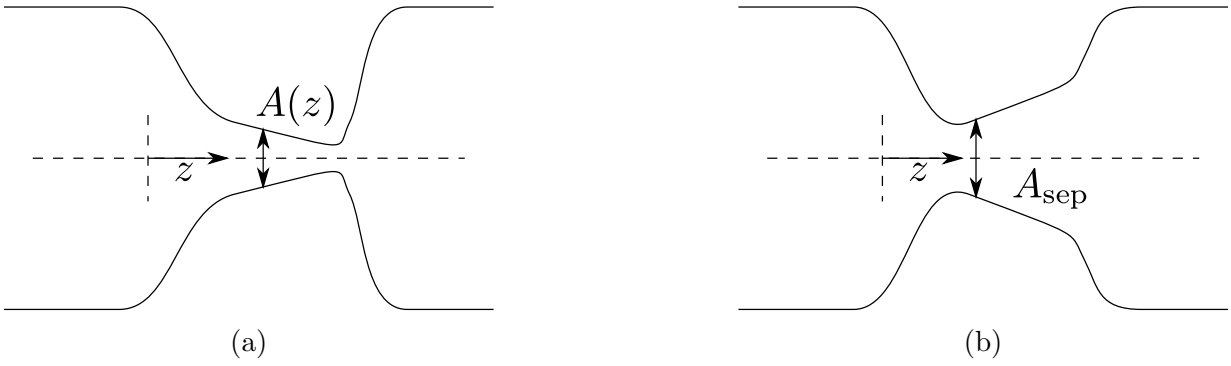


Figure 2.10: Two typical glottal flow channels with a 1D potential flow assumption. (a) A converging configuration enforces the Bernoulli regime throughout the glottis. (b) A diverging configuration leads to separation of the flow at some area A_{sep}.

Under the 1D inviscid assumption, the glottal flow varies only along the inferior-superior direction (the z coordinate in Figure 2.10). Due to the 1D assumption and conservation of mass, the speed of the glottal flow varies only due to the cross sectional area of the glottis, $A(z)$. Thus, the cross sectional area of the glottis can be used to determine the flow velocities and pressures in the glottis through the Bernoulli equation, but only when the inviscid assumption applies. In converging orientations of the glottis (as seen in Figure 2.10(a)), flow separation does not occur in the glottis, which allows application of the inviscid assumption; however, in diverging orientations (as seen in Figure 2.10(b)) flow separation occurs in the glottis [21], which is not naturally modelled in inviscid flows. Thus flow separation must be accounted for by specification of a separation point, and the Bernoulli

approach applied only up to the point where the flow separates. The area at separation is denoted by A_{sep}.

Application of the Bernoulli equation starts with using two known pressures at the subglottal location, and separation location. The Bernoulli equation can be applied between these two points in order to solve for the volumetric flow rate through the glottis:

$$P_{\text{sub}} + \frac{1}{2}\rho v_{\text{sub}}^2 = P_{\text{sep}} + \frac{1}{2}\rho v_{\text{sep}}^2 \tag{2.15}$$

$$P_{\text{sub}} + \frac{1}{2}\rho(\frac{Q}{A_{\text{sub}}})^2 = P_{\text{sep}} + \frac{1}{2}\rho(\frac{Q}{A_{\text{sep}}})^2 \tag{2.16}$$

$$Q^2({A_{\text{sep}}}^{-2} - {A_{\text{sub}}}^{-2}) = \frac{2}{\rho}(P_{\text{sub}} - P_{\text{sep}}) \tag{2.17}$$

$$Q = \sqrt{\frac{2}{\rho}(P_{\text{sub}} - P_{\text{sep}})({A_{\text{sep}}}^{-2} - {A_{\text{sub}}}^{-2})^{-1}}, \tag{2.18}$$

where Q is the volumetric flow rate, P_{sep} is the pressure at separation, P_{sub} is the subglottal pressure, A_{sub} is the subglottal area, and ρ is the density of air. Past the separation point, it is assumed that the glottal flow forms a jet and a constant separation pressure (P_{sep}) acts on the glottis. The pressure in the Bernoulli regime is given in equation (2.19)

$$P_{\text{sub}} + \frac{1}{2}\rho\left(\frac{Q}{A_{\text{sub}}}\right)^2 = P(z) + \frac{1}{2}\rho\left(\frac{Q}{A(z)}\right)^2, \tag{2.19}$$

where $A(z)$ is the area at the location z, and $P(z)$ is the corresponding pressure.

Rearranging equation (2.19) for the pressure at a location z, and noting that the pressure after separation is a constant, the intra-glottal pressure is then given by equation (2.20), where Q is substituted from (2.18)

$$P(z) = \begin{cases} P_{\text{sub}} + \frac{1}{2}\rho Q^2\left({A_{\text{sub}}}^{-2} - A(z)^{-2}\right) & \text{if } z \leq z_{\text{sep}} \\ P_{\text{sep}} & \text{if } z > z_{\text{sep}} \end{cases}. \tag{2.20}$$

Equations (2.18) and (2.20) still require the separation pressure, P_{sep}, and separation location, z_{sep}, to be solved. There are various approaches to doing this. The simplest is to assume that the pressure recovery effect (the tendency of the fast moving jet to have a lower pressure than the slower moving surroundings) in the glottal jet is negligible, thus pressure in the jet is the same as the surrounding supraglottalpressure, P_{sup}. Another

approach is to account for the pressure recovery effect with a pressure recovery factor. Ishizaka *et al.* [33] determined the pressure at separation to be given by equation (2.21)

$$\begin{aligned} P_{\text{sep}} &= P_{\text{sup}} - \frac{1}{2}\rho k_e \left(\frac{Q}{A_{\text{sep}}}\right)^2, \\ k_{\text{e}} &= \frac{2A_{\text{sep}}}{A_{\text{sup}}}\left(1 - \frac{A_{\text{sep}}}{A_{\text{sup}}}\right), \end{aligned} \tag{2.21}$$

where k_{e} is the pressure recovery factor.

The location of separation remains to be determined. When the glottis forms a convergent passage, the pressure gradient is favourable and so separation does not occur. In contrast, when the glottis forms a divergent passage the glottal flow remains attached to the glottal walls until a certain critical condition is reached and the flow separates (see Figure 2.10(b)). For non-aerodynamically smooth models, this is often assumed to occur at the mass with the minimum area [65, 33]. For aerodynamically smooth models, some authors have used ad-hoc methods to determine this location, such as by assuming that separation occurs at a constant ratio of A_{sep} to A_{min} of 1.3 [41, 75].

Body-Cover Flow Model

The flow model used for the BCM, which is used in this work, is derived from Story *et al.* [65], but neglects acoustic pressures. Following similar arguments as discussed previously, they derived the pressures acting on the upper and lower masses when the glottis is open, as shown in Equations (2.22) and (2.23)

$$P_{\text{l}} = \begin{cases} P_{\text{sub}} - (P_{\text{sub}} - P_{\text{sup}})\left(\frac{A_{\text{u}}}{A_{\text{l}}}\right)^2 & A_{\text{u}} \leq A_{\text{l}} \\ P_{\text{sub}} - (P_{\text{sub}} - P_{\text{sup}})\left(\frac{A_{\text{u}}}{A_{\text{l}}}\right)^2 & A_{\text{u}} > A_{\text{l}} \end{cases}, \tag{2.22}$$

$$P_{\text{u}} = P_{\text{sup}}, \tag{2.23}$$

where P_{sup} is the supraglottal pressure and P_{sub} is the subglottal pressure. The areas are given as

$$A_{\text{u}} = 2x_{\text{u}}L_{\text{g}}, \tag{2.24}$$

$$A_{\text{l}} = 2x_{\text{l}}L_{\text{g}}, \tag{2.25}$$

where the factor of 2 is applied for a symmetric BCM. The term L_{g} is the length of the glottis, and comes from the muscle rules specified by Titze *et al.* [77]. During closure of

the glottis, the pressures are given by:

$$P_{\mathrm{l}} = \begin{cases} P_{\mathrm{sub}} & A_{\mathrm{u}} > 0 \text{ and } A_{\mathrm{l}} = 0 \\ 0 & A_{\mathrm{l}} = 0 \end{cases}, \tag{2.26}$$

$$P_{\mathrm{u}} = \begin{cases} P_{\mathrm{sup}} & A_{\mathrm{l}} = 0 \text{ and } A_{\mathrm{u}} > 0 \\ 0 & A_{\mathrm{u}} = 0 \end{cases}. \tag{2.27}$$

Finally the forces on the upper and lower masses can then be obtained by multiplying by the surface area of each mass in contact with the flow. This is given in Equations (2.28) and (2.29)

$$F_{\mathrm{l}} = P_{\mathrm{l}} T_{\mathrm{l}} L_{\mathrm{g}}, \tag{2.28}$$

$$F_{\mathrm{u}} = P_{\mathrm{u}} T_{\mathrm{u}} L_{\mathrm{g}} \begin{cases} P_{\mathrm{sup}} & A_{\mathrm{l}} = 0 \text{ and } A_{\mathrm{u}} > 0 \\ 0 & A_{\mathrm{u}} = 0 \end{cases}, \tag{2.29}$$

where T is the thickness of the mass in the inferior-superior direction. All the remaining constants can be found using the muscle activation rules provided by Titze [77].

The BCM flow model adds 2 parameters to the BCM's 9 parameters: P_{sub} and P_{sup}. In total, the modified BCM used in this study is completely specified by 11 parameters, given in Table 2.1.

Table 2.1: The list of parameters completely specify the BCM. In this work, the parameters are assumed to remain constant throughout time.

Parameter	Description
$\begin{bmatrix} x_{\mathrm{u},0} & x_{\mathrm{l},0} & x_{\mathrm{b},0} \end{bmatrix}$	Vector of 3 initial positions, one for each of the three masses in the BCM
$\begin{bmatrix} \dot{x}_{\mathrm{u},0} & \dot{x}_{\mathrm{l},0} & \dot{x}_{\mathrm{b},0} \end{bmatrix}$	Vector of 3 initial velocities, one for each of the three masses in the BCM
P_{sub}	subglottal pressure
P_{sup}	supraglottal pressure
a_{ct}	A measure of cricothyroid activation; ranges from 0 to 1
a_{ta}	A measure of thyroarytenoid activation; ranges from 0 to 1
a_{lc}	A measure of lateral-cricoarytenoid activation; ranges from -1 to 1

Note that for the current model, the glottal flow equations are not needed. This is due to the neglection of acoustics. For the model considered in this case, acoustics only add additional computational complexity but have no influence on characterizing the error inherent in high-speed glottal video.

2.4.3 Acoustic Models

Acoustics can play an important role in phonation due to the fact that acoustics influence the end result of what others hear (through the filtering effect of the supraglottal tract), as well as because acoustic pressures can influence the driving pressures on the VFs [72] (some modes of speech such as resonant speech take advantage of this [72]). Similar to how FEM and CFD can be used to generate high-order models of the VFs and glottal flow, Computational Aero-acoustics techniques exist that can be used to simulate acoustics from a first-principles model, based on simulating the propagation of pressure waves [93]. Such simulations however, are extremely computationally intensive, due to the small spatial scales and small time scales required to simulate the propagation of pressure waves travelling through the fluid. As a result, simplified models of the acoustics are preferable.

In general there are two approaches to simplified modelling of acoustics in phonation. In an 'electrical analog' approach, the propagation of acoustic pressure waves is computed through electrical analogs of acoustic components [21] (also termed the 'transmission line model' [33]). This approach comes from the ladder topology of circuits used for designing electronic filters, and models the acoustic tract as a series of acoustic tubes represented by electric elements. The 'electrical analog' approach allows the modelling of acoustics in either the frequency domain or time domain. While the electrical analog approach can be used with a known source tone to study the filtering effect of the supraglottal tract in the frequency domain, when coupled with reduced order models a time domain solution is used. Another approach is the wave-reflection analog approach, developed by Kelly and Lochbaum [37], which similarly models the acoustic tract as a series of acoustic tubes, but models the pressure wave propagation through physical principles (rather than an electrical-analog) and strictly in the time-domain. Due to the usage of discrete acoustic tubes, this model assumes a 1D planar wave propagation in each section. In contrast to the the 'electrical analog' approach, the time domain formulation makes it difficult to model frequency-dependent, and other losses [21].

The addition of acoustics adds computational complexity to a reduced order model due to the need to track acoustic pressures. In the context of Bayesian inference, these acoustic pressures become additional parameters that must be estimated. Furthermore, in

this work Bayesian inference is applied on measurements obtained from a simulated HSV experiment, which is not affected by acoustics. As a result, acoustics are neglected in this work.

2.5 Inverse Analysis

In many problems associated with modelling real-world phenomena, a set of parameters for a model are specified, and the resulting behaviour (this could entail any set of quantities of interest that the model produces) of the model is solved. Such a problem is called a forward problem. Typically the behaviour of the forward model is well described, meaning that given a set of parameters, the behaviour of the model can be solved uniquely. Forwards problems are useful because they allow the prediction of results based on an understanding of some phenomena. For example, the weather can be predicted using models of fluids.

In contrast to the forward problems, in some cases it is the parameters that generate a result that are of interest. In this case, it is desired to estimate what parameters created a certain result. This is known as an inverse problem. Inverse problems frequently arise when direct measurements of a quantity (a model parameter) cannot be made. In this case, the value of the measurement can be inferred through indirect measurements. A trivial example is a model of a spring. There is no measurement tool, that directly measures the stiffness of a spring. Instead, measurements of forces and displacements of the spring can be taken. Using the well known Hooke's law for springs, these forces and displacements can be related to the stiffness of the spring, through Hooke's law. In this problem, Hooke's law is the forward model. This is a trivial inverse problem, since Hooke's law is a linear model. Calculation of the spring stiffness is easily computed given a force and displacement. Most inverse problems involve more complicated forward models, which typically cannot be inverted analytically.

An example of such an inverse problem, is in medical imaging through x-ray computed tomography, which can generate cross-sectional images of the body. This is based on projecting thin x-ray beams through the body and measuring the x-ray intensity that passes through to the other side. Based on where the beam passes through the body, the intensity of the x-ray measured will change, due to the absorption properties of different tissues. In this problem, physicians do not care about the intensity of the measured x-ray, but rather what the x-ray was attenuated by, for example a tumor. Thus the parameter actually desired, is a map of tissues that the x-rays passed through. Since there is no known way to map the properties of internal tissues without dissecting the patient, indirect measurements must be made, such as through x-rays. Here, the forward problem gives

the attenuation of x-rays based on the distribution of tissues inside the body. What is desired, is the map of tissues in the body, given x-ray measurements, in other words an inverse problem [36, pg.198]. Similarly in the field of phonation, clinicians can take various measurements of patients (such as HSV or audio recordings), however measurements of the actual VF properties (such as asymmetric muscle activations, which indicate VF paralysis) are more valuable information about patients. Since these properties cannot be directly measured, a natural approach is to infer their values through inverse analysis applied to indirect clinical measurements.

There are two general approaches to solving inverse problems. In optimization methods, a measure of error between the observed behaviour and model behaviour, called the objective function, is minimized [68]. The solution to the inverse problem is then the parameter set that minimizes the objective function. While such methods can approximately account for the effects of measurement errors [68], they are only approximations. In contrast, Bayesian inference is an approach to inverse analysis that models the inverse problem in a probabilistic framework [36]. In contrast to optimization methods, the solution to the inverse problem in Bayesian inference, is a description of which parameter sets have higher probability densities. This allows uncertainties in measurements to be naturally accounted for.

2.5.1 Optimization Methods

Optimization methods in inverse analysis, aim to find the parameter set that reduces a measure of the amount of error between a model and given measurements. One of the earliest examples of optimization methods for inferring reduced order model parameters was conducted by running simulations over a discretized set of parameters and comparing the difference in glottal width amplitude between a measurement and a model in a manual fashion [47]. Motivated by this approach, further studies have incorporated more advanced automatic optimization procedures.

These studies rely on minimizing a measure of error, called the objective function Γ, between an observed measurement and an equivalent output from a model [68, Chapter 3]. One common objective function is the Euclidean-Norm (also known as the L_2 norm), given in equation (2.30)

$$\Gamma(\boldsymbol{x}) = \|\boldsymbol{y}_{\text{error}}\|_2 = \sqrt{(\boldsymbol{y}_{\text{model}}(\boldsymbol{x}) - \boldsymbol{y}_{\text{measured}})^T (\boldsymbol{y}_{\text{model}}(\boldsymbol{x}) - \boldsymbol{y}_{\text{measured}})}, \tag{2.30}$$

where $\boldsymbol{y}$ are measurements, and $\boldsymbol{x}$ are parameters.

Döllinger *et al.* [16] performed one of the earliest inverse analysis studies with automatic optimization of an objective function, for a two-mass model. They optimized an objective function in the frequency domain, based on minimizing the sums of norms of the Fourier transform amplitudes and phase shifts for each of the left and right VF displacements of a two-mass model. Using this procedure, they were able to estimate three parameters: subglottal pressure, and left and right asymmetry factors, referred to as Q factors. These left and right asymmetry factors (Q_l and Q_r), multiplied the default masses of Ishizaka and Flangan's [23] original two mass model, while dividing the spring constants. Thus if the two Q factors are the same, the reduced order model is symmetric. They applied this objective function to glottal widths measured from HSV. Minimizing the objective function, they were able to determine parameters for both healthy and pathological VF recordings. Interestingly, Döllinger *et al.* found approximately equal asymmetry factors for the left and right VFs when they applied the technique to healthy VF recordings, and different asymmetry factors when they applied the technique to pathological VFs. Other studies have used different objective functions and also found success estimating different parameter sets. For example, Wurzbacher *et al.* [83] based their objective function on minimizing wavelet decomposition coefficients of glottal trajectories from high-speed video of the glottis. This was done to estimate 8 time dependent Q factors for a 3D reduced order model. The model consisted of 12 masses; three rows of masses discretized the anterior-posterior direction of each VF, while each row consisted of two masses to discretize the inferior-superior direction [Fig. 2][83]. Three Q factors were used to modify the spring and mass constants for each of the left and right VFs, respectively. The remaining two Q factors were used to modify the anterior-posterior masses and springs, for each VF, respectively. Wurzbacher *et al.* performed the minimization on a variety of voice recordings, including recordings of left-side and right-side paralyzed VFs as well as healthy VFs. Their results were able to differentiate Q factors on the left and right sides, for example, showing that left-paralyzed VFs corresponded to lower Q factors on the left VF springs and masses. A number of other optimization approaches have also been conducted, all of them giving promising results on the estimated parameters [88, 87].

The actual method for optimizing the objective function, can be done with a variety of algorithms, such as Powell's optimization algorithm [83], Nelder-Mead [16], or genetic optimization [59]. Two common classes of algorithms include gradient techniques and direct search methods (Nelder-Mead) [16]. Gradient based techniques, compute the gradient of the objective function with respect to the optimized parameters in order to compute the optimal 'search' direction so as to minimize the objective function. This means that gradient based techniques can converge quickly. However, complex objective functions can make computation of the gradient expensive, furthermore the gradient is typically not

available analytically, and so must be calculated numerically, thus increasing the computational complexity. The presence of local minima can also make these methods unstable [16]. Direct search methods do not rely on computing gradients of the objective function, but rather directly search for the parameter set that minimized the objective functions. This makes such algorithms more stable, as well as more applicable when gradients cannot be computed [16].

All these optimization techniques only produce point estimates. However, it is known that derived measurements inherently have some error. This is not naturally accounted for in the optimization approach to inverse problems. Furthermore, optimization problems sometimes have local minima, for example, Döllinger *et al.* [16] found local minima in their study (described earlier). The presence of local minima hints that multiple parameter sets could reproduce an observed measurement within the limits of observation error. Whether these local minima are significant, or not, depends on the magnitude of the measurement error. A large measurement error, could mean that a local minimum in the objective function, is simply the result of a random error. This would make the point estimate of an optimization approach invalid, since it neglects the significant local minima. Optimization techniques are unable to naturally account for this. In contrast, Bayesian inference provides a natural way to represent this uncertainty on the estimated parameter sets with a probabilistic framework.

2.5.2 Bayesian Inference

Bayesian inference is an approach to inverse analysis utilizing Bayes' formula, which describes the probability of an event H, based on knowledge of conditions that might be related to the event E

$$\Pr(H \mid E) = \frac{\Pr(E \mid H)}{\Pr(E)} \Pr(H), \tag{2.31}$$

where $\Pr(H \mid E)$ is the posterior, $\Pr(H)$ is the prior, $\Pr(E \mid H)$ is the likelihood, and $\Pr(E)$ is the evidence. For example, consider that H is the probability that a specific coin is weighted, and E represents the fact that the specific coin has come up heads 10 times, and tails only once. $\Pr(H \mid E)$ is the probability that the coin is weighted given the observed tosses, called the posterior. $\Pr(E \mid H)$ is the probability of the observed tosses, given that the coin is weighted, called the likelihood. Finally $\Pr(E)$ is the probability of the given tosses for any coin, and $\Pr(H)$ is the probability that any coin is weighted. Bayes' formula provides the exact way in which these probabilities combine.

While Bayes' formula, in equation (2.31) is formulated in terms of probabilities, this is not the most convenient form. Continuous quantities, such as muscle control parameters

like a_{ct}, are typically characterized by continuous probability distributions, which are usually characterized through probability densities [36]. For a continuous random variable X with probability density π_x the probability that x ranges from $[x_1, x_2]$ is defined as

$$\Pr(x \in [x_1, x_2]) = \int_{x_1}^{x_2} \pi_x(s) \mathrm{d}s. \tag{2.32}$$

Bayes' formula also applies for probability densities [36, Chapter 3]

$$\pi_{\text{posterior}}(\boldsymbol{x} \mid \boldsymbol{y}_{\text{obs}}) = \frac{\pi_{\text{likelihood}}(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{x}) \pi_{\text{prior}}(\boldsymbol{x})}{\pi_{\text{evidence}}(\boldsymbol{y}_{\text{obs}})}, \tag{2.33}$$

where π indicates a probability density, $\boldsymbol{x}$ represents a vector of model parameters, and $\boldsymbol{y}_{\text{obs}}$ represents a vector of measurements. Since parameters used in reduced order models are typically continuous, the continuous form of Bayes' formula will be used in the remainder of the thesis.

Prior

The prior density $\pi_{\text{prior}}(\boldsymbol{x})$ is a probabilistic representation of what is known about the parameters $\boldsymbol{x}$ 'prior' to any measurements. For example, it is known that parameter sets for VF muscle activation, vary only between 0 to 1 for a_{ct} and a_{ta}, and -1 to 1 for a_{lc}. As a result, the prior for these parameters should have 0 probability density when the muscle parameter lies outside its range.

Likelihood

As discussed at the beginning of this section, inverse analysis is complicated by the fact that measurements have inherent uncertainty due to observation noise which is impossible to eliminate. In a deterministic model of the phonation system this implies that a single parameter set can produce a range of different measurements within measurement error, due to the fact that the measurement system will incur some random uncertainty. In other words, measurements of the model are a random variable, with a distribution dependent on the parameter set. For a measurement $\boldsymbol{y}_{\text{obs}}$ and parameter set $\boldsymbol{x}$, this distribution is $\pi_{\text{likelihood}}(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{x})$, and comes from a model of the measurement error and the forward model that maps model parameters $\boldsymbol{x}$ to model measurements $\boldsymbol{y}_{\text{obs}}$.

The form of the likelihood depends on the forward model (reduced order model), and the measurement noise model. One common model for measurement noise is additive noise.

Using an additive noise model, the measurements $\boldsymbol{y}_{\text{obs}}$ can be expressed as:

$$\boldsymbol{y}_{\text{obs}} = \boldsymbol{y}_{\text{model}} + \boldsymbol{e} = F(\boldsymbol{x}) + \boldsymbol{e}, \tag{2.34}$$

where $\boldsymbol{e}$ is a vector of random variables, $\boldsymbol{y}_{\text{model}} = F(\boldsymbol{x})$ is the forward model, and $\boldsymbol{x}$ is a vector of parameters. Since the noise is treated as independent of $\boldsymbol{x}$, by integrating over the measurement error the likelihood is defined as [36]

$$\pi_{\text{likelihood}}(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{x}) = \pi_{\boldsymbol{e}}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x})). \tag{2.35}$$

Observation errors can have many different distributions, however, the most common model (and the model we choose in this work) is to treat $\boldsymbol{e}$ as Gaussian with mean $\boldsymbol{\mu}$ and covariance Σ, i.e.

$$\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma). \tag{2.36}$$

So the likelihood becomes

$$\pi_{\text{likelihood}}(\boldsymbol{y}_{\text{obs}}|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}) - \mu)^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}) - \mu)\right), \tag{2.37}$$

here k is the number of observations in the observation vector, i.e. $\boldsymbol{y}_{\text{obs}} \in \mathbb{R}^k$. In many cases the mean of the additive Gaussian noise is 0 (called unbiased noise), and measurements are independent and identically distributed. As a result the covariance matrix becomes $\Sigma = \sigma^2 I$, where I is the identity matrix and σ is the standard deviation of each of the measurements. In this case, the likelihood for a series of measurements becomes

$$\begin{aligned}
\pi_{\text{likelihood}}(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{x}) &= \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))\right) \\
&= \frac{1}{\sqrt{(2\pi)^k|\sigma^2 I|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))^{\mathrm{T}}(\sigma^2 I)^{-1}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))\right) \\
&= \frac{1}{\sqrt{(2\pi\sigma^2)^k}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))^{\mathrm{T}}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))\right) \\
&= \frac{1}{\sqrt{(2\pi\sigma^2)^k}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x})\|^2\right),
\end{aligned} \tag{2.38}$$

where k is the number of measurements. An additive, unbiased, independent and identically distributed Gaussian noise model is used in this work, as described in Section 3.3.2. Note the similarity with equation (2.30) of the objective function. Minimization of the objective function, corresponds to maximization of the likelihood, known as the maximum likelihood estimate (MLE). While optimization approaches correspond to the MLE in Bayesian inference, they do not naturally capture the uncertainty of the parameters.

Evidence

The final term needed in computing the posterior is the evidence, $\pi_{\text{evidence}}(\boldsymbol{y}_{\text{obs}})$. This term is simply a normalizing factor which scales the product of the likelihood and the prior (the unnormalized posterior) to a probability. In other words, the evidence is a scaling factor that multiplies the unnormalized posterior such that the integral under the resulting function is 1.

Posterior

Computation of the posterior simply involves multiplication of the prior and likelihood, followed by normalization of the evidence, as shown by Bayes' formula in equation (2.33). While Bayes' formula is simple, computation of the posterior in practice can be computationally costly, especially as the number of parameters being estimated increases. For example, consider inferring a parameter in the range 0 to 1. Direct application of Bayes' formula could be applied at N discrete points along the range. This involves evaluating the forward model N times. If an additional parameter is estimated, N^2 points must now be distributed over two dimensions, such that each dimension is discretized by N points. As the number of dimensions increases, the number of points evaluated increases exponentially [36, Section 3.6]. This exponentially increasing complexity is sometimes referred to as the curse of dimensionality [8].

There are various numerical approaches to the solution of Bayes' formula, some of which are more suited to high-dimensional inference, than the approach of direct application of Bayes' formula. Sample based approaches approximate the posterior with numerous discrete samples, each sample being a parameter set. This avoids numerical integration errors in approximating the posterior through direct computation, which require scaling the number of points exponentially with respect to the dimension [36, Section 3.6]. Variational approaches approximate the posterior analytically in a local region. This avoids stability issues that can occur in sample based methods [29]. While directly computing the discretized posterior over a grid of parameter sets is subject to the curse of dimensionality, in this work only 2 parameters are inferred as described in Section 3.3.2. Thus a direct application of Bayes' formula is used in this work due to its simplicity, and the fact that it computes the posterior explicitly.

Characterizing Posterior Estimates

The posterior obtained from Bayesian inference consists of a probability distribution over the estimated parameter set. When the distribution applies over one or two parameters, it can be interpreted easily in a graphical format. However, as the number of dimensions grows, these distributions can be difficult to visualize. Thus 'estimators', are used to characterize these distributions [36, Section 3.1.1]. Estimators can be used to characterize point estimates, called point estimators, as well as their associated uncertainties, called spread (or interval) estimators [36, Section 3.1.1].

There are a number of point estimators, two of the most popular being the mean and maximum aposteriori (MAP). The mean estimate of a posterior distribution is given by:

$$\boldsymbol{x}_{\text{mean}} = \int_{\Omega} \boldsymbol{x} \pi_{\text{posterior}}(\boldsymbol{x} \mid \boldsymbol{y}_{\text{obs}}) \mathrm{d}\boldsymbol{x}, \tag{2.39}$$

where $\boldsymbol{x}$ is a vector of parameters, $\boldsymbol{y}_{\text{obs}}$ is a vector of measurements, and Ω is the domain of the parameter space. The MAP estimate is the parameter set that maximizes the posterior probability density:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max \pi_{\text{posterior}}(\boldsymbol{x} \mid \boldsymbol{y}_{\text{obs}}), \tag{2.40}$$

where $\arg\max$ denotes the argument, $\boldsymbol{x}$, that maximizes the posterior density.

The aforementioned estimators characterize the entire posterior. However in some cases, only the probability over a subset of the parameters are of interest, independent of the remaining parameters. For example, in estimating the parameters $(a_{\text{ct}}, P_{\text{sub}})$, one may be interested only in the a_{ct} parameter. These are called marginal distributions, which represent the probability densities of a subset of parameters, independent of the others. A marginal distribution of x_i over $\boldsymbol{x}$ is given by [36, Section 3.1.1]:

$$\pi_{\text{posterior}}(x_i \mid \boldsymbol{y}_{\text{obs}}) = \int_{\Omega} \pi_{\text{posterior}}(x_1 \dots x_{i-1} x_{i+1} \dots x_n \mid \boldsymbol{y}_{\text{obs}}) \mathrm{d}x_1 \dots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \dots \mathrm{d}x_n, \tag{2.41}$$

where Ω is the domain of all parameters except x_i. The above discussed estimators can be identically applied to marginal distributions, and thus give spread or point estimators over the marginalized parameter x_i.

There are a number of spread estimators as well. One primarily used measure is the covariance given in equation (2.42)

$$\Sigma_{\boldsymbol{x}} = \int_{\Omega} (\boldsymbol{x} - \boldsymbol{x}_{\text{mean}})(\boldsymbol{x} - \boldsymbol{x}_{\text{mean}})^T \pi_{\text{posterior}}(\boldsymbol{x} \mid \boldsymbol{y}_{\text{obs}}) \mathrm{d}\boldsymbol{x}, \tag{2.42}$$

where $\Sigma_{\boldsymbol{x}}$ is a covariance matrix, and $\boldsymbol{x}_{\text{mean}}$ denotes the mean of $\boldsymbol{x}$. The covariance is a measure of spread about the mean [36]. Off diagonal entries, (i, j), of the covariance matrix, measure the relative spread between x_i and x_j while entries along the diagonal, measure the spread of the respective element x_i with itself.

Computation of the posterior covariance is an integration problem, as seen in equation (2.42). When $\boldsymbol{x}$ is high-dimensional or the forward model is complex and non-linear, the integral may become computationally infeasible. The Laplace approximation is a way to overcome this, by approximating the posterior density with a Gaussian. This involves approximating the forward model with its Taylor series expansion about the MAP estimate. This is shown in equation (2.43), where J is the Jacobian of $\boldsymbol{y}_{\text{obs}}$ with respect to $\boldsymbol{x}$, and $o((\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}})^2)$ indicates higher-order terms of the Taylor series

$$F(\boldsymbol{x}) \approx F(\boldsymbol{x}_{\text{MAP}}) + J\mid_{\boldsymbol{x}_{\text{MAP}}} (\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}}) + o((\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}})^2). \tag{2.43}$$

Using this linearization, it can be shown that the posterior can be approximated with a Gaussian distribution, when the prior is uniform. This is known as the Laplace approximation [70, 30]

$$\begin{aligned}
\pi_{\text{posterior}}(\boldsymbol{x} \mid \boldsymbol{y}_{\text{obs}}) &\propto \pi_{\text{likelihood}}(\boldsymbol{y}_{\text{obs}} \mid \boldsymbol{x})\pi_{\text{prior}}(\boldsymbol{x}) \\
&\propto \frac{1}{\sqrt{(2\pi)^k \left|\Sigma\right|}} \exp\left(-\frac{1}{2}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))^T \Sigma^{-1} (\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}))\right) \\
&\approx \frac{1}{\sqrt{(2\pi)^k \left|\Sigma\right|}} \exp\left(\begin{array}{c} -\frac{1}{2}(\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}_{\text{MAP}}) + J(\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}}))^T \Sigma^{-1} \\ (\boldsymbol{y}_{\text{obs}} - F(\boldsymbol{x}_{\text{MAP}}) + J(\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}})) \end{array}\right) \\
&\propto \exp\left(-\frac{1}{2}(J(\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}}))^T \Sigma^{-1} (J(\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}}))\right) \\
&= \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}})^T J^T \Sigma^{-1} J(\boldsymbol{x} - \boldsymbol{x}_{\text{MAP}})\right),
\end{aligned} \tag{2.44}$$

where Σ is the covariance matrix of $\boldsymbol{y}_{\text{obs}}$.

From equation (2.44), the covariance matrix of the resulting Gaussian is then given by

$$\Sigma_{\boldsymbol{x}} = (J^T \Sigma^{-1} J)^{-1}, \tag{2.45}$$

where $\Sigma_{\boldsymbol{x}}$ is the covariance matrix of the parameters. For the case of independent, identically distributed measurements, $\Sigma = \sigma^2 I$ and so the covariance matrix is given by equation (2.46)

$$\Sigma_{\boldsymbol{x}} = \sigma^2 (J^T J)^{-1}. \tag{2.46}$$

Chapter 3

Methodology

HSV as a clinical measurement of phonation is useful because many quantitative measurements can be derived from it, from the glottal area to the trajectories of the glottal edges. Furthermore, these are closely related to the actual motion of the VFs. However, HSV is subject to many uncertainties, which come from variables in the HSV process. In particular, off-axis viewing angles, low frame rates, and low spatial resolutions can all contribute to uncertainty in measurements derived from HSV. To investigate the effect of these three HSV imaging variables on Bayesian inference applied to measurements derived from HSV, a simulated HSV experiment was conducted. A consumer digital single lens reflex camera (DSLR) was used to image a pair of actuated VFs, simulating HSV. The VFs were actuated according to a reduced order model with known parameters, held constant throughout the entire motion. Applying Bayesian inference to the glottal width waveform, measured from the simulated high-speed video, allowed comparison with the known parameters from the driving model. This setup was used to capture a series of simulated high-speed videos of the VFs, with varying imaging parameters (spatial resolution, frame rate, and angle of view) to investigate the effect of these imaging parameters on Bayesian inference applied to HSV. In the remainder of this chapter, details of the experimental methodology are presented. Section 3.1 details the experimental setup used, Section 3.2 details the experimental parameters investigated, and Section 3.3 describes the post-processing procedures conducted to perform Bayesian inference.

3.1 Experimental Setup

A schematic of the experiment setup is shown in Figure 3.1. The experimental setup consists of 3 main components: a pair of 2D VFs (extended along the anterior-posterior direction to form a 3D geometry), an actuation system controlled by 4 stepper motors to drive the VFs, and an imaging system mounted on a tripod used to record a simulated superior view of the VFs.

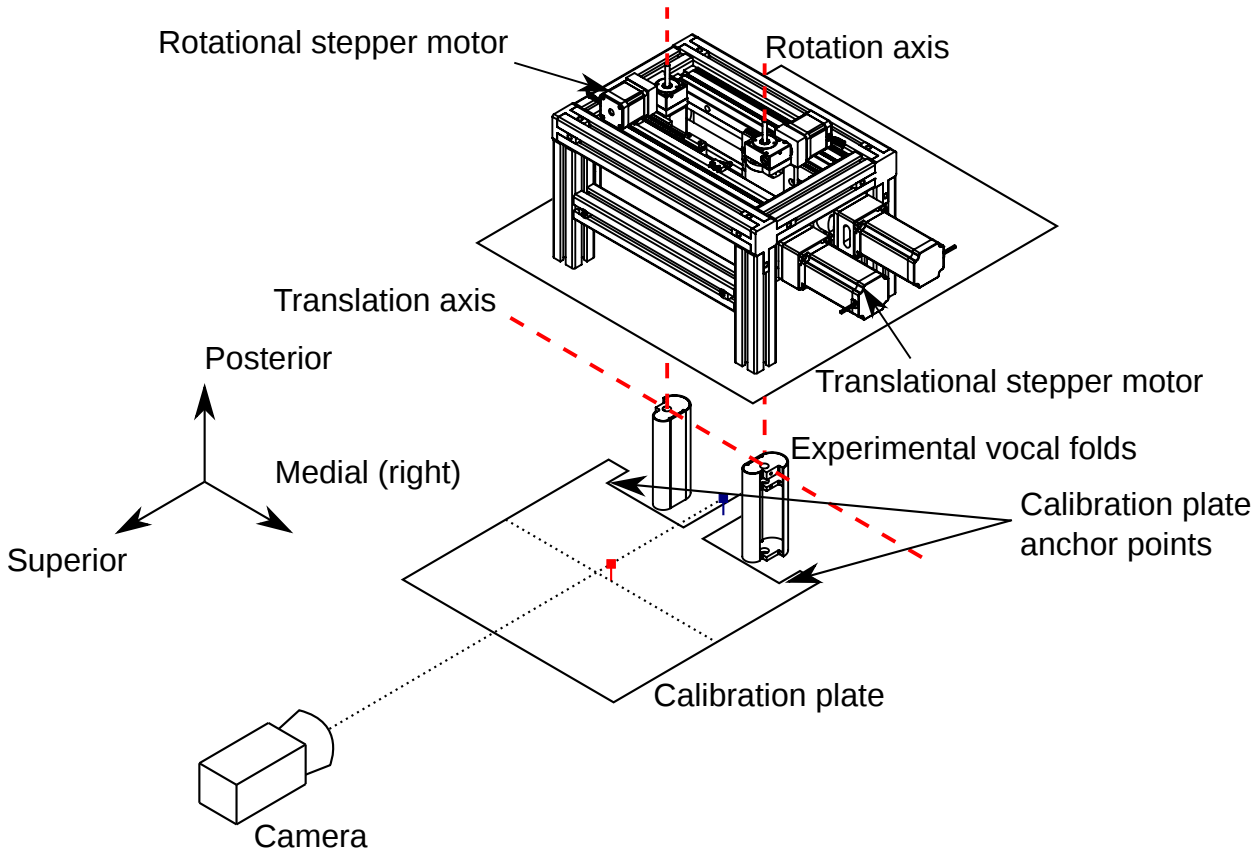


Figure 3.1: A schematic of the experimental setup. A pair of experimental VFs (that can rotate and translate) are driven with a motion system. A camera is used to image the VFs, with the position of the folds initially set with a calibration plate. The calibration plate is also used to set the angle of view of the camera using two dots shown in red and blue. It was removed during the experiment.

Figure 3.1 shows the overall schematic of the experiment. A camera is focused at the

midplane of the VFs to simulate HSV. The 2D VFs translate and rotate in a 2D plane (the coronal plane), actuated by a motion system powered by stepper motors. This motion is captured from a superior view using the camera to simulate HSV. Varying angles of view were obtained by aligning the camera's angle of view through reference points on the calibration plate. Two points located on the plate, are arranged along the desired angle, and subsequently aligned within the image obtained from the camera. The initial position of the VFs was also set with the calibration plate, by moving the VFs against the plate's known width (the central prong in Figure 3.1). The calibration process is detailed later in the section. The anchor points, were two cylindrical rods, attached to the motion system and are shown in Figure 3.2.

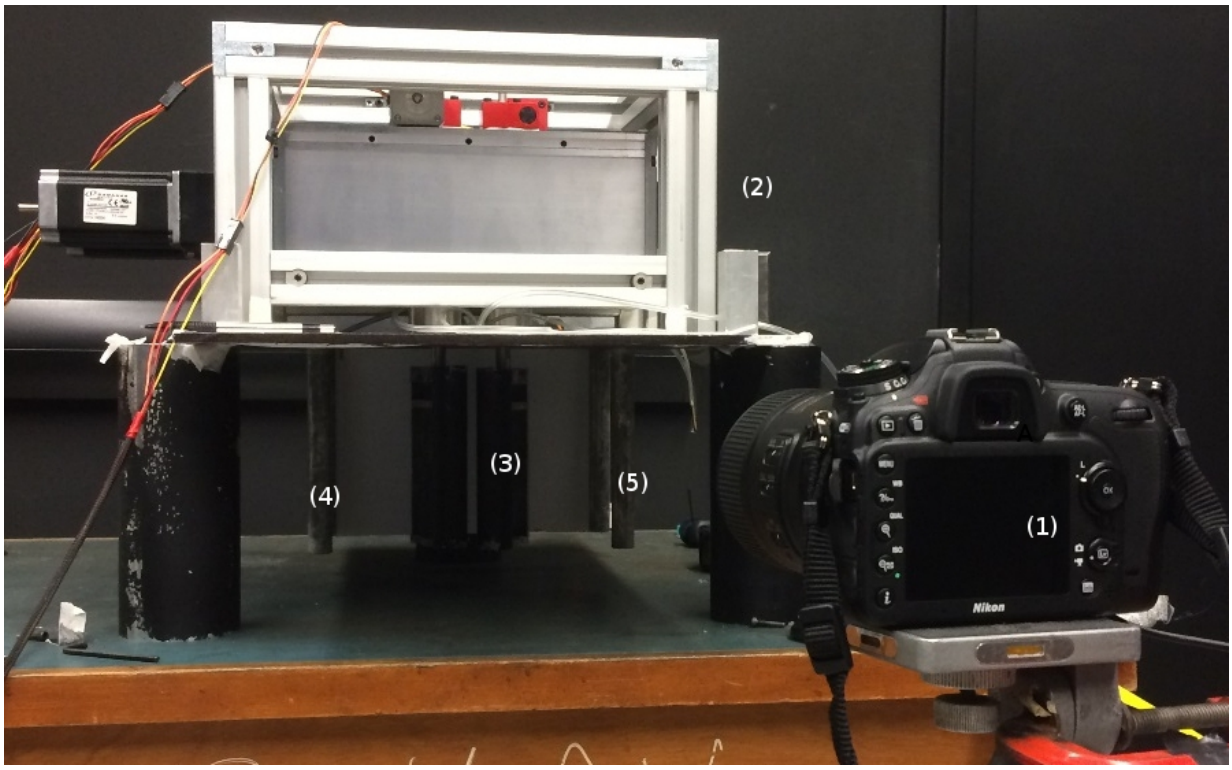


Figure 3.2: A picture of the actual experimental setup used, without the calibration plate. (1) is a NIKON D3200 camera mounted on a tripod used to simulate HSV of the VFs. (2) consists of a stepper motor driven motion system, which actuates a pair of rigid VFs, shown in (3). (4) and (5) are cylindrical rods fixed to the motion system, against which the calibration plate was secured.

Figure 3.2 shows the actual components used in the experiment. The simulated high-speed camera was a Nikon D3200 DSLR. The experimental VFs are constructed from aluminum, and painted black to contrast with a white background (representing the glottis). These are 7.5× larger than life-size. While the DSLR has a low recording frame rate (30 FPS), the motion of the simulated VFs are slowed down to effectively increase the frame rate of the camera. Similarly, the geometry and motions of the VFs are scaled by a factor of 7.5, which increases the effective spatial resolution of the camera. Details of these components are presented in the next sections.

3.1.1 Vocal Fold Geometry

Each of the VFs consists of the medial portion of Scherer *et al.*'s [54] M5 geometry extruded in the anterior-posterior direction. The M5 geometry is shown in Figure 3.3.

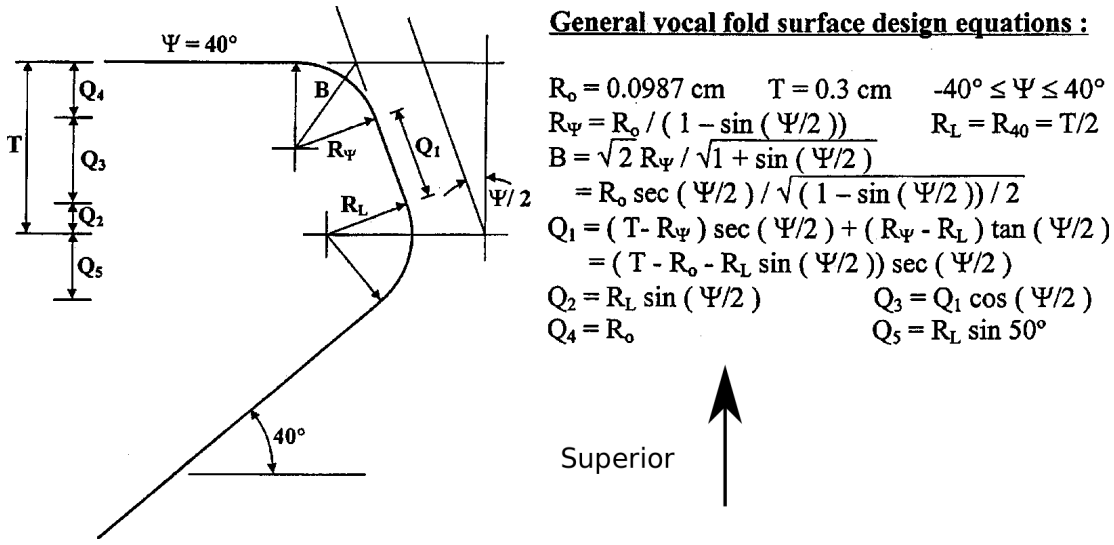


Figure 3.3: The M5 geometry defined by Scherer *et al.* [54] in physiological coordinates. Adapted from [54].

As seen in Figure 3.3, the M5 geometry is governed by a set of equations, parameterized by the glottal angle Ψ. Scherer *et al.* [54, 53, 40, 39] used this glottal angle to simulate the glottis at various converging-diverging angles, in other words to simulate the mucosal wave. This motion is difficult to generate using real actuation systems, since multiple dimensions adjust dynamically in length (for example R_Ψ, Q_1, $Q_2 \ldots$). Implementing this motion

would require actuators to modify each of these lengths throughout a simulated motion, which is both costly and complicated. As a result, an alternative approach to capturing the mucosal wave was used. The M5 geometry was constructed at the glottal angle $\Psi = 0°$ and scaled up by 7.5× [62]. Variation of the actual glottal angle, was then accomplished by rotating the geometry. In addition, a translational degree of freedom allows simulation of a bulk lateral vibration of the VFs. This is illustrated in Figure 3.4.

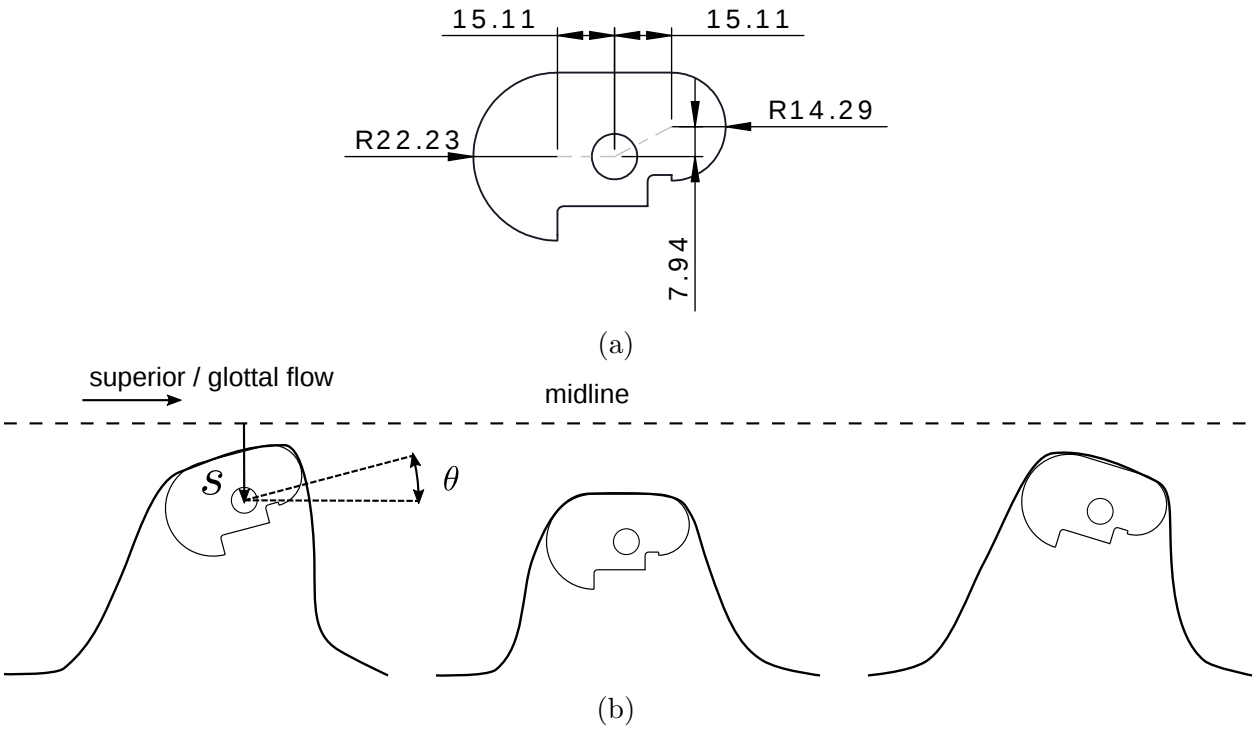


Figure 3.4: (a) The 7.5× life-size medial plane of the M5 geometry used for the experiment, and its dimensions in millimetres. The central circle in (a) is a control shaft that translates and rotates the VFs. (b) Illustration of the translational, s, and rotational, θ, degrees of freedom of the M5 model can simulate two modes of the VF motion. The 3 configurations correspond to phases of the mucosal wave: an opening converging configuration, a maximum opening of the glottis, and a closing diverging configuration.

The VFs are given a translational and rotational degree of freedom to simulate the mucosal wave. Note that the rotation point, shown in Figure 3.4 is located behind the medial surface of VFs. This means that the mucosal wave will introduce some inferior-superior motion of the VFs. When imaged by the camera, this has the effect of moving

the inferior/superior edges closer and further away from the image plane. As a result of perspective projection, this can make the distance between the inferior and superior edges appear either smaller or larger. This inferior-superior motion is small relative to the thickness of the glottis (in the inferior-superior direction) and so has little effect when compared with the effect of imaging the superior edge of the VFs versus the inferior edge of VFs.

3.1.2 Body-Cover Model and Control System

Body-Cover Model

The experimental VFs are driven by a BCM. This model consists of 2 portions, a symmetric BCM and a simplified fluid model, which are coupled togther. These two models were described in Sections 2.4.1, and 2.4.2, respectively. The combination of these two models results in a system of 3 second order differential equations (one for each position $(x_{\mathrm{u}}, x_{\mathrm{l}}, x_{\mathrm{b}})$ given in equation (2.3)), governed by 11 parameters. These parameters are listed in Table 2.1. A numerical solution of the BCM using a 4th order Runge-Kutta algorithm, with a time step of 1/350 ms, was used to drive the experimental VFs. The numerical solution of the BCM was conducted in a control system, that then moved the experimental VFs to appropriate positions through an actuation system. These systems are described in the next section.

Control and Actuation System

The actuation system was numerically controlled through a cRIO-9074 real-time computer, manufactured by National Instruments. A LabVIEW program, written in LabVIEW's graphical programming language with additional libraries from C, was used to solve a BCM, and move the actuation system appropriately. This was implemented through a control loop, that solved the BCM and moved the VFs to corresponding positions (based on a mapping approach detailed later). The loop solved the BCM, and moved the experimental VFs to corresponding positions, once every 200 ms, until 7 min elapsed. This time was chosen since it was believed that it would provide a sufficient amount of data (about 17 oscillations of the VFs were simulated). For every 200 ms of wall clock time, the cRIO simulated 1/14 ms of time for the model. In other words, the motion of the VFs was slowed down in real time by a factor of 2800, which increases the effective frame rate of the imaging system. Thus a 30 FPS recording in real time is effectively a 84 000 FPS recording. At the end of every time step, the calculated positions from the BCM are sent

to the actuation system to move the VFs to the computed positions. The actuation system that controlled the position of the VFs, was a black-box system that generated the motion within each time step based on an undocumented procedure. Thus the exact motion profile within each time step is unknown.

Each VF is actuated through a control shaft. The control shaft is connected to a stage system, driven by stepper motors. A close-up of the stage system is shown in Figure 3.5.

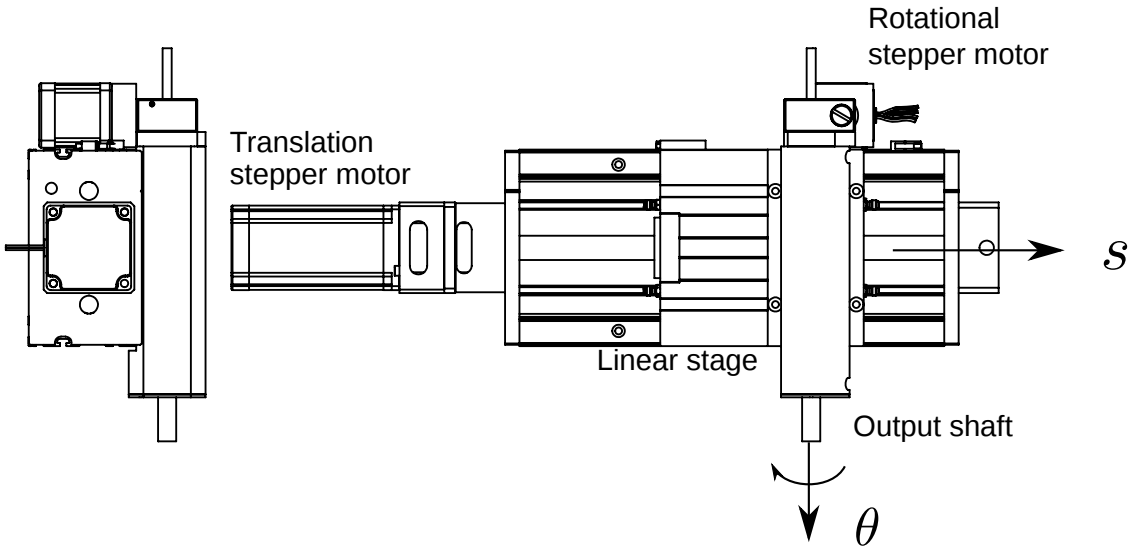


Figure 3.5: The actuation system gives the VFs two degrees of freedom: translation and rotation. Pictured is one of the two identical assemblies. The translation stage is powered by one stepper motor, and moves a second stage that provides rotation, powered by another stepper motor.

The actuation system consists of a shaft mounted through a gearbox, with a 10 to 1 reduction ratio (10 rev of the stepper motor results in 1 rev of the VF). The gearbox is driven by a stepper motor and provides the rotational degree of freedom of the VF. This gearbox is further mounted to a linear stage, which is driven by another stepper motor. The linear stage converts rotation of the stepper to translational motion of the VF at a rate of $5\,\mathrm{mm\,rev^{-1}}$ (the lead of the screw driving the stage). For further details of the actuation system, refer to [62].

Mapping of Body-Cover Model to Experimental Vocal Folds

As mentioned previously, the VFs were controlled through a BCM (specified in Section 2.4.1), however, the experimental VFs do not follow a BCM geometry and are scaled by a factor of 7.5 As a result, an approach to converting BCM displacements into equivalent displacements of the experimental VFs was required. This was accomplished by first scaling all BCM model displacements by a factor of 7.5 , and then applying a mapping of BCM displacements to points on the experimental VF geometry, as illustrated in Figure 3.6.

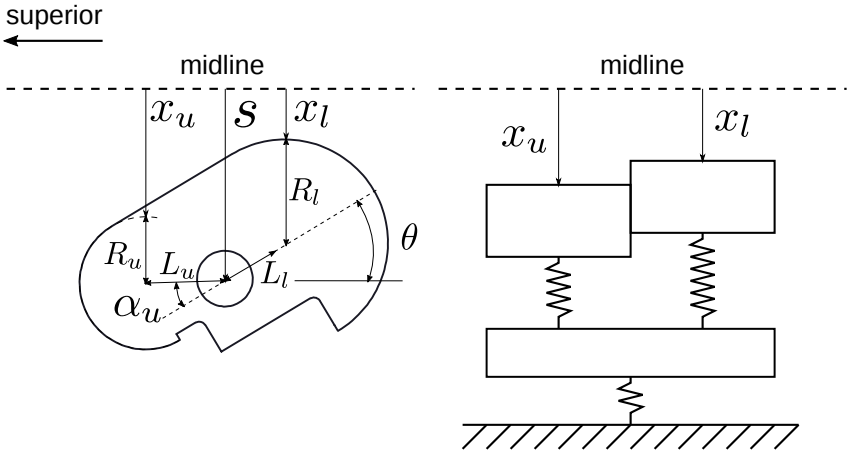


Figure 3.6: The s and θ coordinates of the experimental VFs are computed in order to satisfy x_{u}, and x_{l} from the BCM. The parameters have values as follows: $L_{\mathrm{u}} = 17.07\,\mathrm{mm}$, $L_{\mathrm{l}} = 15.11\,\mathrm{mm}$, $\alpha_{\mathrm{u}} = 27.71°$, $R_{\mathrm{l}} = 22.23\,\mathrm{mm}$, and $R_{\mathrm{u}} = 14.29\,\mathrm{mm}$.

The equation that converts each of the BCM's upper and lower mass displacements into an equivalent translation and rotation of the experimental VFs, is given in equation (3.1). Actuation of the stage system then involves conversion of the BCM displacements into equivalent experimental VF coordinates. This conversion is achieved through numerical solution of equation (3.1)

$$\begin{bmatrix} x_{\mathrm{u}} \\ x_{\mathrm{l}} \end{bmatrix} = \begin{bmatrix} s - R_{\mathrm{l}} - L_{\mathrm{l}} \sin(\theta) \\ s - R_{\mathrm{u}} - L_{\mathrm{u}} \sin(\alpha_{\mathrm{u}} - \theta) \end{bmatrix}. \tag{3.1}$$

In this work, the BCM was symmetric, however the mapping procedure can be applied to individual VF motions. Thus it is also possible to simulate asymmetric VF oscillations.

Collision Modelling of the Experimental Vocal Folds

As mentioned in Section 2.4.1, reduced order models model collision of the VFs by allowing them to pass through each other and activating collision springs. However, the experimental VFs are obviously unable to pass through each other. As a result, the motion system implements an additional algorithm which prevents collision of the experimental VFs when the driving BCM experiences collision. This is illustrated in Figure 3.7.

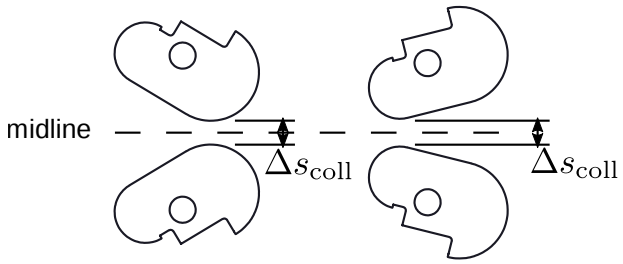


Figure 3.7: To prevent collision of the experimental VFs, a gap of constant width of 1 mm was maintained between the VFs, while the underlying BCM collided. The angles of the VFs adjust according to the BCM simulation, so the VFs translate in order to maintain a constant gap. This ensures that the VFs can smoothly transition out of the collision state.

During collision, the VFs are separated by a constant gap, Δs_{coll}, however the angle of VFs continues to adjust according to the BCM mapping equations. This causes the translational degree of freedom s to adjust in order to maintain the constant gap. When the VFs transition out of collision, the collision rule ensures that no sudden motions need to be made to obey the mapping equations.

3.1.3 Imaging System

Camera and Image Information

To simulate a high-speed endoscopic recording of the VFs, a commercial digital camera (Nikon D3200) was used. This allowed the capture of video at a resolution of 1920 px by 1080 px corresponding to a sensor area of approximately 18.8 mm by 12.5 mm. The frame rate of the imaging system is much less than those of typical cameras used in HSV (frame rates of up to 20 000 FPS are possible), however as a result of the slowed motion of the experimental VFs, the effective frame rate of the system was increased by a factor of 2800.

Similarly the size of the experimental VFs was scaled to be 7.5 times larger than real life. This allowed the usage of a relatively inexpensive consumer grade camera, to simulate a much more expensive high-speed camera.

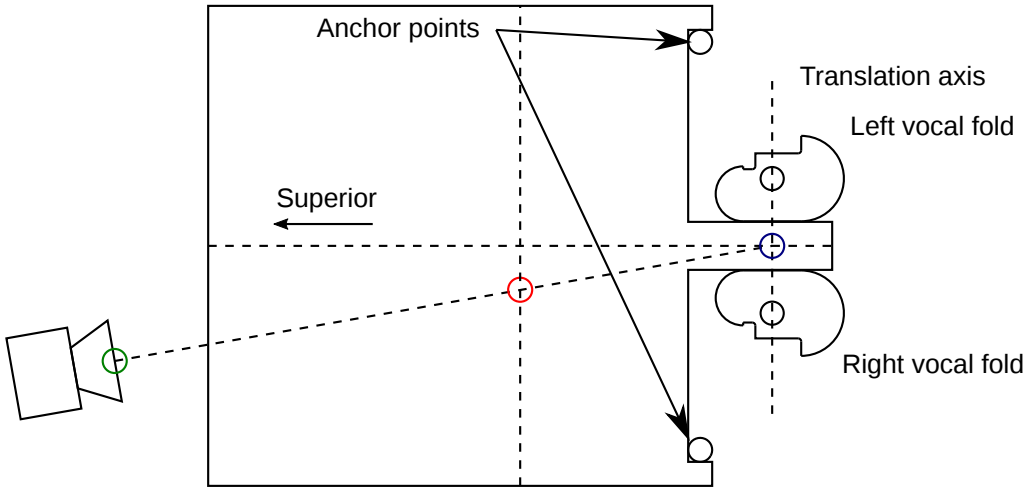
Distance Calibration and Perspective

Extraction of quantitative measures from high-speed video of the glottis, requires a conversion from length measures in the images, which are in pixels, into physical distances. In real HSV recording situations, this has been done with parallel laser beams with a fixed separation distance, projected onto the VFs [56, 57]. To calibrate pixel distances into physical distances in the current experiment, a reference scale was placed in the view of each simulated high-speed video, which provided a reference length. This reference scale was located at the superior edge of the VFs. For each case the resulting calibration was approximately $100\,\mathrm{px\,mm^{-1}}$. However, since the VFs were constructed at $7.5\times$ life size, the effective resolution is approximately $750\,\mathrm{px\,mm^{-1}}$.

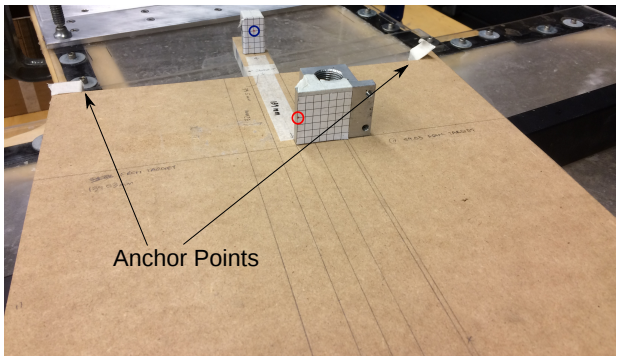
The above calibration applies only at the plane of the reference scale, at the superior edge of the VFs. However, the glottal width can be measured not only at the superior edge, but also at the inferior edge. Due to perspective projection of images from a true 3D scene onto a 2D plane, the calibration value at the inferior edge will be less than the calibration value at the superior edge. In other words the ratio of pixels to mm should decrease as objects are imaged farther away from the image plane. For the present study, the px to mm conversion at the superior edge of the VFs was chosen to convert pixel measurements into physical measurements, even at the inferior edge. This is a similar situation in real HSV recording situations. As mentioned earlier, an approach to calibrating pixel measurements to physical distances is through projecting parallel laser beams with a fixed separation distance onto the VFs. These parallel beams can only be projected onto the superior surfaces of the VFs, or adjacent superior tissues. Thus the calibration provided does not apply at the inferior edge of the VFs. Since there is no direct depth information in a 2D image, it is difficult to account for this error.

3.1.4 Angle of View and Position Calibration

To increment the angle of view of the camera in specified angles around the VFs, a calibration plate was used with a point alignment method. Successive dots were placed along lines of sight corresponding to each angle. These dots were then aligned to the center of the image plane within the camera's view in order to obtain the desired angle of view. This is illustrated in Figure 3.8.



(a)



(b)

Figure 3.8: (a) Schematic of the calibration plate used to set the angle of view, and the initial position of the VFs. The blue and red points are two points in the transverse plane, along a line with the desired angle of view, α. The red dot moves a long a line 139 mm away from, and parallel to, the translation axis. The green point is the center of the camera's image plane. The anchor points are secured against cylindrical rods, fixed to the motion system controlling the VFs. These correspond to (4) and (5) from Figure 3.2. (b) Image of the actual calibration plate (made of medium density fibreboard) used to set the initial position of the VFs.

The linear stages were not referenced to a home position, so they were placed against the central prong manually, as shown in Figure 3.8 where the VFs are aligned against the central prong of the calibration plate. When placed flat against the central prong, the VFs

have known positions and angles, $(s, \theta) = (15\,\text{mm}, 0^\circ)$, since the prong is 30 mm wide and machined with parallel sides. The block with the red dot, shown in Figure 3.8(b), was positioned along a line (shown in Figure 3.8(a)) according to the desired angle of view. The calibration plate was constructed of medium density fibreboard. The plate was placed on a horizontal surface, which the VF apparatus was also placed on, thus making them parallel. The points (1) and (2), were placed at the same height, coloured on 5 mm grid paper.

3.2 Experimental Parameters

The experimental parameters investigated can be split into imaging parameters, and parameters for controlling the experimental VFs. Imaging parameters affect Bayesian inference by changing the quality of the simulated HSV. The parameters used to control the BCM driving the experimental VFs, serve as a reference to compare with the parameters inferred through Bayesian inference. In the next two sections, the imaging parameters, and BCM parameters investigated are given and justified.

3.2.1 Imaging Parameters

While usage of HSV for measuring glottal area introduces many imaging variables (such as lighting of the scene, perspective effects of the lens, lens aberrations...) into the inverse analysis process, it would be impossible to investigate all of them. Thus for this study, only three factors are investigated. As stated in Chapter 1 these are: frame rate, resolution, and off-axis viewing angle. The off-axis viewing angle was selected since it was forecast to be an important effect on quantitative measurements derived from HSV; it is known that off-axis viewing angles will decrease the apparent size of a surface in the resulting image (in this case the glottis). Frame rate and spatial resolution were chosen because they have a direct impact on the resulting video; a low spatial resolution can make the glottis blurry at every frame of the video, while a low frame rate can make it difficult to resolve the motion of the glottis through an oscillation. Furthermore frame rate, and spatial resolution impact the cost of an HSV system. Typically, higher frame rates and higher spatial resolutions correspond to more costly systems. Additionally, some high speed cameras have the ability to increase frame rate at the cost of reduced spatial resolution and vice-versa.

Frame Rate

To provide some guidance on the effects of HSV frame rate, a range of likely frame rates were simulated. In clinical settings, frame rates range from 2000 FPS to 5000 FPS, while in research settings, some researchers may record at frame rates up to 20 000 FPS [15]. In order to evaluate this range of frame rates, an effective frame rate of 84 000 FPS was chosen as the highest temporal resolution. This provides a baseline case that is over 4 times faster than cameras employed in clinical settings. This corresponds to an actual recording frame rate of 30 FPS in the experimental facility. To examine the effect of lower frame rates, the source video at the highest frame rate was downsampled by a factor d_{temporal}. To downsample the frame rate by a factor of d_{temporal}, only every d_{temporal}'th frame was used, with all others being dropped. This aliases the resulting video in time. The highest temporal downsampling chosen was 32, resulting in a worst case frame rate of 2625 FPS.

Spatial Resolution

High-speed cameras are available with varying spatial resolutions, with larger spatial resolutions requiring larger storage capacity per image. For the purposes of inference and glottal width extraction however, it is not the absolute resolution of the camera that matters but the the number of pixels spanning the glottis during the imaging (magnification factor). A low resolution camera with the glottis occupying the entire field of view may be more detailed than a high-resolution camera in which the glottis occupies a small portion of the image. Thus it is important to vary the ratio of number of pixels spanning a characteristic length of the glottis. In physiological scenarios, this ratio is estimated to vary from about $50\,\text{px}\,\text{mm}^{-1}$ to $100\,\text{px}\,\text{mm}^{-1}$. This is based on a typical glottal length of 10 mm, and typical resolutions of high-speed cameras ranging from 1000 square pixels to 200 square pixels. It is assumed that clinical practitioners aim to direct the camera such that the VFs occupy the majority of the image, and therefore the glottal length approximately spans the image. For this study, the ratio of glottal length in pixels to millimetres was approximately $750\,\text{px}\,\text{mm}^{-1}$ in the source video (accounting for the $7.5\times$ scaling on the VF geometry, this corresponds to an actual scaling of $100\,\text{px}\,\text{mm}^{-1}$ on physical distances in the video). This is $7.5\times$ better than a clinical recording with the best spatial resolution. Note that the exact value varied slightly depending on the distance of the camera to the VFs, which could only be controlled approximately.

To investigate this range of spatial resolutions, the reference video was spatially downsampled. To downsample the resolution by a factor of d_{spatial}, the image was grouped into downsampled pixels, each downsampled pixel containing a square of d_{spatial} by d_{spatial} of the

source image's pixels. The intensity of the downsampled pixel was the average intensity of all d_{spatial}^2 pixels contained in it. As a result, the spatial resolution of the downsampled imaged becomes $1920/d_{\text{spatial}}$ by $1080/d_{\text{spatial}}$ pixels. To span the range of clinically encountered spatial resolutions, the largest downsampling factor was 16, corresponding to a spatial resolution of $47\,\text{px}\,\text{mm}^{-1}$.

Viewing Angle

Non-orthogonal viewing angles can lead to under-estimation of the extracted glottal width due to parallax error. Given the practical issues associated with clinical imaging, including the arytenoid hooding (which obstructs the view of the glottis), varying vocal tract geometries, and patient tolerance to procedures, to name a few, clinicians are rarely concerned with obtaining an orthogonal view. As such, we aim to determine the sensitivity of inferred parameters to off-axis viewing angles. While in real HSV, the angle of view can be offset in 3D, the experiment in this study used 2D VFs. Thus the angle of view was adjusted only in one plane (the coronal plane). Specifically, viewing angles ranged from a perfectly aligned view (angular offset of $\alpha = 0°$) to a view offset by $\alpha = 10°$, which we forecast to be an extreme off-axis viewing angle encountered in clinical studies. The angle of view, α, is shown in Figure 3.9.

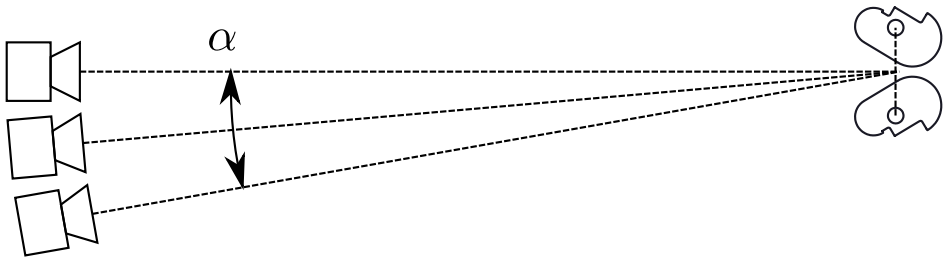


Figure 3.9: The angle of view was controlled by rotating the camera about the intersection of the glottal midline and translational axis of the VFs.

Summarizing the previous discussions, the following parameters were chosen as imaging parameters for the experiment. A total of 5 angles, 4 effective frame rates, and 4 effective physiological resolutions were tested. These are given in Table 3.1. Every combination of these parameters was used to generate simulated HSV videos of the VFs.

Table 3.1: The angles of view, frame rates, and spatial resolutions tested in the experiment.

Angle of View [°]	0	2.5	5.0	7.5	10.0
Frame Rate [FPS]	84 000	10 500	5250	2625	
Resolution [px mm^{-1}]	750	188	94	47	

3.2.2 Body-Cover Model Parameters

For each of the previous video parameters, a simulated video was taken of the experimental VFs, controlled by a BCM with set parameters. As mentioned in Section 2.4.2, the symmetric BCM is governed by a total of 11 parameters (given in Table 2.1). Four different reference BCM parameter sets were tested. These are specified in Table 3.2.

Table 3.2: Parameters used for the experimental BCM. 11 parameters are needed to completely specify the symmetric BCM model. The parameters, a_{ct}, a_{ta}, and a_{lc} are muscle activation parameters and were described in Section 2.4.1. The initial velocities were set to be 0. The initial positions were set at the rest positions of the springs in all cases. The supraglottal pressure, P_{sup}, was set to be 0 Pa in all cases.

Case	Parameters
A	$P_{\mathrm{sub}} = 1800\,\mathrm{Pa}$, $a_{\mathrm{ct}} = 0.20$, $a_{\mathrm{ta}} = 0$, $a_{\mathrm{lc}} = 0.5$
B	$P_{\mathrm{sub}} = 1800\,\mathrm{Pa}$, $a_{\mathrm{ct}} = 0.15$, $a_{\mathrm{ta}} = 0$, $a_{\mathrm{lc}} = 0.5$
C	$P_{\mathrm{sub}} = 2000\,\mathrm{Pa}$, $a_{\mathrm{ct}} = 0.20$, $a_{\mathrm{ta}} = 0$, $a_{\mathrm{lc}} = 0.5$
D	$P_{\mathrm{sub}} = 2000\,\mathrm{Pa}$, $a_{\mathrm{ct}} = 0.15$, $a_{\mathrm{ta}} = 0$, $a_{\mathrm{lc}} = 0.5$

These parameters were chosen as a result of their stable oscillations, and relatively large amplitude of motion. Specifically, it was found that the parameter sets displayed no cycle to cycle differences in behaviour once a steady state was reached, as seen in Figure 3.10. Furthermore, the amplitude of motion was on the order of 1 mm in physiological units. This amplitude is in agreement with physiological observations [48]. As discussed in Section 2.4.1, the subglottal pressure provides the driving forces on the VFs. Thus in general, higher subglottal pressures result in larger glottal displacements. Also discussed, was that a_{ct} is related to tension of the VFs, which influences the frequency of oscillation. Thus the combination of parameters tested above were related to different frequencies of oscillation and amplitudes of vibration. The physiological glottal widths calculated from the 4 simulations are shown in Figure 3.10.

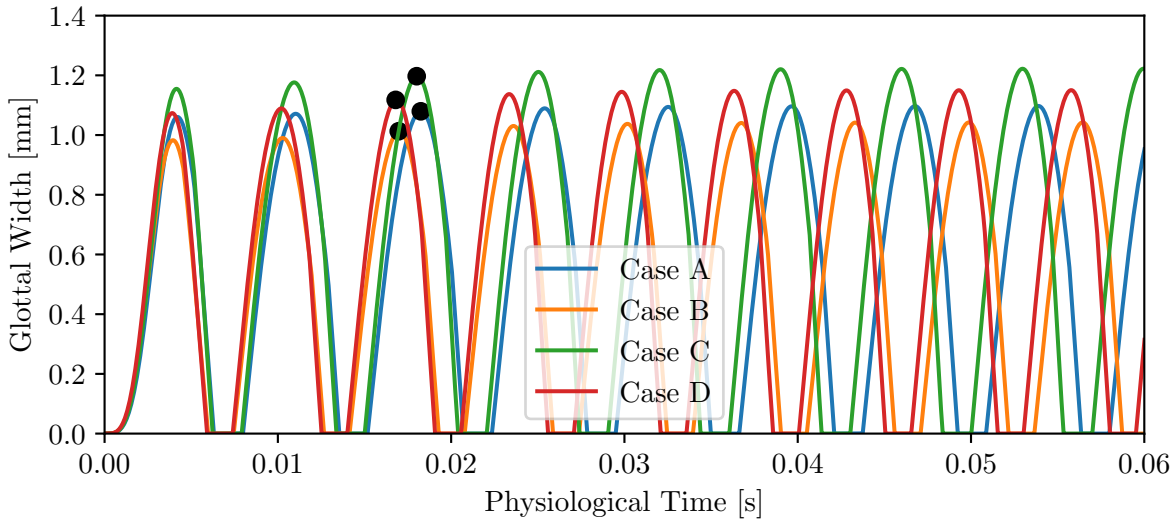


Figure 3.10: The glottal widths computed from each of the 4 parameter sets tested are shown. It is seen that an increased P_{sub} corresponds to an increased amplitude of the glottal width, while an increased a_{ct} corresponds to an increased frequency of oscillation. The black dots illustrate the point at which steady state behaviour was considered to occur. Only video recorded past this point was used in the inference procedure.

As seen in Figure 3.10, P_{sub} and a_{ct} tend to increase the amplitude of motion, and frequency of oscillation respectively. The initial motion is transient, but steady state oscillations are achieved by approximately the 2^{nd} oscillation, for all models. This can be further shown by plotting the first mode of oscillation, and amplitude of vibration over a number of P_{sub} and a_{ct} values, as shown in Figure 3.11.

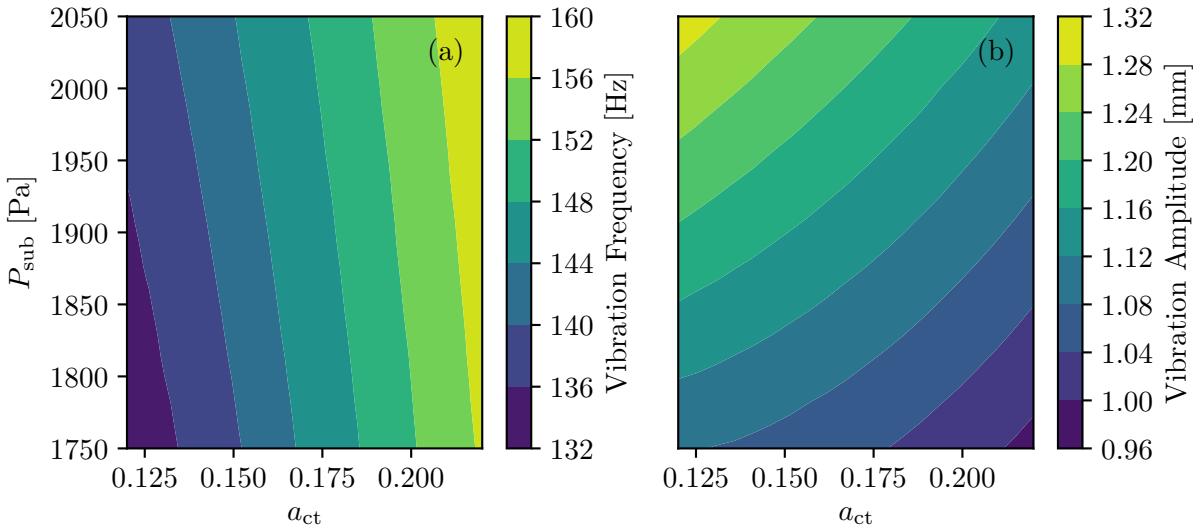


Figure 3.11: Contour plots of (a) the first mode of vibration, and (b) the amplitude of vibration for the BCM over a grid of a_{ct} and P_{sub} parameters.

From Figure 3.11 the effects of a_{ct} and P_{sub} on the glottal width behaviour can be seen. In terms of the frequency of vibration, shown in Figure 3.11(a), it is seen that increasing a_{ct} corresponds to increasing the frequency of oscillation. This makes sense since the cricothyroid serves to tense the VFs, which increases the spring constants in the BCM, and thus increases the frequency of oscillation. There is also a slight effect of P_{sub} on the frequency of oscillation; as P_{sub} increases, the frequency slightly increases. In terms of the amplitude of vibration, shown in Figure 3.11(b), it is seen that a_{ct} and P_{sub} both have an effect on the amplitude of vibration. As a_{ct} increases, the amplitude of vibration tends to decrease, and as P_{sub} increases the amplitude of vibration tends to increase. This makes physiological sense; cricothyroid activation stiffens the VFs so that they have lower displacements for the same forcing glottal pressures; subglottal pressure increases the driving forces on the VFs, and as a result the amplitude of vibration increases.

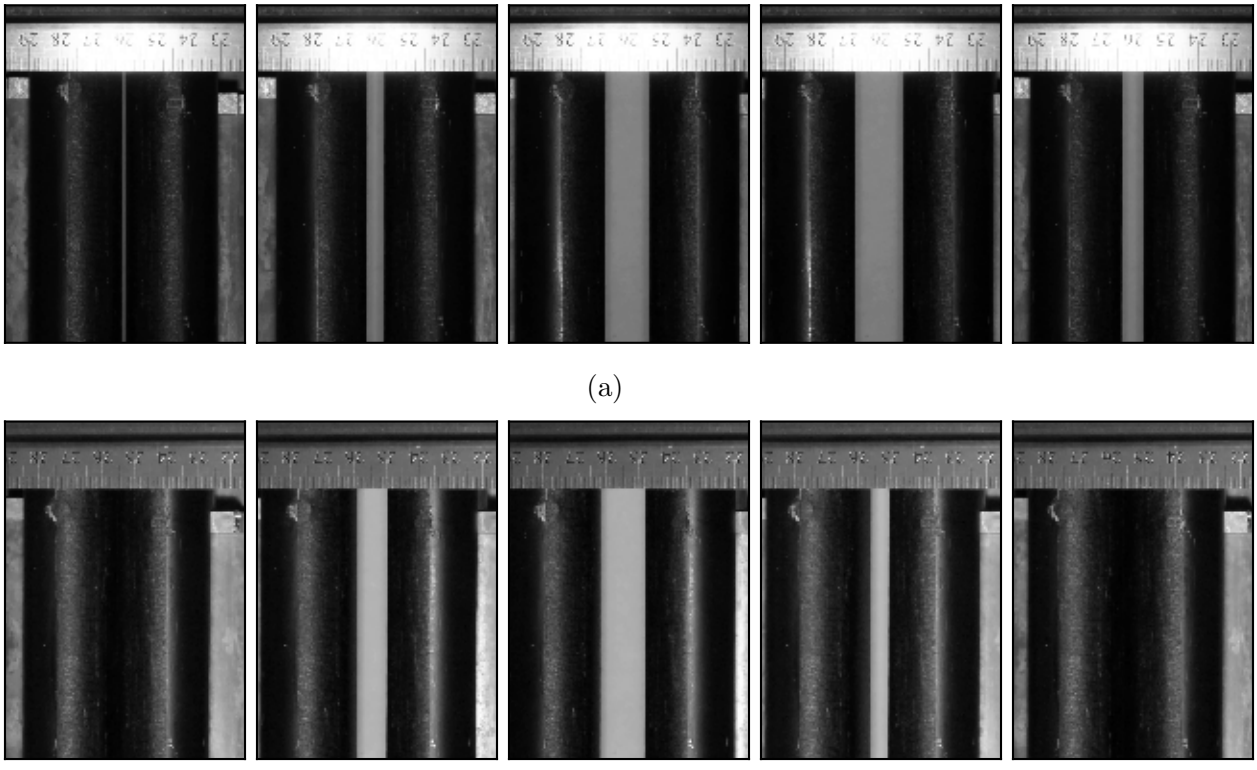
3.3 Post-Processing and Bayesian Inference

For each video, an edge detection procedure was applied in order to find the separation between the left and right VF edges. After conversion from pixels to physical distances,

Bayesian inference was used to estimate the a_{ct} and P_{sub} parameters of the BCM. Details of this process are given in the next two sections.

3.3.1 Edge Detection

There are many approaches to edge detection, such as through computing maxima of the intensity gradient of the image, computing 0 crossings of the Laplacian of the image, fitting of idealized edge intensity curves, or using image moments [27]. Due to the 2D nature of the artificial VFs in this experiment (and orientation of the camera), the edges of the VFs formed a single line along the columns of the image. A series of frames showing the VF edges is illustrated in Figure 3.12.



(a)

(b)

Figure 3.12: Example frames of the driven VFs throughout one cycle of the glottal motion. (a) A view from $\alpha = 0°$; (b) a view from $\alpha = 10°$.

In this work, a zero-crossing Laplacian edge detector was used. This edge detection method is based on the observation that large intensity changes occur over edges. Thus the maximum intensity change, which corresponds to the edge, should be located where the Laplacian (a 2^{nd} order spatial derivative) of the image intensity equals zero [45].

Computation of a glottal width vector, involved finding the glottal width at every frame of a video. To compute a single estimate for the glottal width for each frame, first the upper portion of the frame with the ruler in view was cropped, so that only the VFs occupied the image. In addition, columns of the video were manually cropped to remove any extraneous objects in the image. This resulted in a frame with 847 rows, for all reference videos. Next, zero-crossings of the Laplacian at each row of the cropped frame were found. At every row this results in finding 4 pixel locations, 2 pixel locations are located on each side of an edge for each of the 2 edges. At these pixel locations, the Laplacian transitions from positive to negative or vice-versa. To determine a sub-pixel location of the edge, a linear interpolation was then performed between the two detected pixels. This is illustrated in Figure 3.13.

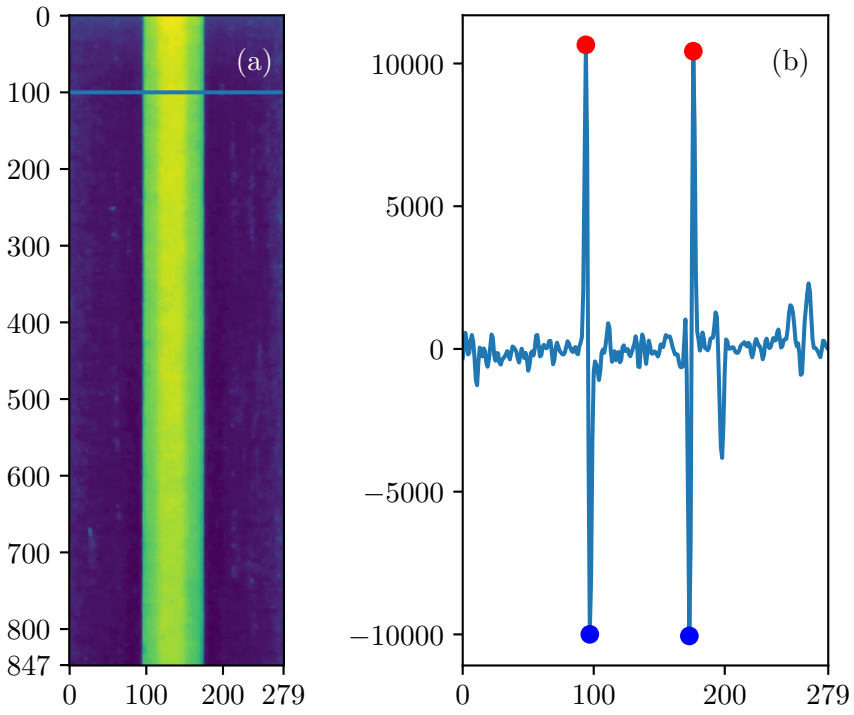


Figure 3.13: The Laplacian plotted along a row of the image. The edge is detected by computing the transition where the Laplacian changes from positive (plotted in red) to negative (plotted in blue).

Figure 3.13 illustrates the Laplacian plotted along a single row of the image. Where the Laplacian transitions from positive to negative, an edge is detected. A linear interpolation of the Laplacian was then performed between the positive and negative Laplacian pixel locations, to determine the location between these two pixels at which the Laplacian is equal to 0, and therefore where the edge is located (this has sub-pixel precision). This was repeated for all rows to produce a sub-pixel accurate glottal width for each row of the frame. The glottal width for the frame was then taken as the average glottal width measured from all rows. Finally, a physical width measurement was generated by multiplying the sub-pixel glottal widths by a calibration constant. This was provided by a ruler in view of each video (see Figure 3.12 for an example). This results in a detected glottal width for each frame of the video, in physical units. An example of this is shown in Figure 3.14 for videos obtained from multiple angles of view.

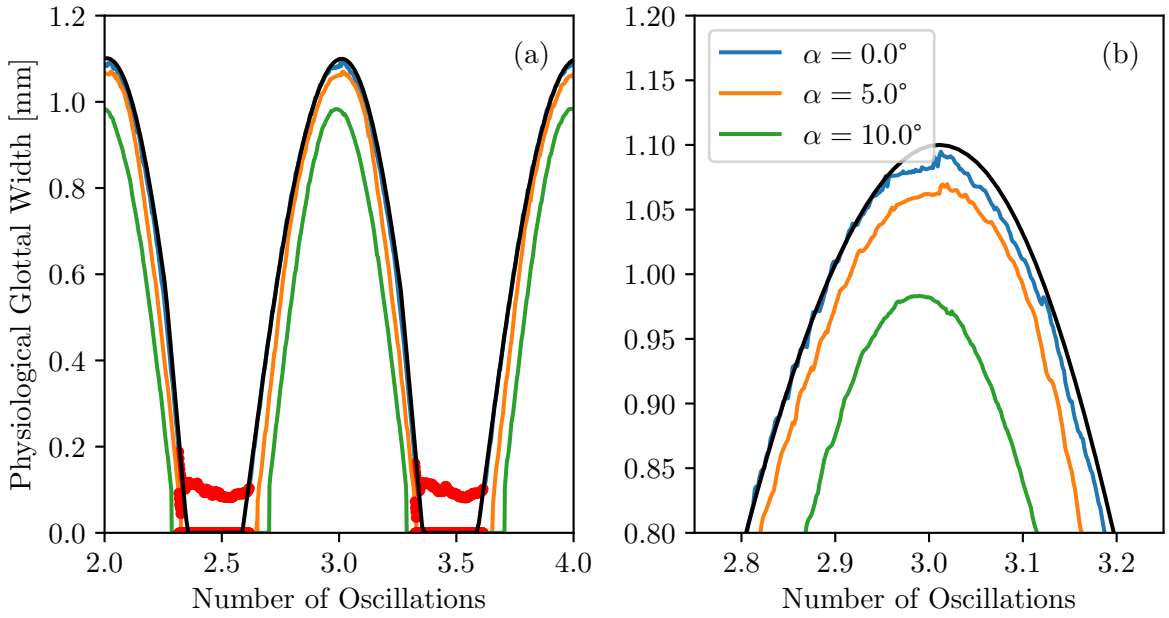


Figure 3.14: (a) The glottal width measured through the edge detection method. Points highlighted in red indicate collision of the BCM driving the VFs occurred. At these points, the motion of the VFs deviates from the BCM, which has overlapping masses. These points are not used in the inference procedure. (b) A closeup of the detected glottal widths at the peak of the oscillation.

In Figure 3.14 the red dots indicate that the collision algorithm stopped the VFs from following the BCM positions. As described earlier, the collision algorithm maintains a constant gap between the experimental VFs while the underlying BCM masses are overlapping. It can be seen in Figure 3.14 that the red dots do show an approximately constant glottal width, except for a slight initial drop at the beginning of collision. This could be the result of a variety of factors, such as extraneous motion commands of the stepper motors. As mentioned previously, the motion profiles of the stepper motors were implemented using undocumented procedures from the manufacturer. These points have little effect on the inference procedure, since they are ignored. At the peak of the oscillation, the measured and commanded glottal widths seem to agree reasonably well. Deviations could be the result of motion command errors, or system calibration errors.

Edge Detection Variance

As the resolution is downsampled, the edge detection algorithm should become less accurate due to the decreased spatial resolution. To measure this trend of decreasing accuracy with decreasing spatial resolution, the variance in the glottal width was computed at different spatial resolutions

$$\sigma^2 = \sum_{i}^{N_{\text{row}}} (w_i - \overline{w})^2$$

where w_i is the glottal width measured at row i of the image, and $\overline{w}$ is the mean of w_i computed over all rows. The glottal width variance was computed for each downsampled resolution, then normalized by the glottal width variance at the reference resolution $d_{\text{spatial}} = 1$. The normalized glottal width variance computed is shown for Case A in Figure 3.15 for each of the angles of view.

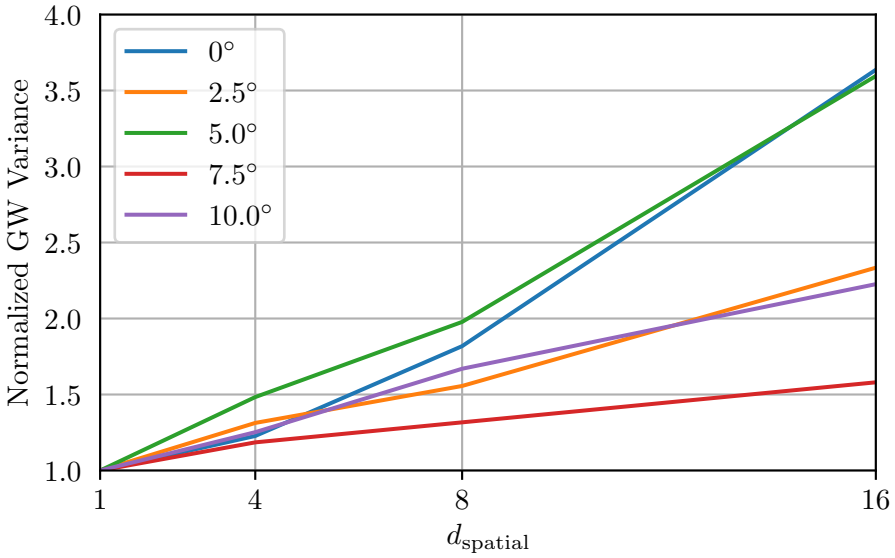


Figure 3.15: Row-wise variance of the the glottal width (GW) at varying spatial resolutions, when imaging Case A. There is no discernible trend with the angle of view. Note that the normalized glottal width variance is 1 at $d_{\text{spatial}} = 1$ by definition of the normalization.

The trends in edge detection variance were used for scaling the measurement error from the glottal width. However, as seen in Figure 3.15, it was observed that these trends displayed little consistency over varying imaging parameters. While the variance in the

detected glottal width does increase as the resolution is downsampled, the increases do not have any consistent form. Furthermore, increasing angles of view do not appear to have a consistent effect on the edge detection variance trends with decreasing spatial resolution. For example, the rate of increase of glottal width variance between the 0.0° and 5.0° angles of view are similar, while the 7.5° case has the lowest rate of increase of glottal width variance. These inconsistent trends are likely due to uncontrolled parameters between videos. In particular, it is known that edge detection depends on the level of noise in the image, which in turn depends on the illumination of the VFs. While the same light source was used to illuminate all cases, the actual noise characteristics of the edges are difficult to control using a single light source. Furthermore, the camera automatically makes adjustments based on lighting in the scene which changes the noise characteristics further. This can be seen clearly in Figure 3.12, where the videos obtained at $\alpha = 0°$ and $\alpha = 10°$ clearly have different lighting characteristics. This is a limitation of the study.

To determine a characteristic scaling on the increase in glottal width variance with respect to spatial downsampling, a linear fit through all the trends was used. This fit was computed to be $1 + 0.094842 \cdot (d_{\text{spatial}} - 1)$, as shown in Figure 3.16.

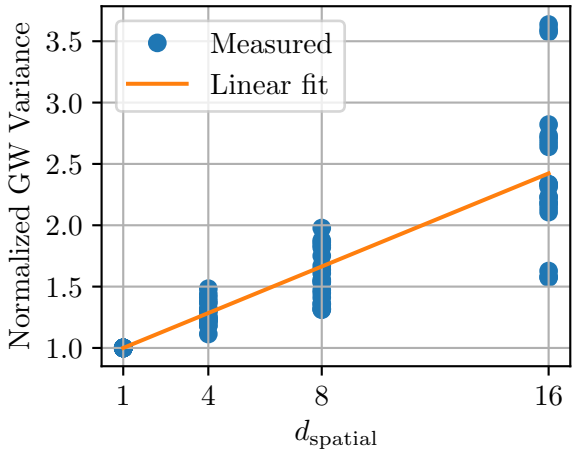


Figure 3.16: Row-wise variance of the the glottal width (GW) at varying spatial resolutions, for all videos. There is no discernible trend with the angle of view, thus a linear fit through all points was chosen.

Each point in Figure 3.16 represents a normalized glottal width variance computed from a video of any of the four BCM cases (A, B, C, or D) at any angle of view. This

corresponds to 20 normalized glottal width variances (5 angles of view for each of 4 BCM cases) for each spatial downsampling factor, d_{spatial}.

3.3.2 Bayesian Inference

With the glottal width extracted from the edge detection method, the last step is to infer parameters of the BCM through Bayesian inference. For this study, the inference model chosen was the same as the model used to generate the motion, the BCM (described in Section 2.4). This setup in inference problems, in which the model perfectly explains the observed data, is known as an inverse crime [36, Section 1.2]. Inverse crimes are problematic because they result in inverse estimates that are over-optimistic. In this study, there are a number of details that prevent the BCM from perfectly explaining the observed glottal width. These include: perspective errors between the inferior and superior edges of the glottis, differing integration time steps between the generating BCM (1/350 ms) and inference BCM (1/14 000 ms), as well as errors in the motion control system of the VFs. These errors all contribute to uncertainty in the glottal width measurement. If the absolute parameter values were desired from the inference, then these errors should be modelled. However, in this study it was the effect of the imaging system errors that was desired, and so these errors were not modelled.

For this study, two parameters were inferred from the extracted glottal width (from a total of 11 parameters): a_{ct} and P_{sub}. The remaining parameters were considered as known, and were obtained from the generating BCM. To infer these 2 parameters, first the generated glottal width waveform was cross-correlated with the observed glottal width waveform, in order to synchronize the time of the cRIO with the time of the video. Next the first 2 transient oscillations of the observed glottal width, were cropped until the peak width of the first steady state oscillation. The starting location for each of Cases A, B, C, and D is shown in Figure 3.10. At the peak glottal widths, the positions $(x_{\text{u}}, x_{\text{l}}, x_{\text{b}})$, and velocities $(\dot{x}_{\text{u}}, \dot{x}_{\text{l}}, \dot{x}_{\text{b}})$, were taken from the cRIO simulation and were considered as known initial parameters for the inference BCM. In addition, the muscle activations were considered as known, $a_{\text{ta}} = 0$, $a_{\text{lc}} = 0.5$. Finally the supraglottal pressure was also considered known $P_{\text{sup}} = 0\,\text{Pa}$. Thus 9 of the 11 parameters specifying the symmetric BCM were considered known, leaving a_{ct} and P_{sub} to be inferred.

From the set of initial parameters, it remains to specify the forward model (as mentioned earlier, this is a BCM). The forward model, $\boldsymbol{y}_{\text{model}} = F(a_{\text{ct}}, P_{\text{sub}})$, comes from solution of the BCM, and subsequent calculation of the glottal width from the solution of the BCM, given the parameters described previously. This starts with a numerical solution of the

BCM for a given a_{ct}, and P_{sub} (plus known initial parameters) which produces a series of displacements at discrete times, as shown in equation (3.2)

$$\begin{bmatrix} x_{\mathrm{u},i} \\ x_{\mathrm{l},i} \\ x_{\mathrm{b},i} \end{bmatrix} = \begin{bmatrix} x_{\mathrm{u}}(i \cdot dt) \\ x_{\mathrm{l}}(i \cdot dt) \\ x_{\mathrm{b}}(i \cdot dt) \end{bmatrix}. \tag{3.2}$$

Since only two parameters are estimated, all remaining BCM model parameters were considered known, and taken from the reference BCM simulation. Solution of the BCM was performed using an RK4 method and a time step of $1/14\,000$ ms. It was found that this time step was a good balance between computational cost and numerical accuracy.

The remaining step in the forward model is computation of the glottal width. This was taken to be the minimum of x_{u} and x_{l} at each time step. Application of this to the BCM displacements results in a series of glottal widths at every solution time of the model, as shown in equation (3.3)

$$w_i = \max(0, 2 \cdot \min(x_{\mathrm{u},i}, x_{\mathrm{l},i})). \tag{3.3}$$

The fact that the initial conditions of the BCM were considered known, is a restrictive condition since the positions and velocities of the BCM masses will vary throughout an oscillation. In a clinical setting, imaging is typically done while the VFs are in steady state oscillation. Obviously, the positions and velocities of an equivalent reduced order model cannot be known apriori, based on HSV of real VFs. In this study, the reference motion generated on the cRIO, made this information available, which reduces the number of parameters needed to be estimated. This is a limitation of the study; in future studies the initial positions and velocities of the masses should be inferred.

Prior

The prior is the expected distribution of model parameters apriori, that is before any information about the system is obtained. In this problem, uninformative priors were used, since it was assumed that no apriori information about a_{ct} or P_{sub} exists. Thus the prior π_{prior} is simply a constant. This makes the likelihood directly proportional to the posterior.

Note that priors would play little to no role in this problem, due to the uncertainty extents of the estimated posterior (see Figure 3.17). As seen in Figure 3.17, the range in likely P_{sub} is on the order of 10 Pa, while the range in likely a_{ct} is on the order of 0.001. Unless the prior varies the probability over these differences in parameters, the prior can be approximated as a constant regardless. A prior would play larger roles if the uncertainty bounds of the posterior were larger.

Likelihood

To compute the likelihood, an independent identically distributed additive Gaussian noise model was considered to model the error in the measured glottal width (see Section 2.5.2 for details of the additive Gaussian noise model). The likelihood can thus be written as

$$\begin{aligned}\pi_{\text{likelihood}}(\boldsymbol{w} \mid a_{\text{ct}}, P_{\text{sub}}) &= \pi_{\text{likelihood}}(\boldsymbol{w} \mid \boldsymbol{w}_{\text{model}}) \\ &= \frac{1}{(2\pi\sigma_{\text{obs}}^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma_{\text{obs}}^2} \left\|\boldsymbol{w} - \boldsymbol{w}_{\text{model}}\right\|^2\right), \end{aligned} \tag{3.4}$$

where the vector $\boldsymbol{y}$ is a vector containing glottal width observations for each frame of the high-speed video, $\boldsymbol{y}_{\text{model}}(a_{\text{ct}}, P_{\text{sub}})$, called the forward model, is the set of model predicted glottal widths at corresponding time instances with frames of the video, and σ_{obs} is standard deviation of the measured glottal width for each frame. The forward model is specified by equation (3.2) and (3.3), and where equation (3.2), requires numerical solution of the BCM, as specified in Section 2.4.

For this work, a standard deviation of 0.12 mm was chosen as the uncertainty in measuring a single glottal width. This choice of standard deviation was not based on variance in the edge detection procedure for each case, rather a uniform value was chosen so that uncertainty in the posterior could be compared across all cases. This uniform value is about 10 % of the maximum glottal width observed. In the case of resolution downsampling, a multiplicative factor was applied to the observation noise, based on the increase in glottal width variance with decreased resolution, $\sigma_{\text{obs}} = 0.12\sqrt{1 + 0.094842 \cdot (d_{\text{spatial}} - 1)}$.

Posterior and Evidence

There are many approaches to compute the posterior. The most straightforward is simply to directly evaluate the posterior distribution. This involves evaluating Bayes' formula over a grid of the parameters (in this case different values of a_{ct} and P_{sub}). For a low number of parameters this approach is computationally feasible. Due to the relatively low number of dimensions in this problem, a direct computation of the posterior was used. First the unnormalized posterior

$$\pi_{\text{posterior}}(a_{\text{ct}}, P_{\text{sub}} \mid \boldsymbol{w}) \propto \pi_{\text{likelihood}}(\boldsymbol{w} \mid a_{\text{ct}}, P_{\text{sub}}) \pi_{\text{prior}}(a_{\text{ct}}, P_{\text{sub}}) \tag{3.5}$$

$$\propto \frac{1}{(2\pi\sigma_{\text{obs}}^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma_{\text{obs}}^2} \left\|\boldsymbol{w} - \boldsymbol{w}_{model}\right\|^2\right), \tag{3.6}$$

is computed over a grid.

Calculation of the evidence then involves normalizing this unnormalized posterior. This is done by computing the area integral of the likelihood over the parameter grid (since an uninformative prior was used in this work)

$$\pi_{\text{posterior}}(a_{\text{ct}}, P_{\text{sub}} \mid \boldsymbol{w}) = \frac{\pi_{\text{likelihood}}(\boldsymbol{w} \mid a_{\text{ct}}, P_{\text{sub}})}{\int_{\Omega} \pi_{\text{likelihood}}(\boldsymbol{w} \mid a_{\text{ct}}, P_{\text{sub}}) \mathrm{d}a_{\text{ct}} \mathrm{d}P_{\text{sub}}}, \tag{3.7}$$

where Ω is the domain of the parameter space. In theory this implies integrating over an infinite span of P_{sub}, due to the uninformative prior, and from 0 to 1 for a_{ct}. In practice however, the posterior uncertainty bounds spanned small ranges of the parameter space, on the order of 10 Pa in P_{sub}, and on the order of 0.001 for a_{ct}. Thus the grid, over which the posterior was computed, was chosen manually for each inference case in order to capture the posterior probability mass. Furthermore, the prior used for this problem is an uninformative prior. This means the likelihood is directly proportional to the posterior. While this makes the posterior and normalized likelihood equivalent, the term posterior will be used to refer to the probability distribution of the estimated parameters. An example of some computed posteriors are shown in Figure 3.17.

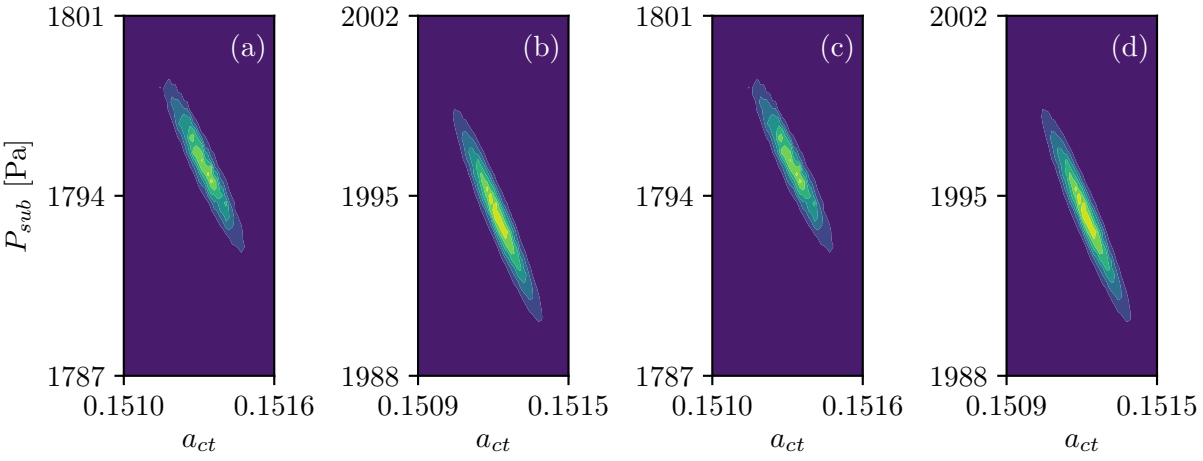


Figure 3.17: The posteriors for a 0° viewing angle, and the highest spatial resolution and frame rate are shown. From left to right, they illustrate Cases $(P_{\text{sub}}, a_{\text{ct}})$ = (a) (1800 Pa, 0.15), (b) (2000 Pa, 0.15), (c) (1800 Pa, 0.20), and (d) (2000 Pa, 0.20). Brighter colours indicate larger probability densities.

The posteriors shown in Figure 3.17 show the probability densities of different parameter sets. The dark blue colour, indicates the minimum probability density (very nearly 0), while

the lightest green colour indicates the maximum probability density. It can be seen that the range of likely parameter combinations occupies a small region of the parameter space. For example in Figure 3.17(a), the range of probable P_{sub} spans from approximately 1791 Pa to 1794 Pa. On first examination of the posterior, the densities appear to be 2D normal distributions, evidenced by the characteristic elliptical shape. Closer examination reveals that the iso-contours of probability density are not smooth ellipses, but rather exhibit some irregularties. This is the product of numerical integration errors. These irregularities and further details of the posteriors estimates, are discussed in Chapter 4.

Chapter 4

Results and Discussion

The results of the study consist of inferred parameters from simulated HSV videos for each of the 4 different numerically driven VF model cases introduced in Chapter 3. Each simulated HSV video is denoted by its spatial downsampling (d_{spatial}), temporal downsampling (d_{temporal}), and angle of view (α), relative to a reference case where $\alpha = 0°$, $d_{\text{temporal}} = 1$, and $d_{\text{spatial}} = 1$ (see Table 3.1 for the tested parameters). The four experimental VF motion cases have driving BCM model parameters summarized in Table 3.2 and will be referred to as Cases A, B, C, and D.

This chapter begins with a discussion of effect of the duration of the high-speed video on the inferred parameters. Using this, an ideal video length is chosen to explore the effects of changing α, d_{spatial}, and d_{temporal} independently, on the posterior uncertainty and MAP estimates for each of the 4 BCM simulations. Next the effect of combined changes in α, d_{spatial}, and d_{temporal} on the posterior uncertainty are investigated. Finally some considerations and guidance for implementing Bayesian inference on reduced order VF models using HSV are given.

4.1 Effect of Glottal Width Time Series Length

The posterior is dependent on the number of measurements in the glottal width time series. In general, the more points, the more information, and therefore the less the uncertainty on the estimated parameters. There are two ways to vary the number of measurements in HSV: either the frame rate can be varied with a fixed video length, or the video length can be varied with a fixed frame rate. Here we investigate the effect of reducing the video

length at a fixed frame rate ($d_{\text{temporal}} = 1$, $d_{\text{spatial}} = 1$, $\alpha = 0°$) on inferred parameters for Case A. Six different video durations, illustrated in Figure 4.1, are explored. Specifically, the HSV is divided into 20 ms segments, resulting in a glottal area waveforms comprising approximately 3, 6, 8, 11, 14, and 17 oscillation cycles.

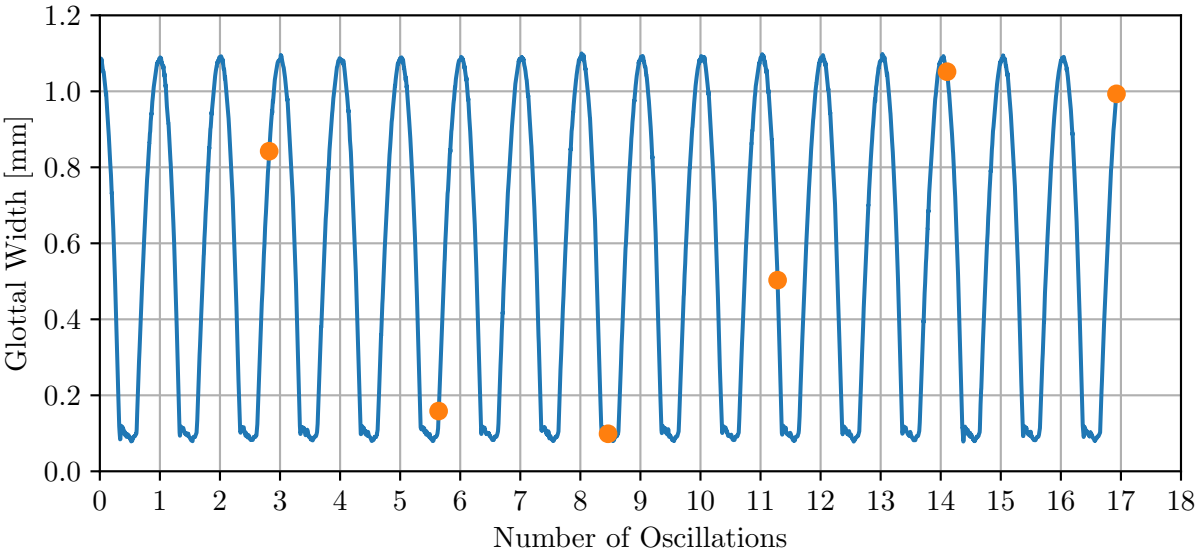


Figure 4.1: Glottal width extracted from HSV for Case A with $\alpha = 0$, $d_{\text{spatial}} = 1$, and 4 $d_{\text{temporal}} = 1$. Orange dots indicate durations of truncated waveforms used for analysis. The points correspond to measurement durations of 20 ms, 40 ms, 60 ms, 80 ms, 100 ms, and 120 ms.

Figure 4.2 illustrates the effect of varying the duration of the time series on the posterior for three of the video lengths (shown in Figure 4.1, 3, 8, and 17 oscillations). For short time series comprising a small number of oscillations, the relatively small number of samples reduces the level of confidence in the estimate in comparison with the longer time series (given a fixed sampling rate), as indicated by the relatively large contours in Figure 4.2(a). As the signal duration increases, so does the number of measurements, and so the size of the posterior shrinks (see Figures 4.2(b) and (c)).

Additionally, it can be seen in Figure 4.2, that the posterior transitions from being positively correlated to negatively correlated as the observation time increases. The angle θ, indicating the orientation of the first eigenvector with respect to the positive a_{ct} axis, decreases as the number of captured oscillations increases, as shown in Figure 4.2. This is

due to the increasing importance of the oscillation frequency with longer time series. With a short time series, any frequency mismatches in the response of the model will not result in significant errors in the glottal width. Therefore for short time series, the estimated parameters will have greater certainty when they maintain an approximately constant amplitude of vibration. This means we expect parameters along the first eigenvector, to approximately preserve amplitudes of vibration. As seen in Figure 3.11(b), this occurs when a_{ct} decreases as P_{sub} decreases; this is what we observe at low observation times. As the length of the time series increases, small frequency mismatches will result in large beat errors analogous to the beat-frequency phenomena. These beat-errors will outweigh any errors in mismatched amplitude. Thus, the inferred parameters with greater certainty should produce approximately the same frequencies of vibration; in other words parameters along the first eigenvector correponds to models with similar frequencies of vibration. As seen in Figure 3.11(a), constant frequencies of oscillation are obtained when a_{ct} is increased while P_{sub} is decreased. Correspondingly, this is also the direction of the first eigenvector at longer observation times.

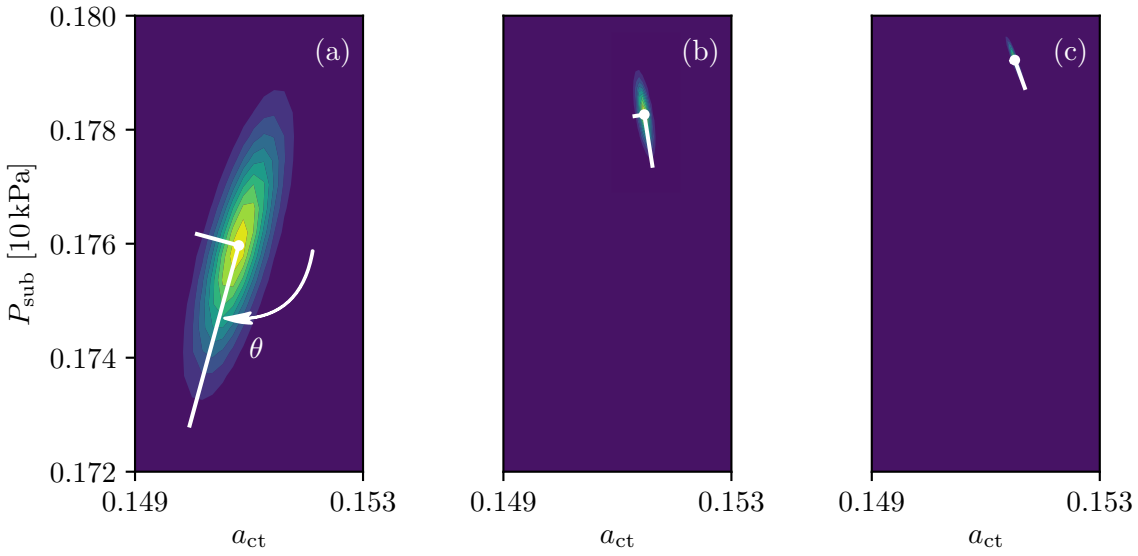


Figure 4.2: The effect of the time series duration on the posterior. Figures (a), (b), and (c) correspond to about 3, 8, and 17 oscillations (or 20, 60 ms, and 120 ms) of the glottis (see Figure 4.1). The angle, θ, is the angle from the a_{ct}-axis to the first eigenvector.

As the duration of observation increases, it can also be seen that the location of the MAP estimate is biased in Figure 4.2. This can be related back to the frequency-amplitude

behaviour of the model with respect to changes in the a_{ct} and P_{sub} parameters. Figure 3.11 demonstrated that amplitude is affected by a combination of a_{ct} and P_{sub}; lower a_{ct} and higher P_{sub} tend to produce larger amplitudes of vibration. As discussed previously, lower observation times correspond to matching amplitude behaviour. At higher observation times (Figures 3.11(a) and 3.11(b)), the frequency of vibration plays a more important role. Thus a_{ct} must increase to match the frequency. Since the amplitude of vibration is still the same, P_{sub} must also increase to balance the effect of increasing a_{ct}. Thus we see in Figure 3.11, that the MAP estimate increases in P_{sub} and a_{ct}. It can further be seen that at these longer observation durations (Figures 4.2(b) and (c)), the known parameters that generated the VF motion are not contained within the certainty bounds of the posterior. This means that there are errors that have not been accounted for. These modelling errors are discussed further in the next section.

Figure 4.3 shows that, over the range of signal lengths considered, θ monotonically decreases. The rate of decrease reduces as more oscillations are considered, however, suggesting that the trend will eventually reach an asymptote. It is noted that during clinical acquisition, HSV captures hundreds of cycles during a typical sustained vowel phonation.

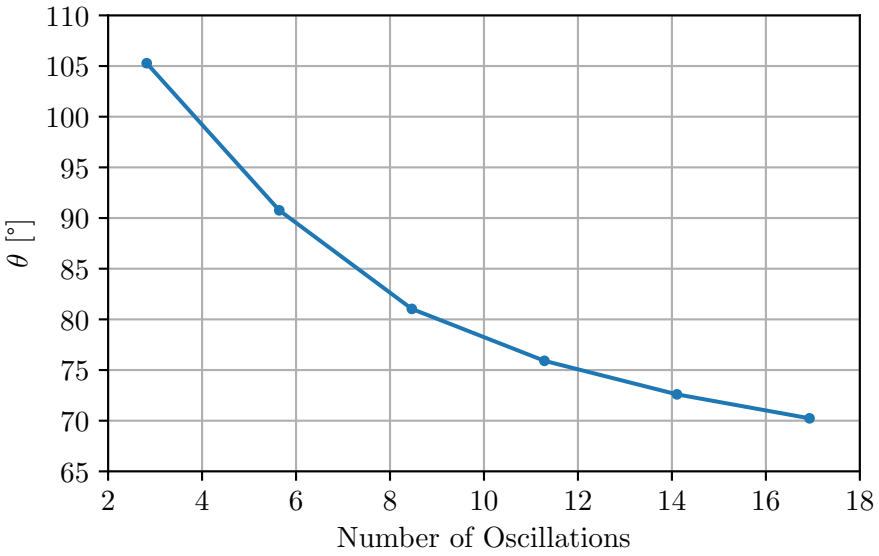


Figure 4.3: The angle of the first eigenvector of the covariance matrix, computed from the posteriors shown in Figure 4.2, for varying numbers of oscillations in the original time series.

For optimal estimates of stationary behaviour (that is, fixed model parameters, such as

a_{ct} and P_{sub}), it is clear that longer time series provide greater certainty on the estimated parameters. In addition, it is expected that when the direction of the correlation plateaus, only the size of the posterior will be affected, not its shape. This suggests that the duration of the video should be chosen to be sufficiently long, such that the direction of correlation remains constant. While ideally the number of oscillations investigated should exceed 18 or more (as shown in Figure 4.2), due to the fact that only 7 minutes of video were recorded in this study (about 17 oscillations of the glottis) larger numbers of oscillations could not be investigated. This initial duration of 7 minutes was chosen because it was initially believed that it would be sufficient for inference, as well as for experimental time constraints. For this reason, a duration of about 17 oscillations of the glottis was chosen.

4.2 Modelling Errors

Bayesian inference accounts for the measurement errors in a model to infer the parameters that generated a set of observations. Thus, if a model perfectly accounts for a measurement along with its errors, the Bayesian inference technique should predict the model parameters within statistical bounds. Closer observation of Figure 4.2(c) shows that the actual parameters of the model driving the VF motion in Case A, do not lie within the probable region of the posterior. This is reproduced in Figure 4.4(a) with the addition of a dot showing the ground truth values. This disparity means that there are errors in the modelling or in the measurement that have not been incorporated into the analysis.

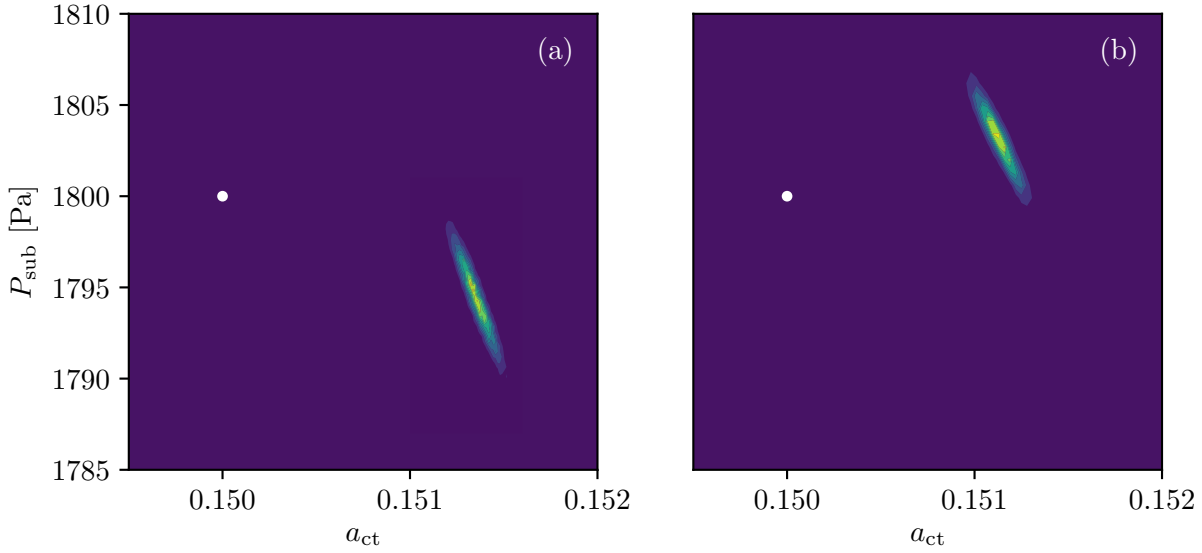


Figure 4.4: (a) posterior estimates for Case A using the full glottal area time series. The white dot indicates the ground truth parameter values used to drive the model. (b) modified posterior obtained by correcting the amplitude bias intrinsic to the experimental system.

Such modelling errors can come from a variety of sources. One source is the difference between the experimental VF motion and that prescribed by the reduced order model controlling the system. Specifically, the actual motion of the experimental VFs may not exactly match that of the governing BCM. This would create a bias between the BCM generated glottal width and the measured glottal width. Figure 4.5 plots the measured glottal gap versus the prescribed value over several oscillations, as well as the error between the two signals. Included as a thick green line is the time averaged error. Evident from the figure is that there is structure in the error; the error in the glottal width does not vary randomly about zero, but rather varies in a repeatable manner across oscillations. Specifically, the error appears to be lower when the glottal width is at a maximum and increases as the VFs enter or exit a minimum. The hypothesized reason behind this structure is that the motion control system needs only make small adjustments to the stepper motor positions to make the positioning accurate when the glottal width is near its maximum. In contrast, while the VFs are rapidly transitioning between a closed and maximally open phase, the control system must constantly adjust the stepper motor positions, since the desired position is, relatively rapidly, changing with time. The details of exactly how the motion control system generates positions are unknown since they are internal to the cRIO software.

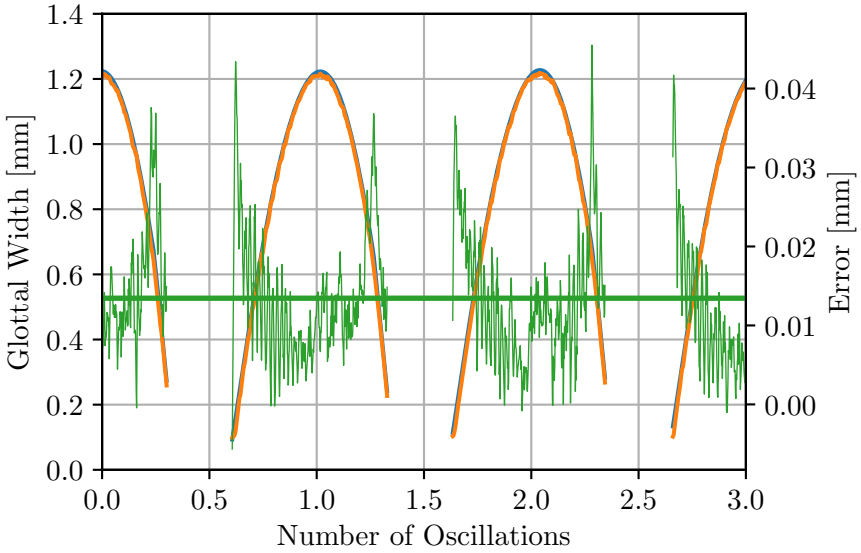


Figure 4.5: Plot of the measured (orange line) and prescribed (blue line) glottal width over several oscillations. The green lines indicate error in the measurement (thin line is instantaneous, while the thick line is the time average error).

There are further effects that can lead to the errors reported in Figure 4.5, such as how the image noise is affected by the light source throughout the motion. When the VFs are wide apart, the white background between the folds (see Figure 3.12) increases the contrast across the VF edge, which could improve the edge detection. In contrast, when the VFs are close, they can block the light source from reaching the background, which decreases the level of contrast. As a result, this should increase the level of error. These errors are a product of this specific experimental setup. As a result, the structure in the error here is not considered since it is not a direct product of the imaging system.

To demonstrate the effect of this bias error source on the posterior in Figure 4.4(a), the measured glottal width can be adjusted to account for the error. The posterior estimated from this corrected glottal width is shown in Figure 4.4(b). The effect of the bias correction is to shift the posterior towards higher supraglottal pressures, with minimal change to the a_{ct} values. As explained in Section 2.4.1, the P_{sub} and a_{ct} parameters primarily affect the amplitude of the glottal width and the frequency of vibration, respectively. Since the bias is corrected by modifying the amplitude of the glottal width, P_{sub} now appears to lie within the span of probable P_{sub} values. The fact that the estimated a_{ct} value still lies far outside the probability mass, is thus likely a result of a mismatch in frequency behaviour.

A mismatch in the frequency behaviour between the generating BCM and the fitting BCM is shown in Figure 4.6. This difference in behaviour was the result of the different integration times used to numerically solve the experimental BCM (1/350 ms) and the fitting BCM (1/14 000 ms). Figure 4.6 illustrates that the same set of parameters (a_{ct}, and P_{sub}) will result in different frequencies of oscillation, depending on the integration time step used to numerically solve a BCM. This leads to differences in the frequency behaviour between the two models. This explains why adjusting for the bias in amplitude of the glottal width, can ameliorate the error in the estimated P_{sub}, but not a_{ct}. Since a_{ct} primarily affects the frequency of oscillation, correcting the bias in this estimated value must correct the bias in frequency. Adjusting the glottal width, obviously cannot adjust the frequency of vibration. As a result, the known experimental a_{ct} does not lie within the certainty bounds of the estimated parameters when the glottal width is corrected.

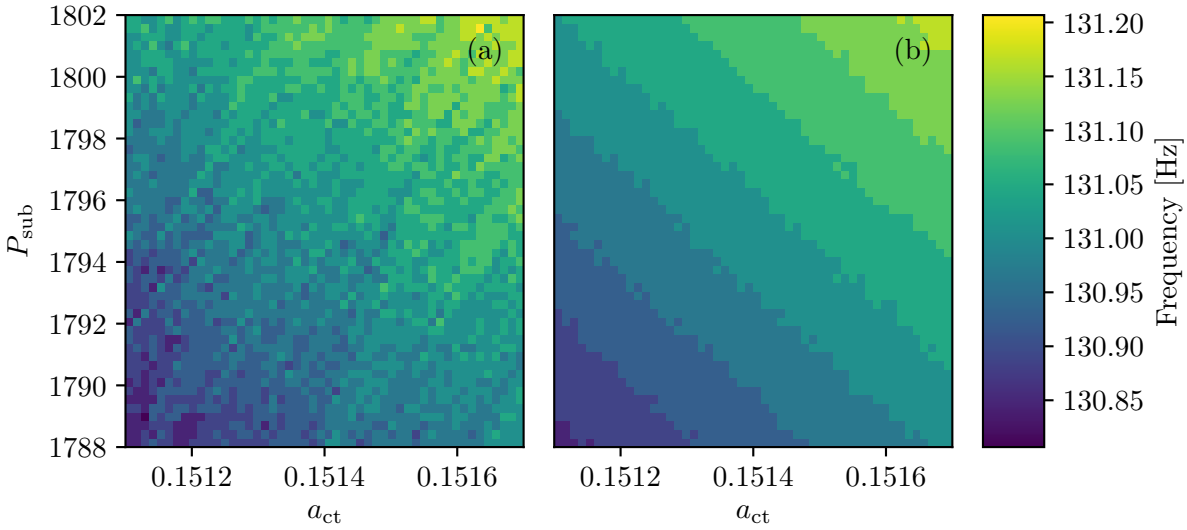


Figure 4.6: A glottal width time vector is computed at every pixel shown (corresponding to combinations of a_{ct} and P_{sub}) and a FFT is then performed on each of these glottal widths to calculate the first mode of vibration. This first mode (Hz) is then plotted as contour plot in (a) and (b) for each of the a_{ct} and P_{sub} parameters. (a) illustrates the first mode when the BCM is solved with an integration time step of 1/350 ms (corresponding to the experimental BCM controlling the VF motion). (b) illustrates the first mode when the BCM is solved with an integration time step of 1/14 000 ms (corresponding to the fitting BCM used in the inference procedure).

While accounting for all of the modelling errors causing the true parameter values to fall outside of the posteriors would correct the biases, this is not of concern in the present study of how HSV imaging parameters affect Bayesian inference. Furthermore, the bias in estimated parameters in this case is relatively small (deviations in a_{ct} are on the order of 0.001 while deviations in P_{sub} are on the order of 10 Pa). To characterize the effects of HSV imaging parameters on Bayesian inference, the remainder of this thesis deals only with relative changes in the estimated posteriors as a function of angle, frame rate, and spatial resolution relative to the reference video $(\alpha, d_{\mathrm{spatial}}, d_{\mathrm{temporal}}) = (0°, 1, 1)$.

4.3 Parameter Estimates and Uncertainty

The result of Bayesian inference is a posterior distribution, representing the probabilities of different parameter sets explaining the observed data. To characterize these posteriors, a point estimate of the posterior and a measure of uncertainty are used. For the point estimate, the MAP estimate is used. As described in Section 2.5.2 the MAP estimate is the parameter set that maximizes the posterior density, $(a_{\mathrm{ct}}, P_{\mathrm{sub}})_{\mathrm{MAP}} = \arg\max\left(\pi_{\mathrm{posterior}}(a_{\mathrm{ct}}, P_{\mathrm{sub}} \mid \boldsymbol{y})\right)$. Since the posterior densities have a single peak (see Figure 3.17 for some example posteriors), the MAP estimates serve as a good point estimate. To characterize the relative uncertainty, the square root of the determinant of the covariance matrix is used. The determinant gives a 'volume' of the posterior, while the square root then gives the geometrically averaged variance. This is a measure of the entire posterior uncertainty, rather than just marginal uncertainties. The uncertainty under given HSV imaging parameters is measured as:

$$u(\alpha, d_{\mathrm{spatial}}, d_{\mathrm{temporal}}) = \sqrt{\begin{vmatrix} \sigma^2_{a_{\mathrm{ct}},a_{\mathrm{ct}}} & \sigma^2_{a_{\mathrm{ct}},P_{\mathrm{sub}}} \\ \sigma^2_{a_{\mathrm{ct}},P_{\mathrm{sub}}} & \sigma^2_{P_{\mathrm{sub}},P_{\mathrm{sub}}} \end{vmatrix}}, \tag{4.1}$$

where $\sigma^2_{a_{\mathrm{ct}},a_{\mathrm{ct}}}$ is the variance of a_{ct}, $\sigma^2_{a_{\mathrm{ct}},P_{\mathrm{sub}}}$ is the covariance of a_{ct} and P_{sub}, and $\sigma^2_{P_{\mathrm{sub}},P_{\mathrm{sub}}}$ is the variance in P_{sub} (see Chapter 2.5.2 for details). The relative uncertainty is then given by

$$u_{\mathrm{rel}}(\alpha, d_{\mathrm{spatial}}, d_{\mathrm{temporal}}) = \frac{u(\alpha, d_{\mathrm{spatial}}, d_{\mathrm{temporal}})}{u(0°, 1, 1)}. \tag{4.2}$$

As explained in Section 4.2 we compare the relative changes in uncertainty relative to the reference case, which is accomplished here by normalizing the uncertainty measure by the reference case. This measure of uncertainty can be interpreted as a geometric average of the variances, and thus characterizes the variance of the whole posterior distribution,

rather than marginal distributions. This is used in place of marginal uncertainty measures, since it is a measure of uncertainty of both parameters, a_{ct} and P_{sub}, simultaneously. This makes it useful for the purpose of illustrating trends in the posterior uncertainty as a function of HSV imaging parameters. In the remainder of the section, the MAP estimate and relative uncertainties are used to explore the effects of changing angle of view (α), frame rate (d_{temporal}), and spatial resolution (d_{spatial}), relative to the reference video.

4.3.1 Effect of Camera Angle

In this section, the effect of changing the angle of view, α, while keeping other imaging parameters constant ($d_{\text{temporal}} = 1$, $d_{\text{spatial}} = 1$) is investigated. Changing the viewing angle of the camera (see Figure 3.2 for the definition of α) has a pronounced effect on the estimated a_{ct} and P_{sub} for each of the measured models. This was shown in Figure 3.14 where it was seen that a non-zero angle of view makes the apparent glottal width smaller. This apparent decrease in the glottal width at offset angles of view, biases the MAP estimates. Specifically, the estimated a_{ct} tends to increase, while P_{sub} tends to decrease with increased viewing angles. Figure 4.7 illustrates the effect of viewing angle on P_{sub} and a_{ct} for the four models.

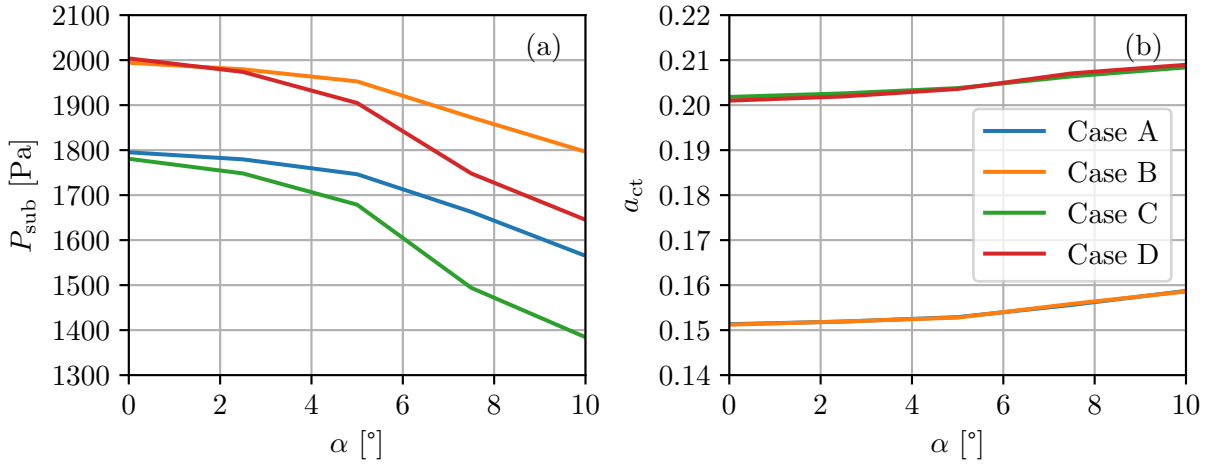


Figure 4.7: MAP estimates of P_{sub} (a) and a_{ct} (b) with a changing angle of view (α) for the 4 experimental cases.

The reason behind the decrease of P_{sub}, is the parallax effect of the camera at offset viewing angles, which reduces the measured glottal width as illustrated in Figure 4.19.

As seen in Section 2.4.2, P_{sub} is related to the the intra-glottal pressure driving the VF motion. In general, a larger P_{sub}, leads to larger intra-glottal pressures, and therefore larger amplitudes of the glottal width waveform. To compensate for the decreased measured glottal width, the estimated P_{sub} also decreases. The increase in a_{ct} is due to the frequency of the glottal width time series. While the amplitude of the observed glottal width changes due to the angle of view, α, the frequency of the waveform does not. As P_{sub} decreases, the frequency of oscillation tends to decrease while a_{ct} increases to compensate. This is shown in Figure 4.8, which illustrates the frequency of the first mode of vibration at different a_{ct} and P_{sub} values, along with the MAP estimates from Case A for each angle of view. Figure 4.8 clearly shows that the MAP estimates tend to lie along contours with constant frequency. The fact that the MAP estimates do not exactly follow the contours, is likely due to the fact that only 17 oscillations were used. As seen in Figure 4.3, the correlation direction has not fully plateaued under this duration.

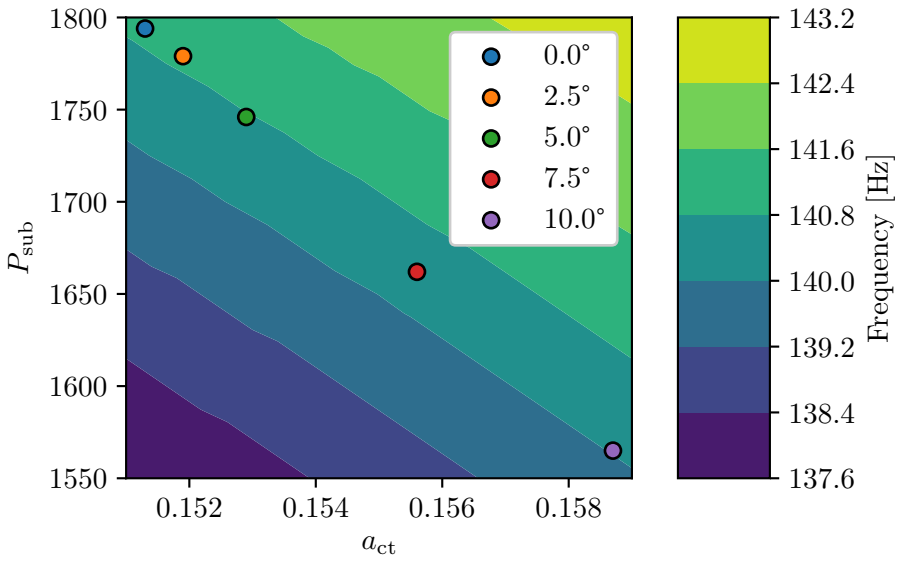


Figure 4.8: The first mode of the glottal width for the inference BCM over a range of a_{ct} and P_{sub} parameter values. MAP estimates of the parameters of Case A are shown as dots for varying angles of view, α.

Changing the angle of view also has a large effect on the uncertainty of the estimated parameters. As the angle of view increases, so does the relative uncertainty, as seen in Figure 4.9. The reason for the trends in relative uncertainty can be explained through the Laplace approximation of the posterior. As described in Section 2.5.2, the Laplace approxi-

mation of the posterior can be used to approximate the posterior as a Gaussian distribution. This results in the approximate covariance of $\Sigma = \sigma_{\text{obs}}^2 \left(J^T J\right)^{-1}$, see equation (2.46). Here the Jacobian, J, represents the changes in glottal width relative to changes in the parameters about the MAP estimate. Since the measurement noise (σ_{obs}) remains constant, the uncertainty must be due to changes in model dynamics, represented by J. Furthermore, since the model dynamics are influenced purely by the governing equations of the BCM, the changes in uncertainty shown in Figure 4.9 are a product of the BCM model and its local behaviour about the estimated parameter set. For a general MAP estimate it is not guaranteed that the relative uncertainty would also increase. To understand how this uncertainty varies for a general MAP estimate, the Jacobian would have to be computed over a range of parameters.

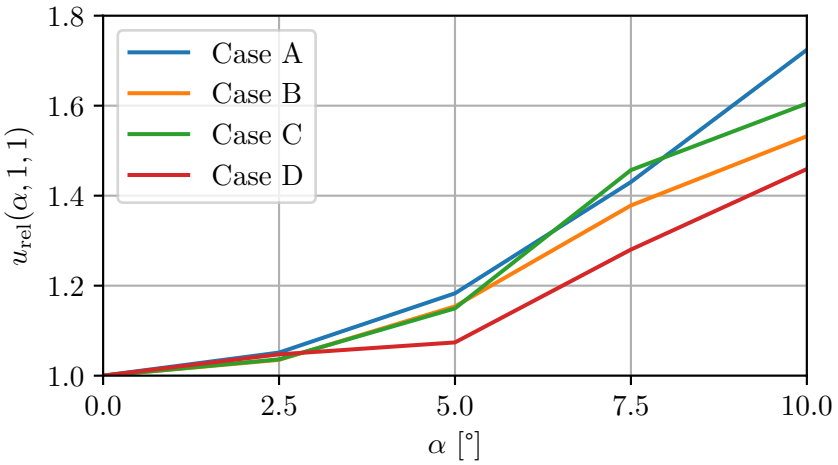


Figure 4.9: Change in relative uncertainty in the posterior with changing angle of view for the 4 considered cases.

To illustrate the applicability of the Laplace approximation, the approximate posteriors, were computed through the Laplace approximation, and the actual posteriors, with varying α, are compared in Figure 4.10. Note that the approximate covariance was computed by numerically calculating the Jacobian with a 2^{nd} order central difference. The step sizes were chosen as $\Delta a_{\text{ct}} = 0.001$ and $\Delta P_{\text{sub}} = 10\,\text{Pa}$. These step sizes were found to minimize numerical error, due to the unsmooth behaviour seen in Figure 4.8, yet preserve linearity of the forward model over the step size.

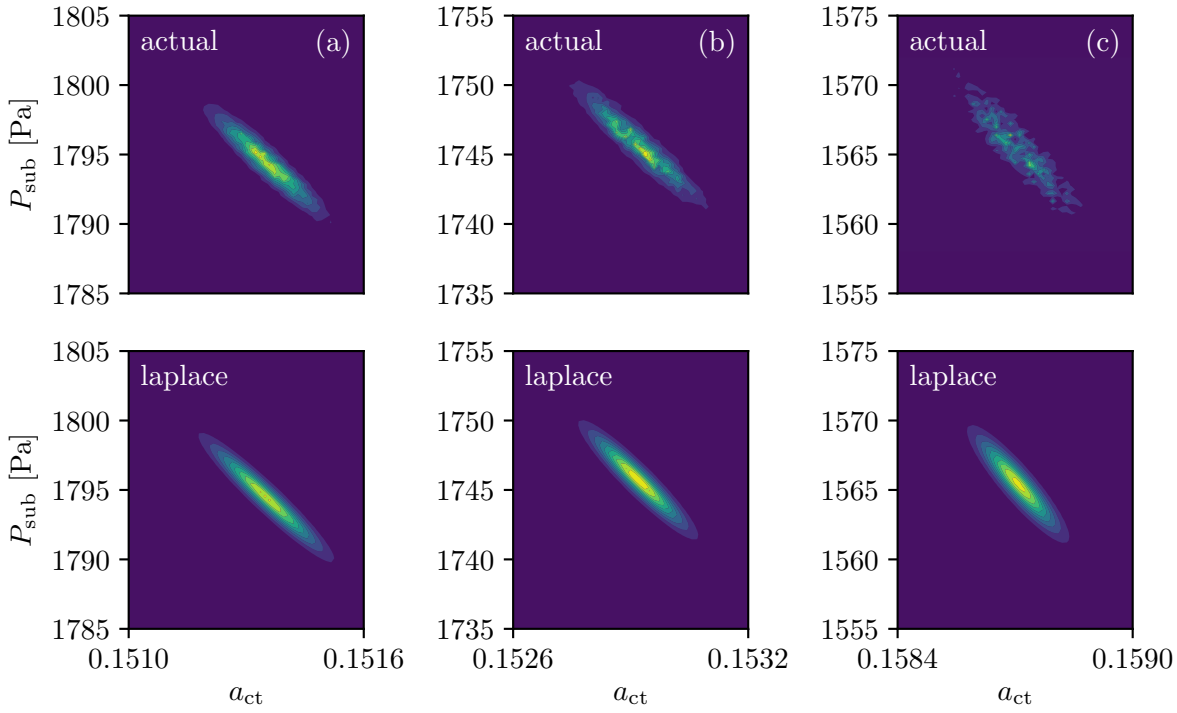


Figure 4.10: A comparison of actual posterior distributions (top row) versus their Laplace approximations (bottom row). Columns (a), (b), and (c) show the posterior distributions at $\alpha = 0°$, $\alpha = 5.0°$, and $\alpha = 10.0°$, respectively.

Visually, the two posteriors seem to agree reasonably well for all three angles represented. One detail to note is that the actual posteriors in Figure 4.10(b) and Figure 4.10(c) are not smooth. This is a product of the numerical accuracy of solving the inference BCM. As seen in Figure 4.6, the BCM output is sensitive to the integration time step. Therefore the accuracy of resolving the posterior is dependent on the integration time step used to solve the BCM. This decreasing accuracy in computing the actual posterior is shown in a comparison of the variances (the diagonal elements of the covariance matrix, see Section 2.5.2 for details) in Figure 4.11. While the general trends in the variance computed by the Laplace approximation and the actual posterior are similar, at larger angles of view, the Laplace approximation becomes less accurate. This is largely due to poor resolution of the posterior (as seen in Figure 4.10(c)), which is a product of the numerical accuracy of solving the BCM.

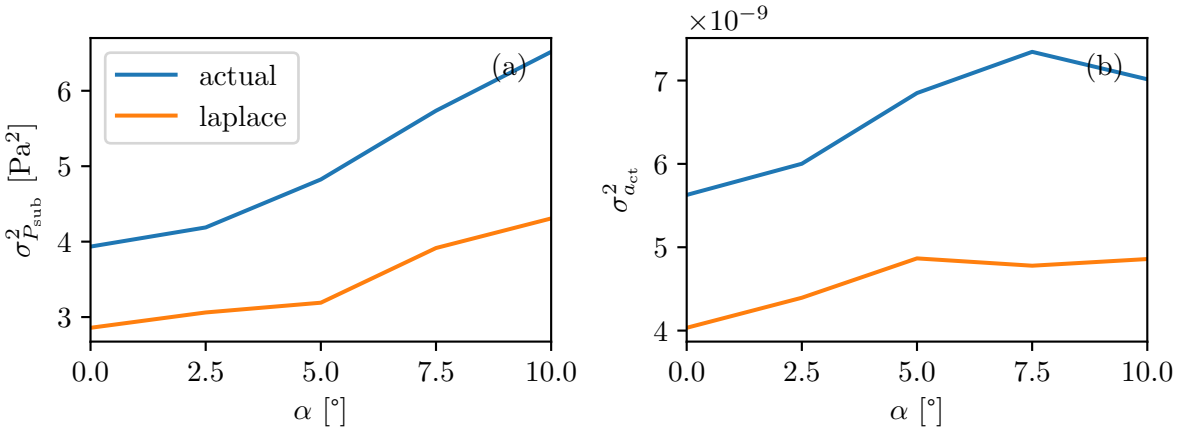


Figure 4.11: A comparison of the variances for (a) P_{sub} and (b) a_{ct} between the Laplace approximation of the posterior and the actual posterior.

The fact that variances computed from the Laplace approximation underestimate those computed from the actual posterior is reasonable. As shown in the top row of Figure 4.10, numerical accuracy issues give the actual posterior a spotty appearance. Compared to a normal distribution, where the probability mass is concentrated in a central ellipse, this increases the spread of the distribution about the mean, which increases the variance (a measure of spread about the mean of the distribution). Furthermore, we see that the posteriors become increasingly spotty at higher angles of view (this is a product of the BCM requiring finer time steps at the biased estimates). As a result, approximate covariances should underestimate covariances computed from the actual posterior, especially as the angle of view increases.

4.3.2 Effect of Frame Rate

In this section we investigate the effect of changing frame rate, d_{temporal}, while maintaining other imaging parameters fixed ($d_{\text{spatial}} = 1$, $\alpha = 0°$). Frame rate has no significant effect on the MAP estimates as shown in Figure 4.12 below. This is because frame rate downsampling was achieved through splicing of frames. This simply removes elements from the glottal width vector at specific times. For N initial glottal width measurements in the reference case, only N/d_{temporal} measurements exist when the frame rate is downsampled by a factor of d_{temporal}. While this reduces the number of glottal widths to be compared,

the glottal width waveform represented remains unchanged. In other words, while each oscillation will be represented by a fewer number of points, the shape of the oscillations are still captured. Thus MAP estimates remain the same as shown in Figure 4.12. It is important to note that this will not hold for arbitrarily large levels of downsampling. At extreme downsampling factors, the waveform would no longer be represented due to the temporal aliasing, which would alter the MAP estimates.

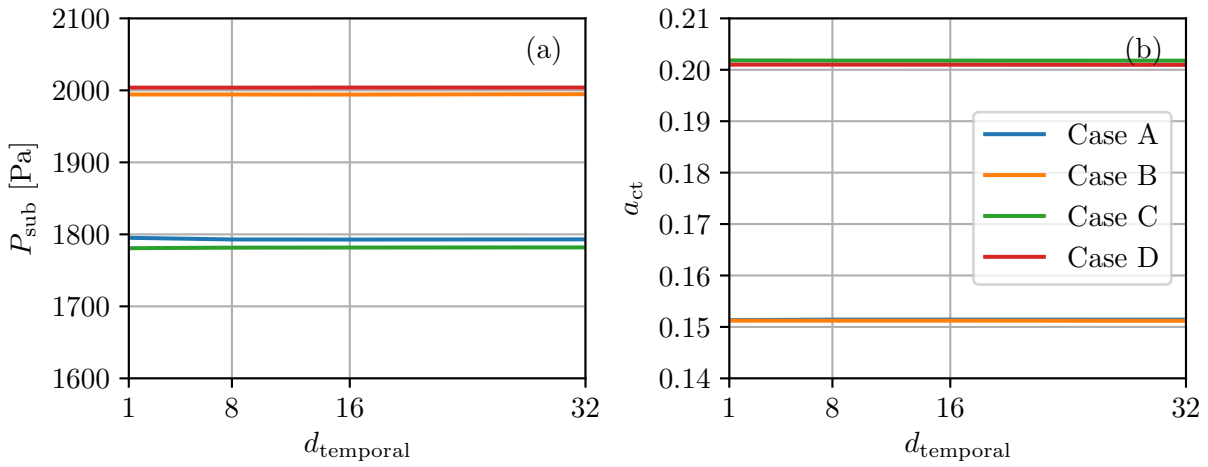


Figure 4.12: MAP estimates of (a) P_{sub} and (b) a_{ct} with increased levels of downsampling.

In contrast to its effect on the MAP estimates, frame rate has a large effect on the uncertainty of the estimated parameters, as seen in Figure 4.13. This is due to the reduction in information from the reduced number of points in the glottal width vector. Most noticeably in Figure 4.13, there is a linear increase in the magnitude of the uncertainty with respect to the downsampling factor of the frame rate.

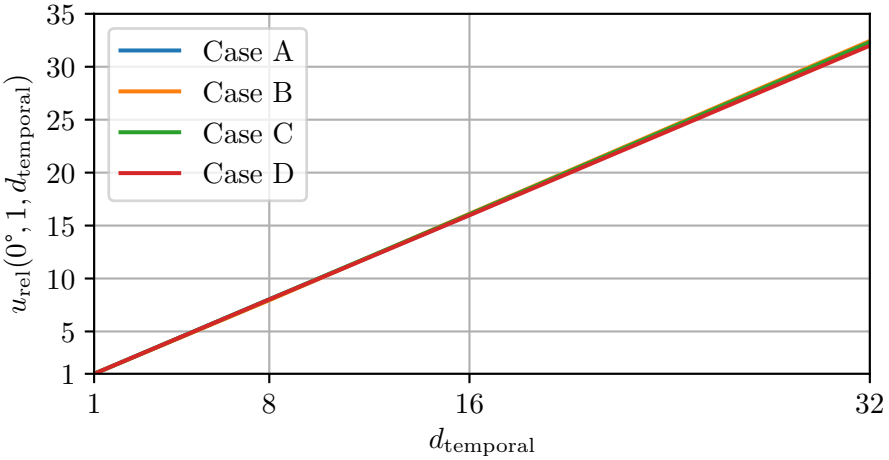


Figure 4.13: Relative uncertainty in the posterior with increased temporal downsampling (decreasing frame rate).

Considering again the Laplace approximation of the covariance (see Section 2.5.2 for details), reduction of the frame rate corresponds to removing rows of the Jacobian. The Jacobian computed in the reference case ($d_{\text{temporal}} = 1$) will be denoted as J_{ref}. In the reference case video, the Jacobian maps the change in the parameters relative to the MAP, ($\Delta a_{\text{ct}} = a_{\text{ct}} - a_{\text{ct, MAP}}$, and $\Delta P_{\text{sub}} = P_{\text{sub}} - P_{\text{sub, MAP}}$) to a change in each of the N glottal width measurements. Therefore, it is of size $N \times 2$. The Jacobian that is generated when the frame rate is downsampled, corresponds to using only equally spaced rows of the original. This Jacobian will be denoted as J_{ds}. The reference Jacobians, and downsampled Jacobians can thus be related as $J_{\text{ds } i} = J_{\text{ref } d_{\text{temporal}} \cdot i}$. We can approximate elements of the resulting covariance according to the approximate covariance shown earlier.

$$\Sigma^{-1}_{\text{ref } i,j} = \frac{1}{\sigma^2_{\text{obs}}} \sum_{k=1}^{N} J_{\text{ref } i,k} J_{\text{ref } k,j}$$

$$\Sigma^{-1}_{\text{ds } i,j} = \frac{1}{\sigma^2_{\text{obs}}} \sum_{k=1}^{N/d_{\text{temporal}}} J_{\text{ref } i,d_{\text{temporal}} \cdot k} J_{\text{ref } d_{\text{temporal}} \cdot k,j}$$

Since row n of the Jacobian represents the mapping of model parameters to changes in glottal width at frame n, the Jacobian at row n and row $n+1$ should not vary greatly. This is a consequence of the smooth behaviour of the BCM in time; over small changes in time, we expect small changes in behaviour. In fact, for high frame rates, we can extend

this and say that the Jacobian at frame n is approximately the same as the Jacobian at frame $n+k$, provided that the time change over k frames is small. Thus the previous covariance can be rewritten as:

$$\Sigma^{-1}_{\text{ref } i,j} \approx \frac{1}{\sigma^2_{\text{obs}}} \sum_{k=1}^{N/d_{\text{temporal}}} d_{\text{temporal}} \cdot J_{\text{ref } i,d_{\text{temporal}}\cdot k} J_{\text{ref } d_{\text{temporal}}\cdot k,j}$$

$$\Sigma^{-1}_{\text{ds } i,j} = \frac{1}{\sigma^2_{\text{obs}}} \sum_{k=1}^{N/d_{\text{temporal}}} J_{\text{ref } i,d_{\text{temporal}}\cdot k} J_{\text{ref } d_{\text{temporal}}\cdot k,j}$$

Further simplifying, we obtain:

$$\begin{aligned} \frac{1}{d_{\text{temporal}}} \cdot \Sigma^{-1}_{\text{ref } i,j} &\approx \frac{1}{\sigma^2_{\text{obs}}} \sum_{k=1}^{N/d_{\text{temporal}}} \cdot J_{\text{ref } i,d_{\text{temporal}}\cdot k} J_{\text{ref } d_{\text{temporal}}\cdot k,j}. \\ &\approx \Sigma^{-1}_{\text{ds } i,j} \end{aligned} \tag{4.3}$$

Clearly each element of the downsampled covariance is increased by a factor of d_{temporal} compared to the reference covariance (the factor is reserved through the inverse operation when obtaining the covariance). This is the reason for the linear scaling, on the relative uncertainty. This is only true because the reduction of the number of frames was implemented through reduction of the frame rate, and because the frame rate reduction does not significantly alias the glottal width. This is in contrast to varying the duration of the glottal width time series as was done in Section 4.1. A plot of relative uncertainty with a reduction in video duration (as opposed to frame rate) of the glottal width time series illustrates this in Figure 4.14. Specifically, in frame rate based downsampling the total number of frames, N, becomes N/d_{temporal} by only considering every ${d_{\text{temporal}}}^{\text{th}}$ frame. In contrast, video duration based downsampling reduces the total number of frames from N to N/d_{temporal} by only considering the first N/d_{temporal} frames of the video. This reduces the duration of the video at the original frame rate, rather than maintaining the duration but decreasing the frame rate.

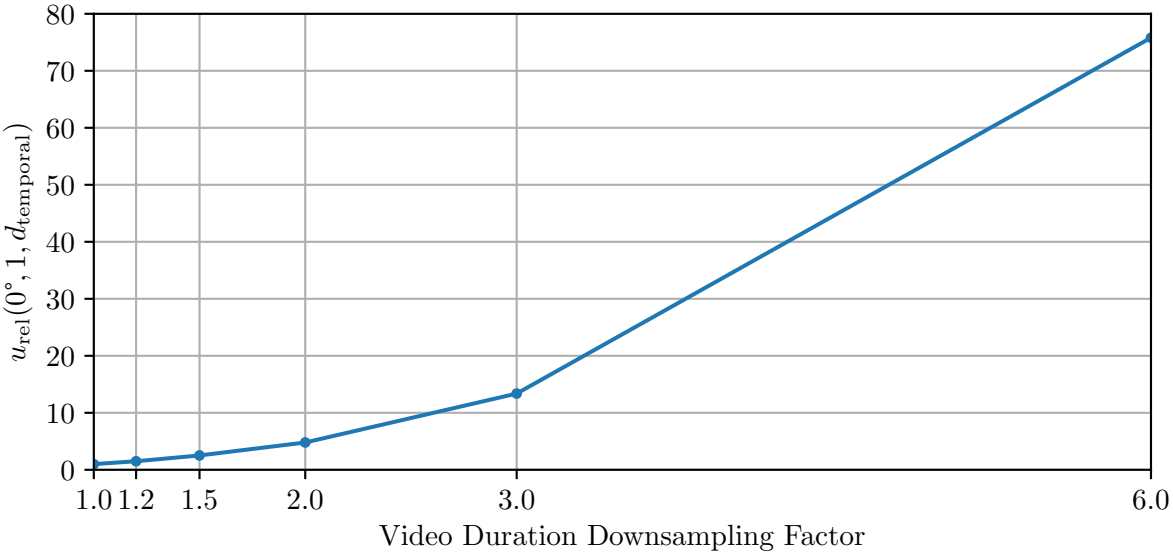


Figure 4.14: Relative uncertainty in the posteriors estimated from Case A (pictured in Figure 4.2) are shown as the video is downsampled by decreasing the duration of the video at a constant frame rate.

The large non-linear scaling on uncertainty, seen in Figure 4.14, is due to the fact that at short observation times, many BCM parameters can produce statistically reasonable behaviour. For example, it was discussed previously that longer glottal width time series lead to estimated parameters that produce similar frequencies of vibration. As was seen in Figure 4.2 however, small observation times relax this constraint, shown in the fact that the correlation direction changes. In other words, parameters that produce similar amplitudes of vibration can also produce reasonable behaviour. This leads to the large scaling on the relative uncertainty.

4.3.3 Effect of Spatial Resolution

Decreasing spatial resolution of the video (see Section 3.2) appears to slightly bias the estimated parameters, as shown in Figure 4.15. This bias in the estimates was due to a bias in the estimated glottal width from the edge detection procedure.

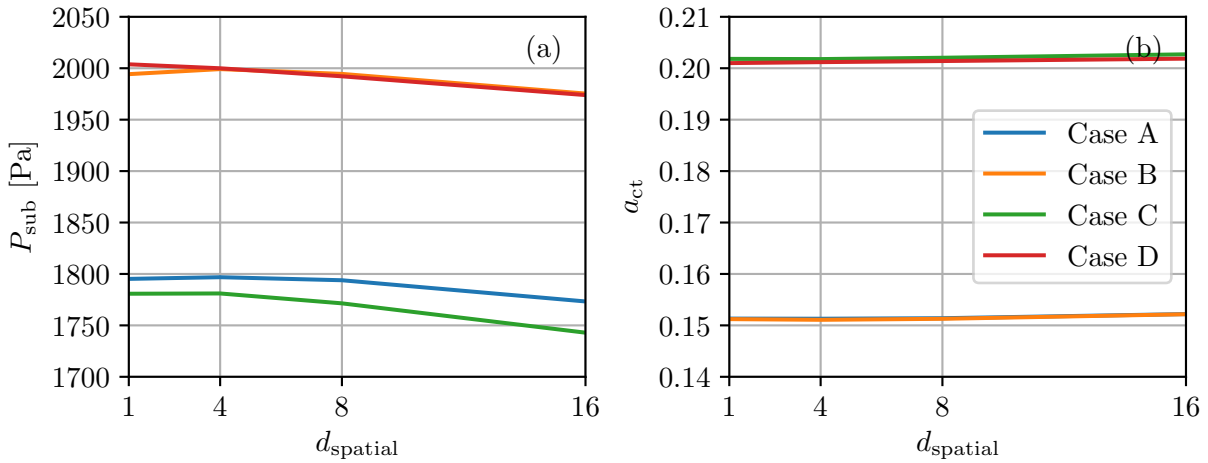


Figure 4.15: MAP estimates of (a) P_{sub}, and (b) a_{ct} with increased spatial downsampling.

Ideally, downsampling of the resolution should only change the level of uncertainty in the estimated parameters, since the detected edges should still be centered around the true edge location. As seen in Figure 4.15 however, the estimated parameters display a consistent bias. This is the result of the edge detection algorithm used in this study. At lower resolutions the edge detection procedure leads to underestimation of the glottal width, which subsequently leads to underestimation of P_{sub}. This is a result of the blurring effect introduced by the spatially downsampled images. The large levels of blur make accurate calculation of the Laplacian highly dependent on the convolution kernel size, as shown in Figure 4.16.

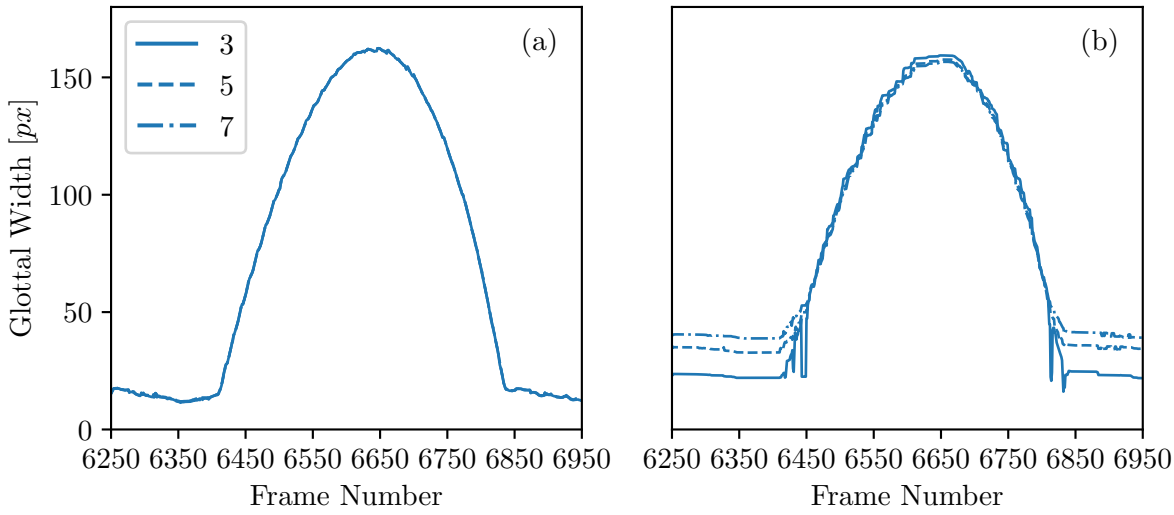


Figure 4.16: The detected glottal width for a series of frames using different kernel sizes, for the non-offset view. (a) The glottal width from the reference spatial resolution video ($d_{\text{spatial}} = 1$). (b) The glottal width from the downsampled spatial resolution video ($d_{\text{spatial}} = 16$). Note that the glottal width measured in (b) is multiplied by 16 to account for the spatial downsampling, allowing for direct comparison with (a).

In Figure 4.16(a) the glottal widths calculated with different kernel sizes are nearly identical. In this case, there is zero spatial downsampling applied to the video. In Figure 4.16(b) the measured glottal widths generally decrease as the size of the convolution kernel increases. The most accurate glottal width is obtained using a kernel size of 3, however, even this kernel size leads to a glottal width that underestimates that shown in (b). This suggests that the level of spatial downsampling is too great to allow accurate calculation of the Laplacian. The exact reason for why the inaccuracy in calculating the Laplacian leads to a consistent underestimation of the glottal width, requires further investigation.

The uncertainty of the estimates increase as the resolution decreases as seen in Figure 4.17. This increase in uncertainty is due to the increased measurement noise used in the inference procedure to account for the edge detection (see Figure 3.15).

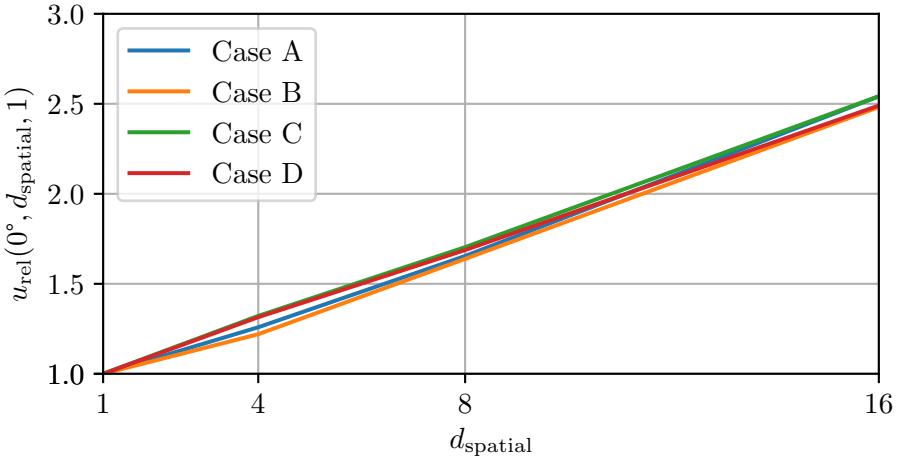


Figure 4.17: Relative uncertainty in the posterior with increased spatial downsampling.

Using the Laplace approximation of the covariance, $\Sigma = \sigma_{\text{obs}}(J^T J)^{-1}$, it is clear why there is a linear increase in the relative uncertainty. From Figure 3.15 it was seen that the variance in the edge detection procedure scales approximately linearly with decreasing spatial resolution. A linear scaling, relative to the reference case, of $\frac{\sigma^2_{\text{obs}}}{\sigma^2_{\text{ref}}} = 1 + 0.09484(d_{\text{spatial}} - 1)$ was used (as described in Section 3.3.1). The normalized scaling in the uncertainty, shown in Figure 4.17, reflects this linear trend.

It is interesting to note that if the increased measurement noise had not been accounted for, there would be no change in the relative uncertainty at different spatial resolutions. As seen in the approximate covariance, $\Sigma = \sigma^2_{\text{obs}}(J^T J)^{-1}$, the posterior covariance depends only on the BCM behaviour through the Jacobian term, and the measurement noise through the measurement error standard deviation σ_{obs}. Since downsampling of the resolution does not significantly change the MAP estimates (as shown in Figure 4.15), it follows that the model behaviour does not change, and thus the covariance should not change with a constant measurement noise. The only case in which this is not true, is if the measurement noise is sufficiently large so that the BCM no longer behaves linearly over the posterior probability mass.

4.3.4 Combined Effects

When the parameters α, d_{spatial}, and d_{temporal} are combined, the changes in relative uncertainty follow a multiplicative rule. In Sections 4.3.1, 4.3.2, and 4.3.3, the effects of viewing

angle, varying frame rate, and spatial resolution, were investigated independently.

When the independent effects are combined over the investigated parameter space, the total relative uncertainty u_{rel} is found to be the product of the relative independent changes in uncertainty:

$$\begin{aligned} u_{\text{rel}}(\alpha, d_{\text{spatial}}, d_{\text{temporal}},) &= \frac{u(\alpha, d_{\text{spatial}} d_{\text{temporal}},)}{u_{\text{ref}}} \\ &= u_{\text{rel}}(\alpha, 1, 1) \cdot u_{\text{rel}}(0^\circ, 1, d_{\text{temporal}}) \cdot u_{\text{rel}}(0^\circ, d_{\text{spatial}}, 1) \end{aligned} . \tag{4.4}$$

This is illustrated in Figure 4.18 where the relative changes in uncertainty as video parameters are simultaneously varied are shown for the posteriors estimated from Case A. The surface in orange represents the approximation of the relative uncertainty, according to equation (4.4). The surface in blue shows the actual computed relative uncertainty based on an calculation of the posterior under the combined video effects.

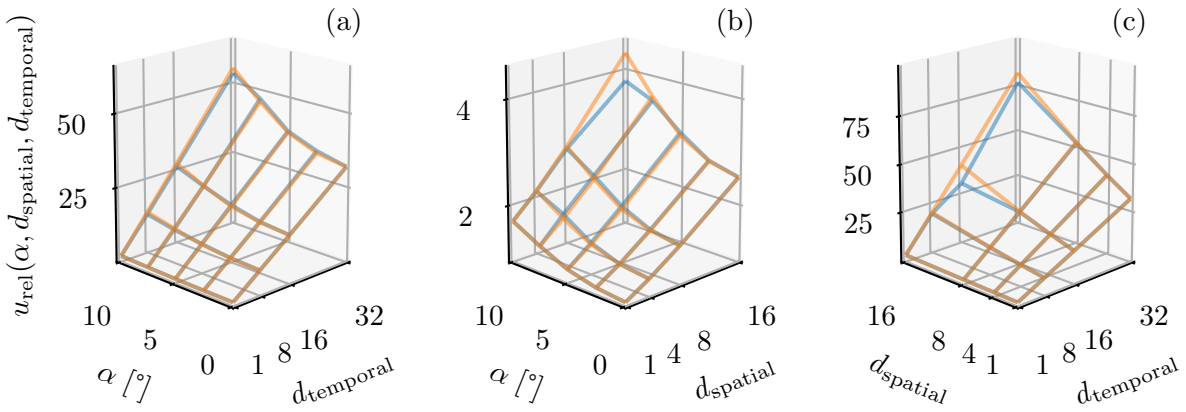


Figure 4.18: The combined effects of varying angle of view, spatial resolution, and temporal resolution on the uncertainty of the estimates compared to the reference case. This is illustrated for the posteriors estimated from Case A. Each subfigure shows the relative uncertainty under combined effects of (a) angle of view and temporal downsampling, (b) angle of view and spatial downsampling, and (c) spatial and temporal downsampling.

From Figure 4.18 it is clear that the combined effects of changes in angle (α), spatial resolution (d_{spatial}), and frame rate (d_{temporal}) on the relative uncertainty combine multiplicatively. For example, if downsampling the frame rate by a factor of 2 increases relative

uncertainty by a factor of 2, and increasing the angle increases relative uncertainty by a factor of 1.2, then applying both effects will increase the relative uncertainty by a factor of $2 \times 1.2 = 2.4$.

The reason for this becomes apparent when considering the Laplace approximation of the likelihood. Changing angle changes the uncertainty by biasing the MAP estimate, which changes the linearized model behaviour, J, by changing the linearization point (the MAP). Changing frame rate changes the uncertainty by removing rows from J. The change in uncertainty from this effect comes from the smooth behaviour in time of the Jacobian. This is independent of the point about which the Jacobian is computed. Thus if a change in angle changes the uncertainty (by modifying the point about which the Jacobians computed), further applying a reduction in frame rate will simply apply on top of this, as can be seen in equation (4.3). Lastly, modification of the spatial resolution simply modifies the measurement error parameter σ_{obs}. Clearly an increase in uncertainty, due to modifying the spatial resolution, will simply apply multiplicatively on top of the changes in uncertainty due to frame rate and resolution.

These trends cannot be expected to continue forever. As the size of the posterior becomes larger, the linear approximation will no longer hold, and thus neither will the trends. This linear approximation can be invalidated in a few ways. For example, if the frame rate is downsampled to the degree that the vibration of the glottis becomes temporally aliased, it is expected that the posterior estimate will not only be biased, but behave non-linearly over the mass of the posterior. Similarly if the resolution is downsampled too greatly, the error in measuring the glottal width may be made so large, as to spatially alias the measured glottal width. This would increase the size of the probability mass, invalidating the linear assumption. Larger viewing angles would likely not invalidate the linear assumption, since the viewing angle does not directly contribute to the uncertainty in the posterior, but only through modifying the linearized behaviour of the BCM's glottal width (the Jacobian) due to biasing the MAP estimate. However, at large viewing angles, biases in the estimate would also be large, and therefore render the estimates not useful.

4.4 Significance for High-speed Videoendoscopy in Bayesian Inference

HSV is a useful indirect measurement of the VFs, from which many quantitative measures of the VFs can be derived. In this work, the glottal width was derived from a simulated HSV experiment, in which three imaging parameters were controlled: spatial resolution,

angle of view, and frame rate, and used to estimate two parameters from a BCM (a_{ct} and P_{sub}). From this experiment, the changes in the estimated parameter values and their uncertainties with respect to the imaging parameters were determined. In this section these results, and their applicability to conducting Bayesian inference on clinical HSV are discussed.

4.4.1 Optimal Frame Rate and Resolution

Many high-speed cameras are able to improve frame rate at the cost of resolution, by reading out the charges of binned pixels. This is due to the total bandwidth (the data rate at which frames of the camera can be read) limitation of high-speed cameras. A reduction of the spatial resolution allows for a proportional increase in the temporal resolution [14]. From the effects of frame rate and resolution on the estimated posteriors, we can see that increasing frame rate by a factor of d_{temporal} reduces uncertainty in the estimated posteriors far more than increasing spatial resolution by a factor of d_{spatial}. This makes it clear that for stationary estimates, HSV studies should opt for higher frame rates at lower resolutions. Alternatively, HSV studies can also record longer time series. Provided that the duration of the recording is sufficient for the posterior covariance eigenvector directions to plateau (see Figure 4.3), recording at a higher frame rate or extending the duration of the video should have the same effect. Recording longer durations of video however, does have clinical feasibility issues. For stationary estimates, this would require that the patient maintain the same position and phonation characteristics over the longer duration.

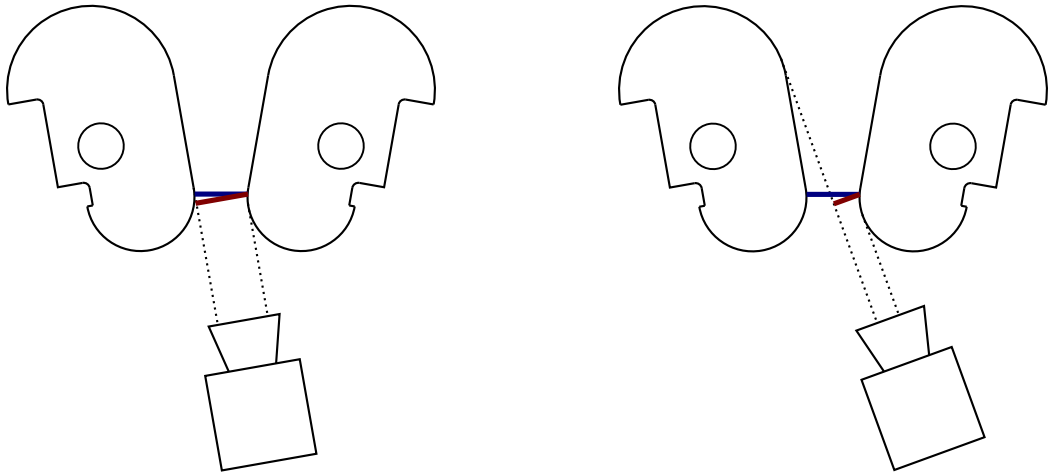
The importance of frame rate over spatial resolution may not always hold, however, depending on the edge detection algorithm. The reason that spatial resolution did not have the same effect as frame rate in this study is due to the effect of the sub-pixel precise edge detection algorithm. For example, when binning groups of 2 by 2 pixels into a single pixel, measured distances at the pixel level are scaled by a factor of 2. For a pixel-accuracy edge detection method, errors in pixel measurements are thus also $2\times$ larger. This scaling on the error is identical to the case where the frame rate is decreased by a factor of 2.

Currently there are numerous edge-detection algorithms used for detecting the glottal contours [89, 85, 86, 42, 18]. Pixel level edge detection techniques include 'snake' based contour techniques, region growing techniques, and classical edge detection techniques, such as Canny edge detection. Sub-pixel edge detection methods have used Zernike moments [85], as well as curve fitting techniques [88, 26]. While sub-pixel edge detection techniques may be possible, pixel level techniques are generally computationally less intensive and are better established. This makes a clear recommendation on an optimal trade off between frame rate and spatial resolution less distinct.

Videostroboscopy is often considered the 'gold standard' in clinical studies of phonation and is the most widely used videoendoscopy technique [12]. For perfectly periodic VF vibrations, videostroboscopy can produce videos of the VFs that appear to be moving in slow motion with excellent image quality [12], where the apparent slow motion is the result of temporal aliasing (see Section 2.3 for a review of videoendoscopy techniques). As a result, when the VF vibrations are periodic, videostroboscopy and HSV should be nearly identical, and we expect stationary estimates from the two to be the same. Videostroboscopy has the advantage of superior image quality, which is relatively difficult to achieve in HSV [12]. However, videostroboscopy also relies on the assumption of true periodicity of the VFs, which may not be the case over long recording times, due to patient and camera movement, and is certainly not the case for patients with VF disorders. Nevertheless, provided these non-periodic effects are small, videostroboscopy provides an alternative to HSV, that provides higher image quality, and is more popular in the clinic.

4.4.2 Angle of View Errors

It was seen that the angle of view of the camera can have a significant effect on the observed glottal width, by reducing the apparent glottal width amplitude. For the present study, the errors induced by this angle of view were not modelled, and as a result it was seen that estimates of model parameters are biased from their true values due to non-orthogonal viewing angles. For HSV to be applicable for Bayesian inference, it would be necessary to account for such errors in the error model for the observed glottal width. These angle of view errors are mainly the result of two effects: the projection of lengths onto the angled image plane, and measuring the glottal width between out-of-plane points on the VFs. These two effects are illustrated in Figure 4.19.



(a) (b)

Figure 4.19: Two sources of error in the measured glottal width as a result of angles. The blue lines show the transverse 'ideal' glottal width and the red lines show the glottal width measured from an offset view. (a) Angle of view error due to projection of the 'true' area onto an off-axis plane. (b) Angle of view error due to imaging of different edges.

When relating the glottal width of imaged VFs to the glottal width of a reduced order model, it is assumed that the glottal width is measured in the transverse plane. When the angle of view is not aligned with the normal of the transverse plane, this assumption is no longer valid. In Figure (a), the width in the transverse plane is still measured, however due to the shifted angle of view, the projected length on the image plane is less than the true length. In Figure 4.19(b), the width is no longer measured in the transverse plane. This can significantly reduce the measured glottal width.

Two approaches that could be used to account for such errors include: accounting for the angle of view as an additional parameter, or alternatively considering the angle of view as a random parameter which results in additional uncertainty on the glottal width. In the first approach, the angle of view parameter is inferred as an additional parameter in the inference procedure. This 'information' about derived measures from the glottis could be contained in the shape of the time varying measurement. Since it is known that offset angles of view reduce the measured glottal width, there are two types of parameters that can correspond to a given measurement. Some parameters may produce similar glottal width waveforms as the measurement, which assumes that the the angle of view is not offset. Other parameters may produce glottal width waveforms with reduced amplitudes relative to the measurement; such parameters are also possible if the angle of view was

offset, so as to reduce the measured glottal width. Thus how well these glottal waveforms match the shape of the observed glottal waveform, can provide information as to what the angle of view is.

In the second approach, the angle of view is not estimated as an additional parameter. Instead, one can assume that the angle of view is a random variable. Since the glottal width measured is a function of the angle of view, the glottal width also becomes a random variable. The uncertainty in the glottal width would then become part of the likelihood, which would affect the uncertainty in the posterior, without having to infer another parameter. Computation of the likelihood then involves marginalizing over this additional random parameter. The resulting likelihood due to the angle would vary depending on the particular configuration of the VFs.

Chapter 5

Conclusions and Recommendations

In this work, the effect of HSV imaging parameters on Bayesian inference applied to HSV videos was investigated. This was performed using an experiment designed to simulate a HSV recording procedure. In the experiment, a set of scaled up, artificial, VFs were driven by a motion system in slow-motion, according to a BCM with known parameters. Imaging these VFs with a consumer DSLR camera resulted in a simulated high spatial and temporal fidelity HSV recording due to the increased size, and slow motion movement of the VFs. Three HSV imaging parameters, angle of view, spatial resolution, and temporal resolution were then investigated by shifting the physical camera, numerically downsampling the spatial resolution, and numerically downsampling the temporal resolution (frame rate) of the videos, respectively. Bayesian inference applied to glottal widths, detected from these simulated HSV videos, then allowed estimation of two parameters (cricothyroid activation and subglottal pressure) for a reduced order model; the same reduced order model, a BCM, was chosen in this work. Comparison of the results showed general agreement between the known parameters of the BCM on the experimental setup and estimated parameters through the inference procedure. Bayesian inference applied to video obtained under different combinations of the imaging parameters, allowed for detection of trends on the estimated posteriors.

5.1 Conclusions

This work aimed to determine the effect of HSV imaging parameters on Bayesian inference, in particular, the effect of offset angles of view, tradeoffs between frame rate and resolu-

tion, and the effects of varying frame rate, resolution, and angles of view simultaneously. Findings on these three effects are listed:

- Offset angles of view in HSV can greatly influence the parameter estimates. In comparison to the non-offset view, parameter estimates from offset views are biased, so that they produce lower amplitudes of vibration. This is due to the effect of parallax errors, that reduce the apparent size of the glottis. Modifying the angle of view does not, in itself, change the uncertainty of the estimated parameter sets. Changes in uncertainty induced by the angle of view (when compared to a non-offset view) are a result of the modified reduced order model dynamics at the biased parameter estimates.

- Frame rate (temporal resolution) and spatial resolution both affect the level of uncertainty in the estimated parameter sets. Frame rate was found to have a linear scaling on the uncertainty in the estimated parameters with respect to a downsampling factor. Reducing the frame rate by a factor of 2 led to an increase in uncertainty by a factor of 2. In contrast, resolution downsampling influences the level of uncertainty indirectly, based on the increased uncertainty involved in extracting quantitative measures from a poorly spatially resolved image. When using sub-pixel accurate edge detection algorithms, this makes the scaling on uncertainty less than that of frame rate, meaning downsampling resolution is preferable to downsampling frame rate. However, this is largely dependent on the edge detection algorithm used. Pixel accurate edge detection algorithms could introduce scaling on the uncertainty comparable with frame rate changes. Furthermore, it was seen that edge detection algorithms can introduce biases into the the estimated parameters, with lower spatial resolutions having larger biases. In contrast, at a fixed spatial resolution, varying the frame rate does not introduce biases. This makes a recommendation on a spatial/temporal resolution trade-off, problem dependent.

- Under the combined effects of downsampling resolution, frame rate, and varying the angle of view, relative to a reference image case, it was found that changes in uncertainty simply apply multiplicatively. This was due to the fact that the small size of the posterior meant that the BCM model behaviour was linear over the probability mass. With larger probability masses, non-linear behaviour could invalidate this trend. This could happen for example, if temporal downsampling became so severe as to cause significant temporal aliasing.

5.2 Recommendations

As seen from the results, imaging parameters in HSV can have significant effects on the estimated parameters. Based on the results in this work, the following recommendations are made:

- The effect of the angle of view in extracting measures of the glottal width should be taken into account. Typical measures from the glottis assume that the projection of the glottis is performed onto a transverse plane; in other words, an on-axis view. To account for the fact that the angle of view of the imaging system can vary, this uncertainty in the angle of view should be accounted for. This could be done in two ways: estimate the angle of view as an additional parameter, or consider the angle of view as an uncertain quantity, which as a result, contributes uncertainty to the measured glottal width at any glottal configuration. In the first approach an additional parameter, the angle of view, must be estimated. Such information could come from the shape of the glottal width waveform. In the second approach the uncertainty in the angle of view simply increases the uncertainty in the estimate.

- The performance of different edge detection techniques should be investigated. In this work, only relative changes in uncertainty due to changes in spatial resolution were reported. This was because the absolute error in the edge detection method was unknown. To report absolute uncertainties on parameter estimates, it is necessary to know the error inherent in different edge detection techniques. As a result, the errors inherent in different edge detection techniques (applied to images of the glottis) should be characterized. These errors could be investigated as function of spatial resolution of the video, as well as other parameters such as the level of noise in the image. These errors for different edge detection approaches, would provide researchers with guidelines on what uncertainties to use in quantitative measures of the glottis, when performing Bayesian inference, and subsequently obtain realistic uncertainty estimates on parameters. Existing edge detection techniques that have been employed in HSV include pixel level techniques such as 'snake' based contour techniques [3], and region growing techniques [42], as well as sub-pixel accurate methods such as Zernike moments [85].

- Videostroboscopy should be investigated as an alternative to HSV when VF oscillations are periodic. Videostroboscopy is a more mature videoendoscopy technique with superior image quality compared to HSV, and is widely used in clinical studies of phonation, however it lacks temporal resolution [12]. Since it is more widely

available than HSV as a clinical measurement tool, considerations of the quality of estimates using Bayesian inference should be assessed.

- Numerical inaccuracies in the BCM should be investigated and more accurate solution techniques developed. In the present study, numerical inaccuracies in solving the BCM led to non-smooth behaviour of the numerically computed glottal width over small parameter changes. This subsequently led to posterior distributions that showed irregularities (a lack of smoothness over small parameter changes). To resolve this issue, very fine numerical integration times were used in solving the forward model, which subsequently made the inference procedure computationally costly. Alternative integration techniques that avoid a fine integration time step would result in improved resolution of the posterior distributions, as well as reduce the computational cost of solving the forward model.

Bibliography

[1] F. Alipour and R. C. Scherer. "Dynamic glottal-pressures in an excised hemilarynx model". In: *Journal of Voice* 14.4 (2000), pp. 443–454.

[2] F. Alipour and R. C. Scherer. "Flow separation in a computational oscillating vocal fold model". In: *The Journal of the Acoustical Society of America* 116.3 (Sept. 2004), pp. 1710–1719.

[3] S. Allin, J. Galeotti, G. Stetten, and S. Dailey. "Enhanced snake-based segmentation of vocal folds". In: *2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821)*. Vol. 2. IEEE, 2004, pp. 812–815.

[4] R. J. Baken and R. F. Orlikoff. *Clinical measurement of speech and voice*. Cengage Learning, 2000.

[5] D. A. Berry and I. R. Titze. "Normal modes in a continuum model of vocal fold tissues." In: *The Journal of the Acoustical Society of America* 100.5 (1996), pp. 3345–3354.

[6] E. Cataldo, C. Soize, and R. Sampaio. "Uncertainty quantification of voice signal production mechanical model and experimental updating". In: *Mechanical Systems and Signal Processing* 40.2 (Nov. 2013), pp. 718–726.

[7] T. Chao, Z. Yu, and J. J. Jiang. "Extracting Physiologically Relevant Parameters of Vocal Folds From High-Speed Video Image Series". In: *IEEE Transactions on Biomedical Engineering* 54.5 (May 2007), pp. 794–801.

[8] F. Daum and J. Huang. "Curse of dimensionality and particle filters". In: *2003 IEEE Aerospace Conference Proceedings (Cat. No.03TH8652)*. Vol. 4. IEEE, 2003, 4_1979–4_1993.

[9] M. de Oliveira Rosa, J. C. Pereira, M. Grellet, and A. Alwan. "A contribution to simulating a three-dimensional larynx model using the finite element method". In: *The Journal of the Acoustical Society of America* 114.5 (2003), p. 2893.

[10] G. Z. Decker and S. L. Thomson. "Computational Simulations of Vocal Fold Vibration: Bernoulli Versus Navier–Stokes". In: *Journal of Voice* 21.3 (May 2007), pp. 273–284.

[11] D. D. Deliyski. "Endoscope motion compensation for laryngeal high-speed videoendoscopy". In: *Journal of Voice* 19.3 (2005), pp. 485–496.

[12] D. D. Deliyski and R. E. Hillman. "State of the art laryngeal imaging: research and clinical implications". In: *Current opinion in otolaryngology & head and neck surgery* 18.3 (2010), p. 147.

[13] D. D. Deliyski, K. Kendall, and R. Leonard. "Laryngeal high-speed videoendoscopy". In: *Laryngeal evaluation: Indirect laryngoscopy to high-speed digital imaging* (2010), pp. 245–270.

[14] D. D. Deliyski, P. P. Petrushev, H. S. Bonilha, T. T. Gerlach, B. Martin-Harris, and R. E. Hillman. "Clinical Implementation of Laryngeal High-Speed Videoendoscopy: Challenges and Evolution". In: *Folia Phoniatrica et Logopaedica* 60.1 (June 2008), pp. 33–44. arXiv: `NIHMS150003`.

[15] D. D. Deliyski, M. E. G. Powell, S. R. C. Zacharias, T. T. Gerlach, and A. De Alarcon. "Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment". In: *Biomedical Signal Processing and Control* 17 (2015), pp. 21–28.

[16] M. Döllinger, U. Hoppe, F. Hettlich, J. Lohscheller, S. Schuberth, and U. Eysholdt. "Vibration parameter extraction from endoscopic image series of the vocal folds". In: *IEEE Transactions on Biomedical Engineering* 49.8 (2002), pp. 773–781.

[17] J. S. Drechsel and S. L. Thomson. "Influence of supraglottal structures on the glottal jet exiting a two-layer synthetic, self-oscillating vocal fold model". In: *The Journal of the Acoustical Society of America* 123.6 (2008), pp. 4434–4445.

[18] M. Echternach, S. Dippold, J. Sundberg, S. Arndt, M. F. Zander, and B. Richter. "High-Speed Imaging and Electroglottography Measurements of the Open Quotient in Untrained Male Voices' Register Transitions". In: *Journal of Voice* 24.6 (Nov. 2010), pp. 644–650.

[19] B. D. Erath, S. D. Peterson, M. Zañartu, G. R. Wodicka, and M. W. Plesniak. "A theoretical model of the pressure field arising from asymmetric intraglottal flows applied to a two-mass model of the vocal folds". In: *The Journal of the Acoustical Society of America* 130.1 (July 2011), pp. 389–403.

[20] B. D. Erath and M. W. Plesniak. "Three-dimensional laryngeal flow fields induced by a model vocal fold polyp". In: *International Journal of Heat and Fluid Flow* 35 (2012), pp. 93–101.

[21] B. D. Erath, M. Zañartu, K. C. Stewart, M. W. Plesniak, D. E. Sommer, and S. D. Peterson. "A review of lumped-element models of voiced speech". In: *Speech Communication* 55.5 (2013), pp. 667–690.

[22] G. Fant. *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. Vol. 2. Walter de Gruyter, 1960.

[23] J. L. Flanagan and L. L. Landgraf. "Self-Oscillating Source for Vocal-Tract Synthesizers". In: *IEEE Transactions on Audio and Electroacoustics* AU-16.1 (1968), pp. 57–64.

[24] L. P. Fulcher, R. C. Scherer, G. Zhai, and Z. Zhu. "Analytic Representation of Volume Flow as a Function of Geometry and Pressure in a Static Physical Model of the Glottis". In: *Journal of Voice* 20.4 (2006), pp. 489–512.

[25] Y. Fung. *Biomechanics: mechanical properties of living tissues*. Springer Science & Business Media, 2013.

[26] G. E. Galindo. "Bayesian Estimation of a Subject-Specific Model of Voice Production for the Clinical Assessment of Vocal Function". Ph.D. Universidad Técnica Federico Santa María, 2017.

[27] S. Ghosal and R. Mehrotra. "Orthogonal moment operators for subpixel edge detection". In: *Pattern Recognition* 26.2 (1993), pp. 295–306.

[28] P. J. Hadwin, G. E. Galindo, K. J. Daun, M. Zañartu, B. D. Erath, E. Cataldo, and S. D. Peterson. "Non-stationary Bayesian estimation of parameters from a body cover model of the vocal folds". In: *The Journal of the Acoustical Society of America* 139.5 (May 2016), pp. 2683–2696.

[29] P. J. Hadwin, T. A. Sipkens, K. A. Thomson, F. Liu, and K. J. Daun. "Quantifying uncertainty in auto-compensating laser-induced incandescence parameters due to multiple nuisance parameters". In: *Applied Physics B* 123.4 (Apr. 2017), p. 114.

[30] L. Held and D. Sabanés Bové. "Applied statistical inference". In: *Springer, Berlin Heidelberg, doi* 10 (2014), pp. 978–3.

[31] D. P. Hill, A. D. Meyers, and R. C. Scherer. "A comparison of four clinical techniques in the analysis of phonation". In: *Journal of Voice* 4.3 (1990), pp. 198–204.

[32] M. Hirano. "Phonosurgery. Basic and clinical investigation". In: *Otologia* 21.1 (1975), pp. 299–303.

[33] K. Ishizaka and J. L. Flanagan. "Synthesis of voiced sounds from a two-mass model of the vocal cords". In: *Bell System Technical Journal* 51.6 (1972), pp. 1233–1268.

[34] S. Iwata, H. Von Leden, and D. Williams. "Air flow measurement during phonation". In: *Journal of communication disorders* 5.1 (1972), pp. 67–79.

[35] J. J. Jiang, A. G. Shah, M. M. Hess, K. Verdolini, F. M. Banzali, and D. G. Hanson. "Vocal fold impact stress analysis". In: *Journal of Voice* 15.1 (2001), pp. 4–14.

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

[37] J. L. Kelly and C. C. Lochbaum. "Speech Synthesis". In: *Proceedings of the Fourth International Congress of Acoustics*. Ed. by J. L. Flanagan and L. R. Rabiner. Dowden, Hutchinson & Ross, Stroudsburg, PA, 1962, pp. 127–130.

[38] B. Q. Kettlewell. "The influence of intraglottal vortices upon the dynamics of the vocal folds". MASc. University of Waterloo, 2014.

[39] S. Li, R. C. Scherer, M. Wan, and S. Wang. "The effect of entrance radii on intraglottal pressure distributions in the divergent glottis." In: *The Journal of the Acoustical Society of America* 131.2 (2012), pp. 1371–7.

[40] S. Li, R. C. Scherer, M. Wan, S. Wang, and H. Wu. "The effect of glottal angle on intraglottal pressure". In: *The Journal of the Acoustical Society of America* 119.1 (Jan. 2006), pp. 539–548.

[41] J. Liljencrants. "Quarterly Progress and Status Report A translating and rotating mass model of the vocal folds". In: *STL-QPSR* 32.1 (1991), pp. 1–18.

[42] J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, and M. Döllinger. "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos". In: *Medical Image Analysis* 11.4 (Aug. 2007), pp. 400–413.

[43] H. Luo, R. Mittal, and S. A. Bielamowicz. "Analysis of flow-structure interaction in the larynx during phonation using an immersed-boundary method". In: *The Journal of the Acoustical Society of America* 126.2 (2009), pp. 816–824.

[44] H. Luo, R. Mittal, X. Zheng, S. A. Bielamowicz, R. J. Walsh, and J. K. Hahn. "An immersed-boundary method for flow-structure interaction in biological systems with application to phonation". In: *Journal of Computational Physics* 227.22 (2008), pp. 9303–9332. arXiv: `NIHMS150003`.

[45] D. Marr and E. Hildreth. "Theory of edge detection". In: *Proc. R. Soc. Lond. B.* Vol. 207. 1167. The Royal Society. 1980, pp. 187–217.

[46] L. Meirovitch. *Fundamentals of vibrations*. Waveland Press, 2010.

[47] P. Mergell, H. Herzel, and I. R. Titze. "Irregular vocal-fold vibration–high-speed observation and modeling". In: *The Journal of the Acoustical Society of America* 108.6 (2000), pp. 2996–3002.

[48] R. Mittal, B. D. Erath, and M. W. Plesniak. "Fluid Dynamics of Human Phonation and Speech". In: *Annual Review of Fluid Mechanics* 45.1 (Jan. 2013), pp. 437–467.

[49] P. R. Murray and S. L. Thomson. "Vibratory responses of synthetic, self-oscillating vocal fold models". In: *The Journal of the Acoustical Society of America* 132.5 (2012), pp. 3428–3438.

[50] K. M. O. Paul C. Bryson. *Videostroboscopy*. Ed. by A. D. Meyers. 2015. URL: `https://emedicine.medscape.com/article/1891175-overview#showall` (visited on 04/10/2018).

[51] B. A. Pickup and S. L. Thomson. "Flow-induced vibratory response of idealized versus magnetic resonance imaging-based synthetic vocal fold models." In: *The Journal of the Acoustical Society of America* 128.3 (2010), EL124–EL129.

[52] B. A. Pickup and S. L. Thomson. "Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models". In: *Journal of Biomechanics* 42.14 (2009), pp. 2219–2225.

[53] R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh. "Intraglottal pressure distributions for a symmetric and oblique glottis with a uniform duct (L)". In: *The Journal of the Acoustical Society of America* 112.4 (Oct. 2002), pp. 1253–1256.

[54] R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh. "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees". In: *The Journal of the Acoustical Society of America* 109.4 (2001), pp. 1616–1630.

[55] R. C. Scherer, I. R. Titze, and J. F. Curtis. "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes". In: *J Acoust Soc Am* 73.2 (1983), pp. 668–676.

[56] S. Schuberth, U. Hoppe, M. Köster, M. Döllinger, and U. Eysholdt. "Optical measurement of the vocal fold length and elongation during phonation". In: *Proceedings of the 6th International Conference Advances in Quantitative Laryngology*. 2001.

[57] S. Schuberth, U. Hoppe, M. Döllinger, J. Lohscheller, and U. Eysholdt. "High-Precision Measurement of the Vocal Fold Length and Vibratory Amplitudes". In: *The Laryngoscope* 112.6 (2002), pp. 1043–1049.

[58] R. Schwarz, M. Döllinger, T. Wurzbacher, U. Eysholdt, and J. Lohscheller. "Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model." In: *The Journal of the Acoustical Society of America* 123.5 (2008), pp. 2717–32.

[59] R. Schwarz, U. Hoppe, M. Schuster, T. Wurzbacher, U. Eysholdt, and J. Lohscheller. "Classification of unilateral vocal fold paralysis by endoscopic digital high-speed recordings and inversion of a biomechanical model". In: *IEEE transactions on biomedical engineering* 53.6 (2006), pp. 1099–1108.

[60] J. A. Seikel, D. G. Drumright, and D. W. King. *Anatomy & Physiology for Speech, Language, and Hearing*. 3rd ed. Thomson Delmar Learning, 2005.

[61] P. Šidlof, J. G. Švec, J. Horáček, J. Veselý, I. Klepáček, and R. Havlík. "Geometry of human vocal folds and glottal channel for mathematical and biomechanical modeling of voice production". In: *Journal of Biomechanics* 41.5 (2008), pp. 985–995.

[62] D. E. Sommer. "Development of a coupled numerical-experimental facility to model the fluid-structure interactions of the human vocal folds by". MASc. University of Waterloo, 2014, pp. 1–106.

[63] I. Steinecke and H. Herzel. "Bifurcations in an asymmetric vocal-fold model". In: *The Journal of the Acoustical Society of America* 97.March (1995), pp. 1874–84.

[64] J. C. Stemple, N. Roy, and B. K. Klaben. *Clinical voice pathology: Theory and management*. Plural Publishing, 2014.

[65] B. H. Story and I. R. Titze. "Voice simulation with a body-cover model of the vocal folds". In: *The Journal of the Acoustical Society of America* 97.2 (Feb. 1995), pp. 1249–1260. arXiv: `arXiv:1011.1669v3`.

[66] C. Tao, J. J. Jiang, and Y. Zhang. "Simulation of vocal fold impact pressures with a self-oscillating finite-element model". In: *The Journal of the Acoustical Society of America* 119.6 (2006), pp. 3987–3994.

[67] C. Tao, Y. Zhang, D. G. Hottinger, and J. J. Jiang. "Asymmetric airflow and vibration induced by the Coanda effect in a symmetric model of the vocal folds". In: *The Journal of the Acoustical Society of America* 122.4 (Oct. 2007), pp. 2270–2278.

[68] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. Vol. 89. siam, 2005.

[69] S. L. Thomson, L. Mongeau, and S. H. Frankel. “Aerodynamic transfer of energy to the vocal folds”. In: *Journal of the Acoustical Society of America* 118.3 Pt 1 (2005), pp. 1689–1700.

[70] L. Tierney and J. B. Kadane. “Accurate approximations for posterior moments and marginal densities”. In: *Journal of the american statistical association* 81.393 (1986), pp. 82–86.

[71] I. R. Titze. “A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice”. In: *Journal of Voice* 18.3 (2004), pp. 292–298.

[72] I. R. Titze. “Nonlinear source–filter coupling in phonation: Theory”. In: *The Journal of the Acoustical Society of America* 123.5 (May 2008), pp. 2733–2749.

[73] I. R. Titze. “Normal modes in vocal cord tissues”. In: *The Journal of the Acoustical Society of America* 57.3 (1975), p. 736.

[74] I. R. Titze. “On the mechanics of vocal-fold vibration”. In: *The Journal of the Acoustical Society of America* 60.6 (1976), pp. 1366–1380.

[75] I. R. Titze. “Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model”. In: *The Journal of the Acoustical Society of America* 111.1 (2002), p. 367.

[76] I. R. Titze. “The physics of small-amplitude oscillation of the vocal folds”. In: *The Journal of the Acoustical Society of America* 83.4 (Apr. 1988), pp. 1536–1552.

[77] I. R. Titze and B. H. Story. “Rules for controlling low-dimensional vocal fold models with muscle activation”. In: *The Journal of the Acoustical Society of America* 112.3 (2002), p. 1064.

[78] M. Triep, C. Brücker, and W. Schröder. “High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds”. In: *Experiments in Fluids* 39.2 (2005), pp. 232–245.

[79] J. Van den Berg. “Myoelastic-aerodynamic theory of voice production”. In: *Journal of Speech, Language, and Hearing Research* 1.3 (1958), pp. 227–244.

[80] K. Verdolini, M. M. Hess, I. R. Titze, W. Bierhals, and M. Gross. “Investigation of vocal fold impact stress in human subjects”. In: *Journal of Voice* 13.2 (1999), pp. 184–202.

[81] C. Vilain, X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, and J. Willems. “Experimental validation of a quasi-steady theory for the flow through the glottis”. In: *Journal of Sound and Vibration* 276.3-5 (Sept. 2004), pp. 475–490.

[82] D. Wong, M. R. Ito, N. B. Cox, and I. R. Titze. "Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases." In: *The Journal of the Acoustical Society of America* 89.1 (1991), pp. 383–394.

[83] T. Wurzbacher, M. Döllinger, R. Schwarz, U. Hoppe, U. Eysholdt, and J. Lohscheller. "Spatiotemporal classification of vocal fold dynamics by a multimass model comprising time-dependent parameters". In: *The Journal of the Acoustical Society of America* 123.4 (2008), pp. 2324–2334.

[84] T. Wurzbacher, R. Schwarz, M. Döllinger, U. Hoppe, U. Eysholdt, and J. Lohscheller. "Model-based classification of nonstationary vocal fold vibrations". In: *The Journal of the Acoustical Society of America* 120.2 (2006), pp. 1012–1027.

[85] Q. Xulei, W. Supin, and W. Mingxi. "Improving Reliability and Accuracy of Vibration Parameters of Vocal Folds Based on High-Speed Video and Electroglottography". In: *IEEE Transactions on Biomedical Engineering* 56.6 (June 2009), pp. 1744–1754.

[86] Y. Yan, X. Chen, and D. Bless. "Automatic Tracing of Vocal-Fold Motion From High-Speed Digital Images". In: *IEEE Transactions on Biomedical Engineering* 53.7 (July 2006), pp. 1394–1400.

[87] A. Yang, D. A. Berry, M. Kaltenbacher, and M. Döllinger. "Three-dimensional biomechanical properties of human vocal folds: Parameter optimization of a numerical model to match in vitro dynamics". In: *The Journal of the Acoustical Society of America* 131.2 (2012), p. 1378.

[88] A. Yang, M. Stingl, D. A. Berry, J. Lohscheller, D. Voigt, U. Eysholdt, and M. Döllinger. "Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model." In: *The Journal of the Acoustical Society of America* 130.2 (2011), pp. 948–64.

[89] Y. Zhang, E. Bieging, H. Tsui, and J. J. Jiang. "Efficient and Effective Extraction of Vocal Fold Vibratory Patterns from High-Speed Digital Imaging". In: *Journal of Voice* 24.1 (Jan. 2010), pp. 21–29. arXiv: `NIHMS150003`.

[90] Y. Zhang, C. Tao, and J. J. Jiang. "Parameter estimation of an asymmetric vocal-fold system from glottal area time series using chaos synchronization". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 16.2 (June 2006), p. 023118.

[91] Z. Zhang. "Characteristics of phonation onset in a two-layer vocal fold model". In: *The Journal of the Acoustical Society of America* 125.2 (Feb. 2009), pp. 1091–1102.

[92] Z. Zhang, L. Mongeau, and S. H. Frankel. "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes". In: *The Journal of the Acoustical Society of America* 112.4 (Oct. 2002), pp. 1652–1663.

[93] W. Zhao, C. Zhang, S. H. Frankel, and L. Mongeau. "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms". In: *The Journal of the Acoustical Society of America* 112.5 (2002), pp. 2134–2146.

[94] X. Zheng, S. A. Bielamowicz, H. Luo, and R. Mittal. "A Computational Study of the Effect of False Vocal Folds on Glottal Flow and Vocal Fold Vibration During Phonation". In: *Annals of Biomedical Engineering* 37.3 (Mar. 2009), pp. 625–642.

[95] X. Zheng, R. Mittal, Q. Xue, and S. A. Bielamowicz. "Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model". In: *The Journal of the Acoustical Society of America* 130.1 (2011), pp. 404–415.

[96] X. Zheng, Q. Xue, R. Mittal, and S. Beilamowicz. "A Coupled Sharp-Interface Immersed Boundary-Finite-Element Method for Flow-Structure Interaction With Application to Human Phonation". In: *Journal of Biomechanical Engineering* 132.11 (2010), p. 111003.

Glossary

abduction Refers to the movement of a body part towards the midline of the body.

adduction Refers to the movement of a body part away from the midline of the body.

anterior Anatomical direction pointing towards the front of the body.

arytenoid A set of paired cartilages of the larynx.

Bayes' formula A formula relating the conditional probabilities of two events. In bayesian inference, this formula provides a way to calculate the probability of two 'events': the model parameters given a measurement.

cartilage A type of flexible structural tissue in the human body.

corniculate A set of paired cartilages of the larynx.

coronal plane Anatomical plane splitting the body into front and back halves.

cricoid A cartilage of the larynx.

cricothyroid A muscle attaching the cricoid to the thryoid cartilages of the larynx. Primarily responsible to stretching and loosening the vocal folds by articulating these two cartilages, since the vocal folds are attached between them.

epiglottis A cartilage of the larynx

evidence The probability of a specific measurement independent of any parameter sets. This serves to normalize the product of the likelihood and the prior in Bayes' formula, to a probability.

extrinsic With reference to a particular anatomical structure, extrinsic muscles have one attachment point on the structure itself, and one an external structure.

glottal flow The air flow between the VFs, driven by contraction of the lungs.

glottis The set of two vocal folds and the immediate space between them, although it often refers to only the immediate space between the folds.

high-speed videoendoscopy A type of videoendoscopy in which high-speed video is recorded from the endoscope using high-speed cameras. This produces high frame rate videos, with a 2D video (as opposed to a line-scan video) at variable spatial resolutions, limited by data transfer rates of high-speed cameras.

inferior Anatomical direction pointing towards the feet.

intrinsic With reference to a particular anatomical structure, intrinsic muscles have attachment points only on the structure itself.

larynx The cartilaginous structure that houses the vocal folds.

lateral Anatomical direction pointing along the sides of the body, to the left and right.

lateral-cricoarytenoid A muscle attaching the cricoid to the arytenoid cartilages of the larynx. Primarily responsible for adducting the vocal folds.

likelihood A term in Bayes' formula, being a measure of the relative chances of a parameter set to produce a measurement.

medial Denotes the middle of a structure, or the portion of it closest to the midline of the body.

posterior A term in Bayes' formula, being the probability of model parameters given a measurement.

posterior Anatomical direction pointing towards the back of the body.

prior A term in Bayes' formula, being the probability of model parameters apriori to any measurements.

reduced order model A model consisting of multiple discrete units that interact with each other. Many of these discrete units are used to represent the behaviour of a more complex system. For vocal fold models, the bulk mass is split into smaller masses, which are connected by springs and dampers, representing the viscoelastic properties of the vocal folds.

sagittal plane Anatomical plane splitting the body into left and right halves.

subglottal Below (inferior) to the glottis.

subglottal tract The tube like structure that leads from below the larynx to the lungs.

superior Anatomical direction pointing towards the head.

supraglottal Above (superior) to the glottis.

supraglottal tract The tube like structure that leads from above the larynx and terminates at the mouth and nose.

thyroarytenoid A muscle attaching the thyroid to the arytenoid cartilages of the larynx. It forms the innermost core portion of the vocal folds.

thyroid A cartilage of the larynx

transverse plane Anatomical plane splitting the body into top and bottom halves.

videoendoscopy A type of endoscopy in which video is recorded of the view obtained through the endoscope. In imaging the vocal folds, this involves coupling a camera with an endoscope whose objective lens is directed at the vocal folds.

videokymography A type of videoendoscopy in which a line-scan video across one lateral line of the glottis is obtained. The low spatial resolution of a line scan video allows videos to be obtained at a high frame rate, such that the vocal fold motion throughout an oscillation can be observed.

videostroboscopy A type of videoendoscopy in which video is through an endoscope in combination with stroboscopic lighting. The stroboscope is synchronized with microphone recordings of phonation to create a slow motion video of the vocal folds by recording different phases of their oscillation over multiple cycles.

vocal fold One of the two whitish coloured viscoelastic tissues that vibrate together during phonation.