2012-07-05

Three Dimensional Characterization of Vocal Fold Fluid Structure Interactions

Joseph R. Nielson
Brigham Young University - Provo

Three Dimensional Characterization of Vocal Fold Fluid Structure Interactions

Joseph Nielson

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Tadd T. Truscott, Chair
Julie C. Vanderhoff
R. Daniel Maynes

Department of Mechanical Engineering

Brigham Young University

August 2012

ABSTRACT

Three Dimensional Characterization of Vocal Fold
Fluid Structure Interactions

Joseph Nielson
Department of Mechanical Engineering, BYU
Master of Science

Voice quality is strongly linked to quality of life; those who suffer from voice disorders are adversely affected in their social, family, and professional relationships. An effort has been made to more fully understand the physics behind how the voice is created, specifically the fluid structure interactions that occur during vocal fold vibration. Many techniques have been developed and implemented to study both the motion of the vocal folds and the airflow that creates the motion. Until recently these techniques have sought to understand a highly three-dimensional phenomenon with 1D or 2D perspectives.

This research focuses on the development and implementation of an experimental technique to obtain three-dimensional characterizations of vocal fold motion and fluid flow. Experiments were performed on excised human vocal fold models at the University Hospital Erlangen Medical School in Erlangen, Germany.

A novel technique for tracking the motion of the vocal folds using multiple camera viewpoints and limited user interaction was developed. Four high-speed cameras (2000 fps) recorded an excised vocal fold model vibrating at 250 Hz. Based on the images from these four cameras a fully 3D reconstruction of the superior surface of the vocal folds was achieved. The 3D reconstruction of 70 consecutive time steps was assembled to characterize the motion of the vocal folds over eight cycles. The 3D reconstruction accurately modeled the observed behavior of vocal fold vibration with a clearly visible mucosal wave. The average reprojection error for this technique was on par with other contemporary techniques (~20 μm).

A whole field, time resolved, three-dimensional reconstruction of the vocal fold fluid flow was obtained using synthetic aperture particle image velocimetry. Simultaneous 3D flow fields, subglottal pressure waves, and superior surface motion were presented for 2 consecutive cycles of oscillation. The vocal fold fluid flow and motion measurements correlated with behavior observed in previous three-dimensional studies. A higher resolution view of one full cycle of oscillation was compiled from 16 time resolved data sets via pressure data. The result was a full three-dimensional characterization of the evolution and disintegration of the glottal jet.

Keywords: Joseph Nielson, SAPIV, synthetic aperture, superior surface, vocal folds, three-dimensional, particle image velocimetry, glottal jet, mucosal wave

ACKNOWLEDGEMENTS

Attending Brigham Young University has been a great privilege for me. The unique gospel centered environment has been ideal for me to succeed in my studies and in my life. I am grateful for the guidance of my Father in Heaven in leading me to this opportunity at BYU and in helping me be successful.

I am grateful for the support and expertise of my two advisors, Dr. Truscott and Dr. Thomson. They have gone above and beyond what is required of them to help me do well. I appreciate their time and effort.

I am also grateful to my graduate committee members, Dr. Daniel Maynes and Dr. Julie Vanderhoff, for their support, flexibility, and guidance throughout the course of my research.

I would like to thank the organizations that have funded this research project. The National Science Foundation for providing the funds necessary to purchase the equipment for my experiments. The National Institutes of Health for funding my graduate studies and the research expedition to Erlangen, Germany, as well as the Rocky Mountain NASA Space Grant Consortium for their financial support.

I had the unique opportunity to work with the Department of Phoniatrics and Phonetics at the University Medical School in Erlangen Germany. I am grateful for the assistance and expertise of Georg Luegmair and Michael Döllinger in our collaborations.

Finally, I am grateful for the support of my wife, Candace, and my son, Joseph Jr. They have given me the encouragement, confidence, and motivation to do my best in all that I do. Their sustaining support was crucial to my success with this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

The voice plays an important role in a person's quality of life. People rely on their voices as the primary method of daily communication with family, friends, coworkers, or customers. Some people, like college professors or singers, rely on their voices for their livelihood. Unfortunately, many people suffer from voice disorders that prevent them from comfortably participating in daily communication. A focus of current voice research has been determining the physics behind voice production with the end goal of developing preventative care, surgical procedures, and post-surgical therapies that can help those suffering from voice disorders to retain their quality of life.

The aim of this study is to experimentally characterize the fluid structure interactions (FSI) of the vocal folds during speech. A novel imaging technique will be employed to characterize the 3D fluid flow and the surface motion of the vocal folds at the same moment in time. It is expected that the ability to visualize these phenomena and their interactions will help researchers and clinicians to better understand the physics of speech production and improve the clinical approach to treating patients with voice related problems.

1.1 Research Objectives

The three main research objectives are outlined as follows:

1. Reconstruct a time resolved 3D flow field of excised human vocal folds.

2. Develop and implement a method for tracking the 3D motion of the superior surface of the vocal folds.
3. Correlate 3D fluid flow, 3D vocal fold motion, and subglottal pressure in time.

1.2 Chapter Overview

Chapter 2 provides a basic overview of voice production FSI research. The anatomy and physiology of speech are briefly outlined and an in depth review of current FSI vocal fold research is presented. The motivation and contribution of this thesis project are discussed in the context of current research needs.

Chapters 3 and 4 are manuscripts, which are to be submitted to the IEEE Transactions on Medical Imaging Journal. Chapter 3 presents a novel method for tracking the superior surface motion of the vocal folds in three dimensions. The computer vision algorithms used to track the vocal folds are outlined and experimental results from data taken on excised human vocal folds are presented. The mucosal wave is clearly visible over eight cycles of oscillation and is projected in three dimensions with an average error on par with current state of the art techniques (~20 μm). Chapter 4 describes the implementation of a new fluid imaging technique to study the fluid flow over an excised human vocal fold model. Synthetic aperture particle image velocimetry (SAPIV) is used to characterize the 3D flow fields during sound production. Vocal fold FSI are further investigated through correlated flow, surface motion, and subglottal pressure data.

Chapter 5 provides a summary of the conclusions gained from this research, as well as, recommendations for future work in each of the areas of research addressed. Appendix A contains the MATLAB code that was written and executed in the development of the 3D vocal fold superior surface tracking algorithms (Chapter 3).

2 BACKGROUND

Phonation is the process by which the vocal folds produce sound. This occurs as air passes from the lungs, through the trachea, and between the vocal folds. As air reaches the vocal folds in their adducted position, pressure begins to build up. This pressure forces the vocal folds to open. The air passes between the vocal folds, the pressure decreases and the vocal folds return to their adducted position. The flow that passes between the vocal folds is known as the glottal jet. A pressure wave results from the interaction between the glottal jet and the vocal folds and this can be heard as the primary voice signal. The airstream is filtered by structures in the neck and mouth to form the sound that is heard when one speaks or sings. A basic reference diagram of the anatomy of the head, neck, and vocal folds can be seen in Figure 2-1.

Human voice production is complex and not all the factors that influence it are well understood. Voice researchers have sought to understand the physics behind vocal fold function, including two main factors, the glottal jet flow and the motion of the vocal folds. The glottal jet has been a source of study for many years and is still largely not well understood. In particular, two important aspects are relatively unknown, a) the three-dimensional nature of the glottal jet, and b) the variation of the jet from cycle-to-cycle [1]. Ideally, glottal jet studies would be performed *in vivo*, because the information gained could be directly applied to voice production and used for treatment of voice disorders. However, experimental measurements of the glottal jet are not easily obtained *in vivo*. Due to the location of the vocal folds within the larynx, methods

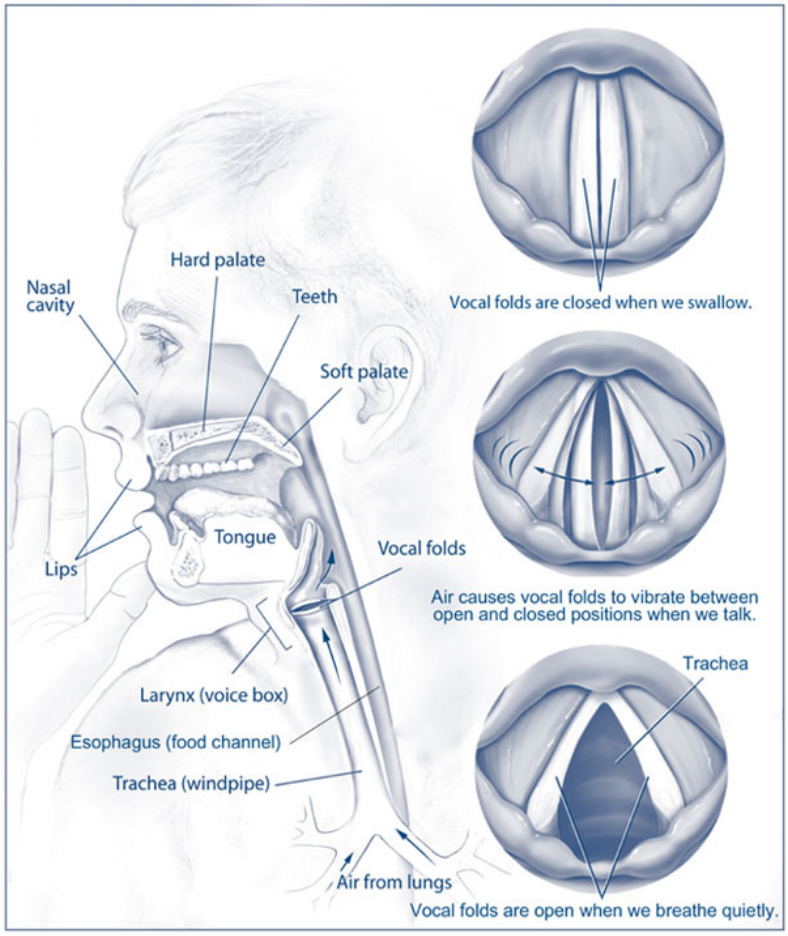


Figure 2-1: ***Left*****, a diagram identifying the important anatomy and structures involved in speech and voice production.** ***Right*****, vocal fold function described from a superior surface view. www.nih.gov.**

for visualizing the airflow in a live human being are very limited. As a result, much of what is known about the glottal jet has been developed from either *in vitro* models or synthetic physical models. These models inherently neglect any interactions of the vocal folds with other structures in the larynx, the effect of living tissue, and the unique environment inside the body. However, they provide useful method to analyze the basic physics of glottal jet evoltion and interactions with the vocal folds. *In vitro* and synthetic models can be used in parametric studies and can be used with complicated experimental techniques that simply aren't possibly *in vivo* such as particle image velocimetry and hemi-larynx motion tracking.

2.1 2D Flow Field Voice Research

Characterizing the glottal jet has been the focus of many voice research studies. Often these studies have used 2D particle image velocimetry (PIV) to reconstruct the flow fields downstream of the vocal folds. These studies have established key characteristics of glottal jet behavior such as axis switching, flow separation, and vena contracta.

Axis switching is a phenomenon that is well understood with free jets from elliptical orifices. As the jet leaves the orifice, traveling in the downstream direction, the shape of the jet is relatively the same as the elliptical orifice. As the jet evolves downstream it switches to the axis perpendicular to the length of the elliptical orifice. This has been observed in several 2D PIV studies of the glottal jet [1] [2] [3]. Flow separation refers to the jet separating from the vocal folds due to an adverse pressure gradient and forming eddies and vortices [8]. Vena contracta occurs along the mid-sagittal plane of the vocal folds; from the mid-coronal plane the jet appears to fan outward, however, the jet actually narrows along the mid-sagittal plane. Vena contracta has been observed in multiple PIV studies [1] [4].

2D PIV Experiments have been performed on a variety of different vocal fold models such as: synthetic driven models, synthetic self-oscillating models, and excised larynx models [1] [2] [3] [4]. In the following sections an overview of each type of model is given as well as a review of recent studies using each model.

2.1.1 Synthetic Vocal Fold Studies

Synthetic driven models are powered by mechanisms that simulate the natural oscillating motion of the vocal folds. This is often achieved with flapping type mechanisms or cam mechanisms. PIV experiments have been done on synthetic driven models in various studies [1] [4]. Both of these studies used scaled up models of the vocal folds (3:1 and 10:1 respectively);

the working fluid was water. An advantage of driven models is that they can be used with water, at low flow rates. In water the oscillation frequency is about 15 Hz, compared to 100-300 Hz in air. Thus time-resolved PIV data can be captured without a high-speed camera and laser. The drawback is that the models and their motion differ from that of natural vocal folds. They are driven mechanically and thus do not experience the flow induced mucosal wave-like vibrations that are essential to true vocal fold vibration. Another challenge with these models when water is used as the working fluid is that the vocal fold motion and flow cannot be correlated with acoustics, a key aspect of true vocal fold operation.

Despite the drawbacks, synthetic driven models have been valuable in understanding the physics of flow through the glottis. Krane et al. [4] used digital 2D PIV on a 10:1 vocal fold model to track the evolution of the glottal jet. This study examined the unsteady nature of the glottal jet and dependence between the volume flow waveform and the frequency of vibration. Triep et al. [1] used a similar type of cam driven model with a vocal tract downstream. The study tracked the 2D evolution of the glottal jet at the center plane and found that the emerging glottal jet attached to one wall of the vocal tract that appeared to be random from cycle to cycle.

Similar studies have been done using synthetic self-oscillating models. Self-oscillating models vibrate due to the fluid structure interaction as the air passes between the synthetic vocal folds. This is more similar to the actual behavior of human vocal folds compared to the driven models. However, the way that these models vibrate does not match exactly with a true vocal fold; studies have cited their lack of mucosal wave, excess lateral-superior motion, and un-naturally high onset pressure [2] [5] [6]. Recently developed four-layer models are closer to true vocal fold anatomy and vibration properties than the previous models yet still lack the features inherent to biological tissues [7].

Neubauer et al. [8] used a one-layer vocal fold model and studied the flow structures immediately downstream of the vocal folds. Starting vortices, vortex convection, and jet flapping were reported as a result of their 2D DPIV study. From the results it was suggested that the "flapping" of the glottal jet in the turbulent region was due to large-scale motion of vortices. Drechsel et al. [2] observed similar phenomena in a 2D PIV study using a two-layer vocal fold model. The study highlighted the improvement of synthetic vocal folds with results of starting vortices, jet flapping, and axis switching similar to those which have been seen in excised vocal fold studies [3].

2.1.2 Excised Vocal Fold Studies

The excised larynx models have the advantage of being real vocal folds. This type of model most closely correlates to real voice production because of the similarity to physiology of vocal folds in the body. However, there are several challenges to working with excised larynges that make them less common in vocal fold flow research. First, the biological tissue has a limited working life of only a few hours due to the tissue drying out. As biological tissue dries out, it responds less and less like tissue found in the body. Second, each larynx is unique and as such the experimental setup has to be adjusted and recalibrated for every new larynx that is used. Third, it is difficult to establish a good experimental setup in order to adequately see the glottal jet. When preparing the vocal fold model, care has to be taken to remove as much cartilage as possible so as to maximize the unobstructed space above the vocal folds without damaging the integrity of the larynx. These and other difficulties have motivated the development of synthetic models like those mentioned above.

Kholsa et al. [3] studied 2D flow fields using three excised canine larynges. The data was phase averaged over 30 phase positions throughout the oscillation cycle. The analysis focused on

characterizing the vortices downstream of the vocal folds and flow separation as the jet passes through the vocal folds. The data was presented with a high-speed image of the superior surface of the vocal folds, as well as the EGG and microphone signal for the given phase of oscillation. They concluded that the location and shape of vortices downstream of the vocal folds depend on the phase of oscillation. The identified vortices as a possible driver for the vocal fold vibrations but encouraged further study of the problem.

2.2 3D Flow Field Voice Research

Recently two studies have presented a 3D glottal jet flow field in the literature. Triep et al. [9] used traditional 2D PIV at various slices of a 3D volume downstream of the vocal folds. A cam driven model was used with water as the working fluid. The time resolved flow field was reconstructed with each 2D slice phase averaged. In additional to some of the common flow observations (axis switching, etc), this study highlighted the 3D and unsteady nature of the glottal jet and the need for fully 3D flow studies of the jet.

Krebs et al. [10] used stereoscopic PIV, to resolve a third component in each slice of the 3D volume. Again, each slice was phase averaged to reconstruct the 3D volume. The study examined, in more detail, the 3D nature of axis switching and noted that the switching does not occur at a fixed distance from the vocal folds and is connected to the flow separation. Despite these recent studies of the 3D flow field there do not appear to be any whole-field, time-resolved, 3D glottal jet descriptions in the voice research literature. The lack of data in this area has motivated the work of this thesis to obtain time resolved 3D flow fields of excised human vocal folds.

2.3 Motion Tracking

The other key component of vocal fold FSI is the motion of the vocal folds. Many techniques have been developed to study this motion using both synthetic and excised vocal fold models. The hemi-larynx technique has been used by several researchers seeking to understand vocal fold motion [11] [12] [13] [14]. This technique utilizes only one vocal fold, which is set up to vibrate against a prism, producing two viewing angles from one camera. Utilizing standard stereoscopic correspondence, 3D information can be extracted from the images. This technique is advantageous because the vocal folds can be viewed from otherwise impossible vantage points, such as from below or from the side [11]. However, this technique is inherently unable to investigate interactions between two vocal folds during the vibration cycle. Other techniques have focused on extracting 3D motion data from the superior surface of a complete set of vocal folds. Luegmair et al. [15] utilized a projected grid of dots and stereoscopic imaging to extract 3D data from the vocal fold surface, which worked well but neglected information from below the folds. No single motion tracking technique has proven to be perfect, the common drawbacks include: low marker resolution, tedious user interaction, and complex calibration. Overcoming these issues will be the topic of methodological improvement for years to come.

2.4 Research Contribution

The work presented in this thesis expands each of the key areas of vocal fold FSI research. Experiments were performed on excised human vocal fold models to develop and improve current motion tracking and fluid imaging techniques. A novel technique was developed for tracking the superior surface of the vocal folds in three dimensions from multiple camera viewpoints. Based on images from four cameras a fully 3D reconstruction of the superior surface of the vocal folds was achieved. The 3D reconstruction of 70 consecutive time steps was

assembled to characterize the motion of the vocal folds over eight oscillation cycles. The 3D reconstruction modeled behavior consistent with other current vocal fold tracking techniques with an error on par with other state of the art experimental methods (~20 μm) [15].

A new fluid imaging technique, synthetic aperture particle image velocimetry (SAPIV) [16], was applied to the study of the glottal jet, to obtain a whole field, three-dimensional characterization of the fluid flow. The flow measurements were synchronized with a structured light motion tracking system and a subglottal pressure probe to obtain a better understanding of the fluid structure interactions that occur during speech.

3 THREE-DIMENSIONAL TRACKING OF VOCAL FOLD SUPERIOR SURFACE MOTION

3.1 Contributing Authors and Affiliations

Joseph R. Nielson, Scott L. Thomson, Tadd T. Truscott
Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602

Georg Luegmair, Michael Döllinger
Department of Phoniatrics & Pediatric Audiology, University Hospital Erlangen Medical School, Erlangen, Germany

3.2 Abstract

A technique using a projected grid of dots imaged simultaneously by four cameras is presented for tracking the 3D motion of the superior surface of vocal folds during sound production. The 2D pixel locations of the dots are extracted from each image and projected into 3D coordinates using epipolar geometry relationships for the camera array. A fully 3D reconstruction of the superior surface of the vocal folds is obtained from these 3D projections for a given moment in time. The 3D motion for one cycle of oscillation is obtained by compiling successive time-steps.

3.3 Introduction

Sound is produced as air passes from the lungs, through the trachea, and between the vocal folds. The air induces vibration in the vocal folds, which then creates a pressure wave. This

fluid-structure interaction (FSI) has been the focus of many studies because it is the primary source of sound during speech. One goal of current voice research is to gain a better understanding of the FSI during vocal fold vibration and an understanding of voice production in general [11]. Although a full characterization of the FSI of vocal fold vibration is still a long-term goal, methods of improved experimentation and expanded data sets can benefit the community at large. The development of computational and analytical vocal fold models relies on accurate experimental observations. Thus, we seek to determine a method of improving the quantitative position of the vocal folds as they vibrate with the thought that the end goal of improving the treatment and prevention of vocal fold related voice disorders will include vocal fold motion analysis [11] [16].

One method that has been used in the characterization of the motion of the vocal folds is stereoscopic imaging using a prism [12] [13] [14]. A camera is set to view a single vocal fold (hemi-larynx) through a prism such that two views of the vocal fold are obtained at a single moment in time from one camera. This method has been used with excised vocal fold models, as well as synthetic self-oscillating models. Döllinger et al. [12] used an excised canine larynx model, with micro sutures as markers, in a hemi-larynx configuration to track both the medial and superior surface motion of the vocal folds. This method for tracking was termed 'semi-automatic'. The first five frames required user selection of all points and the remaining frames were processed automatically. The main challenge with the excised hemi-larynx experiments is getting the vocal fold to vibrate; this is not a trivial procedure. Not only does the larynx have to be carefully prepared by a trained medical professional but also the position of the vocal fold with respect to the prism has to be adjusted precisely so that the vocal fold will actually vibrate [12].

The same hemi-larynx, prism technique has been used with synthetic, self-oscillating models. The procedure is greatly simplified using synthetic models. The behavior of the synthetic vocal folds and the physical setup can be more easily manipulated than the excised vocal folds [13]; consequently less effort is needed to get the vocal fold to vibrate. Also, using synthetic vocal folds enables the use of a variety of different markers for tracking. Some studies have used simple markers such as ink and speckled materials [13] [14]. Murray et al. [13] compared the medial and inferior surface motion of several different synthetic self-oscillating models.

An advantage of the hemi-larynx technique is that the setup can be arranged to view the vocal folds from different vantage points, such as the medial, superior, or inferior surfaces [12] [13]. However, the hemi-larynx setup inherently is unable to investigate the effect of vocal fold to vocal fold interaction during the vibration cycle. Thus any effect that the interaction between the two vocal folds may have on their overall motion during oscillation is missed. Limitations of accessibility and visibility make it difficult to investigate vocal fold motion on the medial and inferior surface while including these interactions.

While medial and inferior surface motion characterization are important to fully understanding vocal fold vibration, an understanding of the superior surface motion can potentially be very useful because this is the point of view from which information can be gathered from a patient in vivo. Current voice diagnostic techniques utilize an endoscope to visualize the vocal folds superior surface. Studying voice production FSI from the superior surface view can yield information directly applicable to current medical practices.

Depth-kymography is one technique used *in vivo* to characterize vocal fold motion [17] [18]. One line across the vocal fold surface, in the coronal direction, is recorded and depth

information is extracted. The result is a kymograph, which displays the elevation of the vocal folds at one coronal cross section over a period of time. The kymograph clearly shows the mucosal wave during oscillation. Mucosal wave properties can then be calculated such as: velocity, horizontal and vertical amplitude, and frequency. The technique can reconstruct the depth profile for a given cross section with an average error of 50 μm [17]. The main drawback of this technique is that only one cross section of the vocal folds is examined and all other 3D surface information is missed.

Another recently developed method utilizes stereoscopic imaging with one continuous light source and one high-speed camera to reconstruct the 3D superior surface motion for a pair of oscillating vocal folds [15]. The technique utilizes a continuous laser directed through a micro lens array projecting a grid of dots onto the superior surface. When the vocal folds oscillate, a high-speed camera images the vocal folds at 4000 fps. The dots are used as markers to track the 3D motion of the vocal folds from one frame to the next. This technique leverages stereoscopic correspondences between the camera and patterned grid to extract 3D information [15] [19]. The 3D reconstruction is obtained with an average error of 25 μm. The technique is limited by the complexity of calibration and image processing. The technique requires custom, high precision calibration grid and a calibration of both the high-speed camera and the laser projection system. The triangulation is very sensitive to image noise and requires significant user pre and post processing.

We propose a new technique for 3D vocal fold superior surface motion tracking that uses four camera viewpoints and a laser projection system. Using four camera viewpoints eliminates the need for a traditional calibration but instead only requires a simple alignment, obtained through auto-correlation on a checkerboard. The four camera viewpoints used are part of a new

system for fluid imaging such that the fluid dynamics and the surface motion that occur during the FSI can be captured at the same moment in time. To track the 3D surface motion, a grid of continuous laser dots is projected onto the superior surface of the vocal folds, similar to Luegmair et al. [15]. The grid is in full view of four of the cameras from the fluid imaging setup and the motion is captured during vocal fold oscillation. The 2D pixel location of the dots are initially estimated with some user input and then automatically located in all successive time steps. Additional user input can be used to correct any missed points. With the 2D point locations from all four cameras, the 3D position is estimated using epipolar geometry relationships [20]. The result is a three-dimensional reconstruction of the grid of laser dots, representing the 3D superior surface of the vocal folds at a given moment in time.

3.4 Methods

3.4.1 Experimental Setup

Excised human vocal folds from a deceased 82 year old male were prepared by removing excess cartilage and tissue surrounding the vocal folds, mounted on a 16 mm diameter stainless steel pipe (trachea tube), and clamped to ensure a tight seal (see Figure 3-1). The vocal folds were symmetrically tensioned with a torque of 2 N-m applied to each fold using a custom tensioning mechanism [21]. The trachea tube was connected to a compressed air reservoir and the flow rate was adjusted to the minimum needed to induce vibration in the vocal folds, 4 L/min. A custom LabVIEW program controlled the airflow rate.

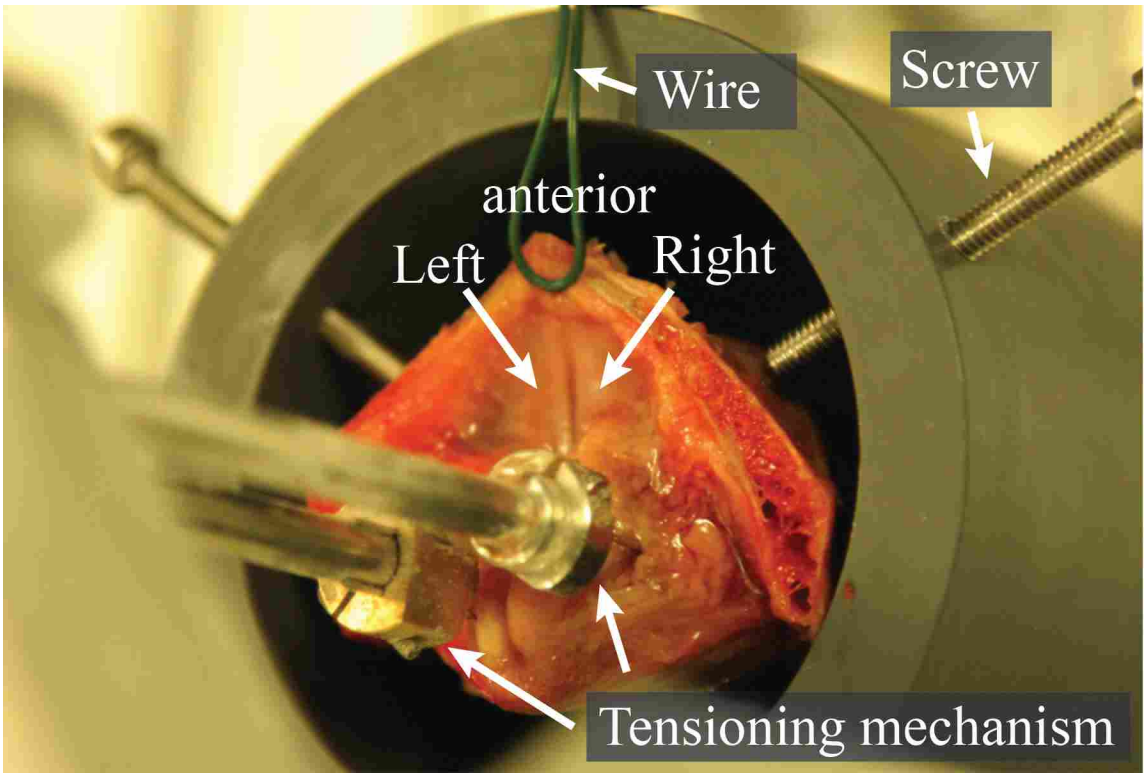


Figure 3-1: Excised human vocal folds used in this study. A mechanism is attached to each side of the model to adjust tension in the vocal folds while two screws and a wire keep it in place.

The superior surface motion tracking measurements utilized four FASTCAM SA3 high-speed cameras (Photron, 2000 fps, 640 x 640 pixels) in combination with a continuous laser system (see Figure 3-2). The laser was directed through a micro lens array that projected a grid of dots onto the superior surface of the vocal folds (see Figure 3-3). The experimental motion tracking setup was coupled with a synthetic aperture particle image velocimetry (SAPIV) fluid imaging setup, consisting of an additional four high-speed cameras and a high-speed (1000 Hz) double-pulsed laser after the manner of Belden et al. [22] and Daily et al. [23].

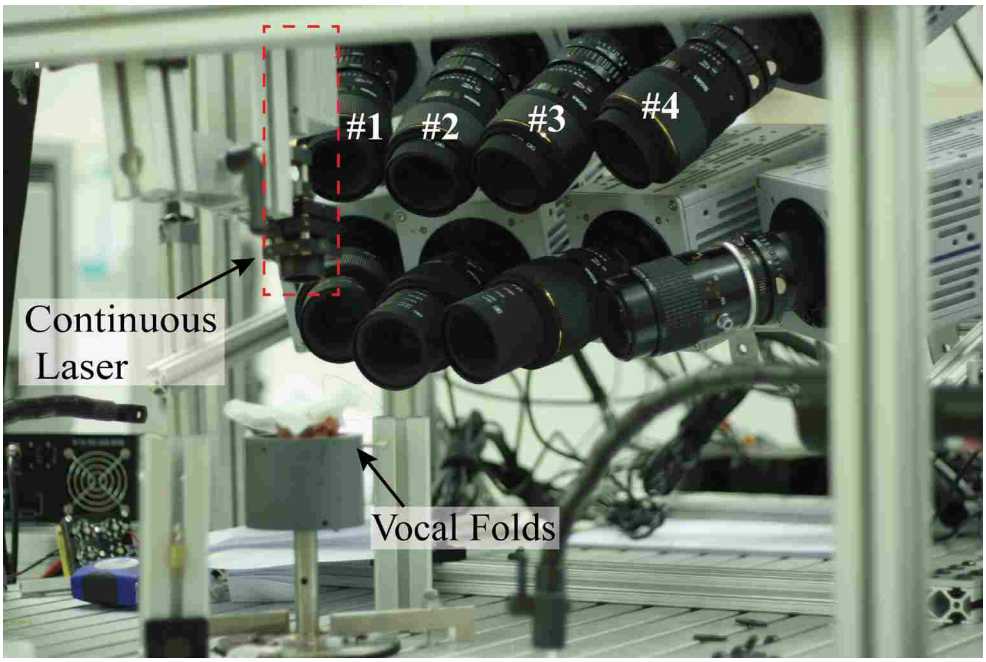


Figure 3-2: The motion tracking and fluid imaging setup. Eight high-speed cameras (Photron SA-3) are focused on the volume immediately above the vocals. A continuous laser, coupled with a micro lens array is used to project a grid of dots onto the superior surface of the vocal folds. The top row of cameras (1-4), utilized for the technique presented herein, is aimed such that the grid of laser dots is in focus within the field of view.

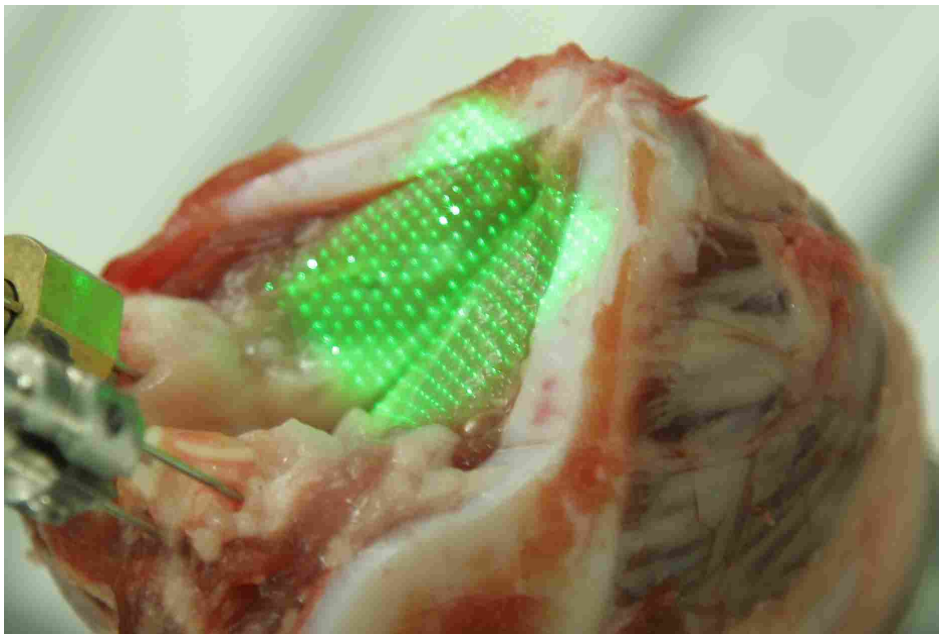


Figure 3-3: A continuous laser was directed through a micro lens array to project a grid of dots onto the superior surface of the vocal folds. This image shows an excised porcine vocal fold model with the grid of laser dots projected onto the surface. This is not the model used in the experiments of this paper.

Particle image velocimetry is a technique used extensively in experimental fluid dynamics to characterize fluid flow velocity fields. Small particles are seeded into the fluid of interest, a double-pulsed laser is fired to illuminate the particles at two moments in time, and images are taken corresponding to each laser pulse. The velocity of the particle field is determined via correlation. SAPIV is a specialized version of PIV used for reconstructing whole-field, 3D velocity fields [22].

The experimental setup, seen in Figure 3-2, shows all 8 cameras, the continuous laser, and the excised vocal fold model. The images from cameras 1-4 can be used both for SAPIV measurements and for motion tracking. In the present study these images are only used for motion tracking.

Although the cameras captured images at a rate of 2000 fps, due to the nature of PIV images the timing scheme is slightly more complicated. Figure 3-4 shows a diagram of the timing scheme used for the experiments. Even though the cameras imaged at 2000 fps, information was only gathered when the laser pulsed. Thus the motion tracking data consisted of an array of image pairs captured at 1000 Hz with spacing of 10 μs within the image pair.

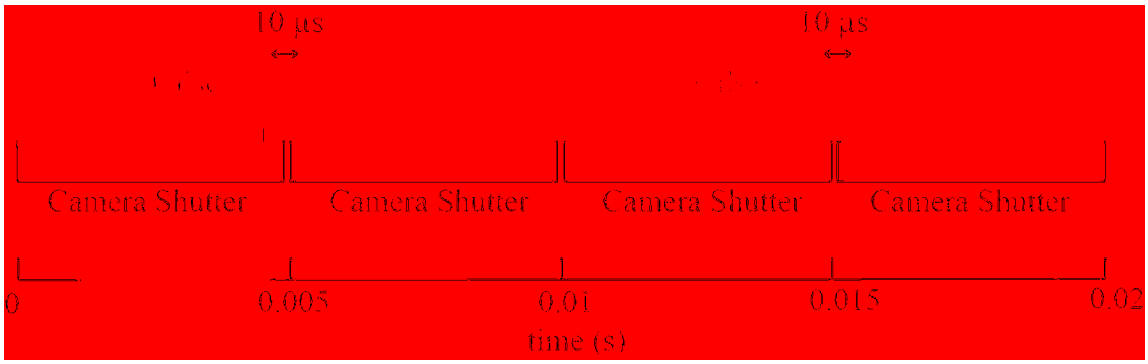


Figure 3-4: The timing scheme for the combine SAPIV and motion tracking setup. The high-speed cameras capture images at 2000 fps, however, they only collect data when the double-pulsed laser fires. The result is an array of image pairs captured at 1000 Hz with spacing of 10 μs within the image pair.

3.4.2 Data Processing

Using the images captured by the top four cameras during the experiments, the motion of the vocal folds can be extracted. Raw images from the top four cameras can be seen in Figure 3-5. The computer vision algorithm for the dot tracking consists of three main components:

(i) Preprocessing

(ii) 2D point location

(iii) Estimation of 3D projection

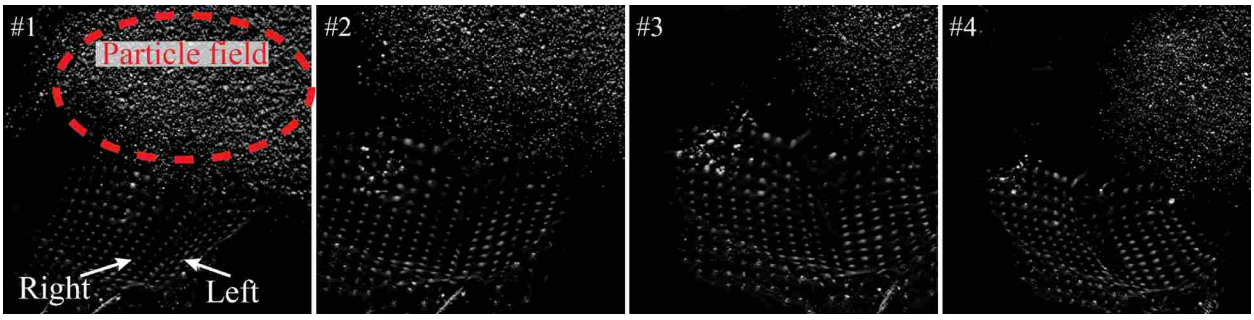


Figure 3-5: Images from the top four cameras, as presented in Figure 3-2, after preprocessing. *Left* – the dashed red ellipse and the red label indicate the particle field. The white labels and arrows indicate the right and left vocal folds. Both the particle field and the grid of dots on the superior surface of the vocal folds are clearly visible in all four cameras.

3.4.2.1 Preprocessing

For volumetric PIV, it is well established that there is a need for preprocessing to reduce the effects of differences between laser pulses and to remove background noise [22] [24]. The images of the laser grid are the same images used for SAPIV and therefore also benefit from similar preprocessing. The specific preprocessing technique used was based on the approach outlined by Belden et al. [22]. The approach consists of the following steps:

(i) Subtract sliding minimum (10 pixel window)

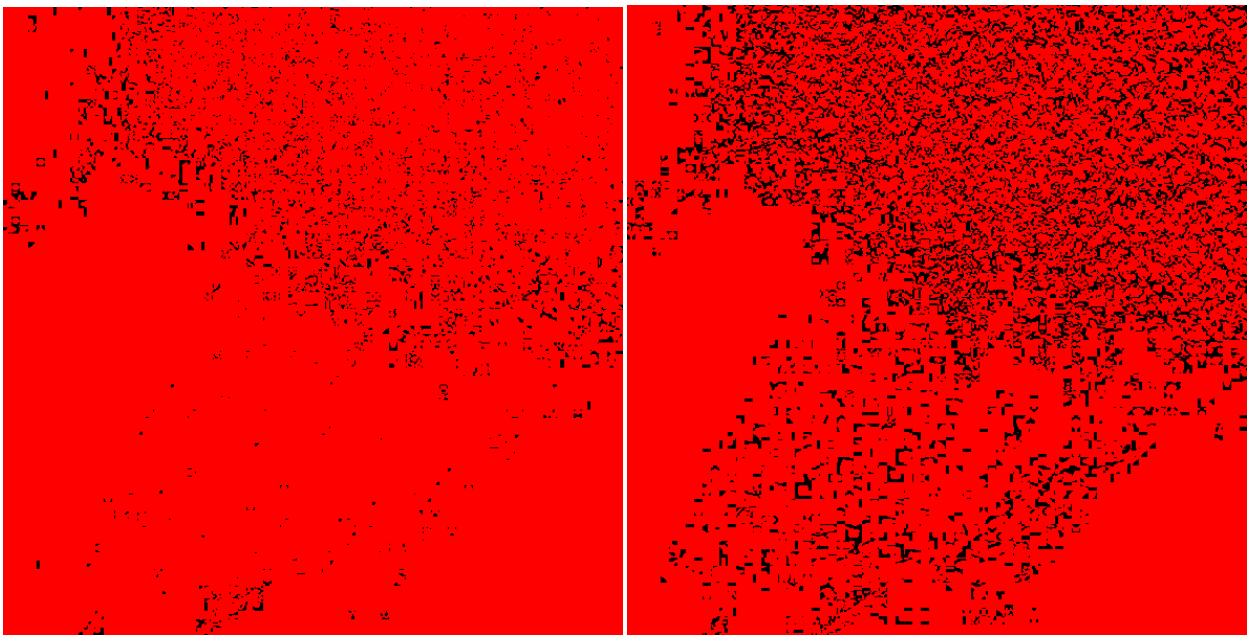
(ii) Blur the image by convolution with 3x3 Gaussian kernel

(iii) Equalize the image histograms to the histogram of the image with the highest contrast

(iv) Increase the contrast by trimming the bottom and top 0.1% of intensity values

(v) Subtract sliding minimum (10 pixel window)

The results of preprocessing show that the background noise is reduced and the laser dots are more distinguished in the preprocessed image compared to the raw one (Figure 3-6). The higher contrast image enables the dots to be extracted more easily when the image is thresholded.



(a) (b)

Figure 3-6: A comparison of a raw image and a preprocessed image. (a) A raw image of the particle field and laser dot grid as seen from camera #2; (b) the same image seen in (a) after preprocessing.

3.4.2.2 2D Point Location

Using the preprocessed images the 2D pixel locations of the dots are found automatically after some initial user direction. The user selects three points in at least three columns of interest and specifies the row and column each point belongs to. A complete grid is interpolated from the

user-defined points (see Figure 3-7). This grid is used as an initial guess for where the dots are located. When the user selects three points in every column of interest this method establishes a very good initial guess, however, it involves a little more time from the user. The results presented in this paper represent an initial guess based on user input from all seven columns of interest. The initial guess estimation would be greatly simplified if there were no particles from the fluid imaging measurements. The particles often appear with similar intensity and size as the laser dots and since the particles overlap with the laser dots in some images, it makes it difficult to distinguish between a particle and a dot without some user interaction.

Figure 3-7: The user selects three points from at least three columns and the rest of the points are interpolated to create an initial guess grid. The green points represent the points that were selected by the user and the red points represent the points that were interpolated to complete the grid. From Camera #2.

Using this initial guess as a starting point a more precise 2D pixel location of each dot is found. This is achieved through the following sequence:

(i) Threshold image

(ii) Binary morphology – modified open

(iii) Crop window around point (user specified size)

(iv) Find white regions and assign weighted score based on: proximity, area, slope, and distance.

First, an image threshold is determined and the image is binarized. For this experiment it was determined that >10/255 was a suitable threshold. Determining a suitable threshold for the images is left up to the user. The threshold of >10/255 was chosen for this experiment by simple trial and error. An image processing technique called binary morphology was used to clean up the binary image by removing small groups of "on" pixels. A modified version of the open operation was used where the image was first eroded with a 3x3 kernel of ones and then dilated with a 4x4 kernel of ones.

To find the precise location of each dot, each initial guess point is systematically visited starting from the top left corner and proceeding in column major order. A small window around the point is cropped from the thresholded image; the user specifies the size of the window. Choosing the appropriate window size involves knowing how good the initial guess is and knowing how much dot motion is expected from one time step to the next.

For the present study it was determined that for any given time step a dot would not move more than 6 pixels. Since the initial guess was very good, a window of 14x14 pixels was chosen. An appropriate window size can be determined by calculating the maximum expected motion of a dot, from one time step to the next, and the maximum distance of a guess point to an actual dot.

The algorithm searches for continuous, 4-connected, white regions within the interrogation window. If no white region is detected, then the initial guess point is erased from the grid, indicating no dot found in that window. If only one white region is found, it is assumed that this

white region corresponds to a dot on the grid. A weighted centroid of the white region is calculated to sub pixel accuracy using the MATLAB regionprops function 'WeightedCentroid', and the grid is updated with the correct dot location. If more than one white region is found a series of calculations are made and each white region is scored to determine which region is the actual dot.

The dot scoring is based on four different parameters. First, the proximity of the white region centroid to the initial guess point is considered. The proximity measure is simply the Euclidean distance between the two points, calculated using the following equation:

$$D = \sqrt{(x_{Centroid} - x_{Guess})^2 + (y_{Centroid} - y_{Guess})^2} \quad (1)$$

The white region that is closest to the initial guess is assigned the highest score, equal to the total number of white regions in the window. The white region furthest from the guess is assigned a score of one. All other white regions are assigned an intermediate score, equal to the average of the high and low score. This same scoring scheme is used for each of the four parameters.

Second, the area of each white region is calculated and scored. Larger white regions are more likely to be "real" dots and are therefore score high while the smaller white regions score lower. This parameter ensures that random speckles or partial dots on the edge of the window are not as strongly considered as other, more likely candidates.

Third, the slope of the line passing through a white region's centroid, relative to the other dots in the column, is considered. The slope of a line between a white region's centroid and the next point in the column, from the initial guess, is calculated and is compared to the slope of the line passing from the initial guess point to the next point in the column. The slope for a white region centroid that matches closely to the initial guess is considered more likely to be a real dot and is given a higher score. The parameter leverages the grid structure of the laser dot array. In

general, it is not likely that a severe change in slope will occur for any given column during normal vocal fold vibration. White regions that are drastically out of line with the rest of the column are penalized and less likely to be detected as the actual point.

Fourth, the distance from a white region's centroid to the previous dot is considered. The Euclidean distance is used for this metric (1). This parameter achieves two purposes: 1) it ensures that the same laser dot will not be located twice by penalizing points that are too close to another point that have already been identified. 2) It ensures that white regions with spacing similar to the initial guess are scored higher than regions that are unrealistically far away from the previous location. Using these four parameter scores, an overall score is calculated for each white region using (2), and the region with the highest overall score is deemed the most likely candidate. The weighted centroid of the most likely candidate is used to represent the precise 2D pixel location for the laser dot of interest.

$$S_i = w_1 s_{i1} + w_2 s_{i2} + w_3 s_{i3} + w_4 s_{i4} \tag{2}$$

Where S_i is the overall score for each white region, w is a user-defined weight, unique to each parameter, and s_i is the score from each of the four parameters. The user can adjust the weights assigned to each of the scores mentioned above. For example, in the present study, a heavier weight was put on the first parameter, the proximity to the initial guess. Since the initial guess was very good, it was expected that the true dot location would be close to the initial guess and so the score was weighted to emphasize that characteristic. For the present study the parameter weights were set at $w_i = [2, 1, 1, 1]$;

After each window has been examined and dot locations have been assigned the finalized grid is exported for use later and used as the initial guess for the next time step. The process is repeated for each camera and each time step for all time steps in the data set.

One problem that commonly occurs during this step is when a dot is in one time step but not in the next time step (dot dropout). The two challenges of dot dropout are making sure that dots are not located if there isn't supposed to be a dot there and finding a dot when it reappears in a future time step.

Thresholding and morphology can ensure that false positives are not identified, however, the thresholding step can also enhance the problem of dot dropout. Since some dots may be dimmer than others the threshold and morphology may erase the dim dots. Alternative methods of thresholding can be implemented to mitigate its effect on dot dropout.

One such method is threshold relaxation [25]. For this method an initial threshold is determined and the dot location algorithm is executed. Then locations where no dot was found are visited again, this time with a slightly relaxed threshold. The process can be repeated until all dots are found. However, at some point false positives will be introduced.

In order to address the second challenge of finding a dot when it reappears, missing dots are filled in with dots from the previous guess to use for the next guess. For example, Figure 3-8 shows a grid in which the red dots were correctly located but 4 (blue) dots were not. The blue dots are values from the previous guess and will now be used as part of the grid for the next guess but will not be included in the 3D projection of points for this time step.

At this point additional user interaction can help to identify dots that have been missed. The user is given the option of manually selecting dots that have either been misidentified or have been missed altogether. When the user selects the dot the same scoring scheme described above is used to determine the precise location of the laser dot.

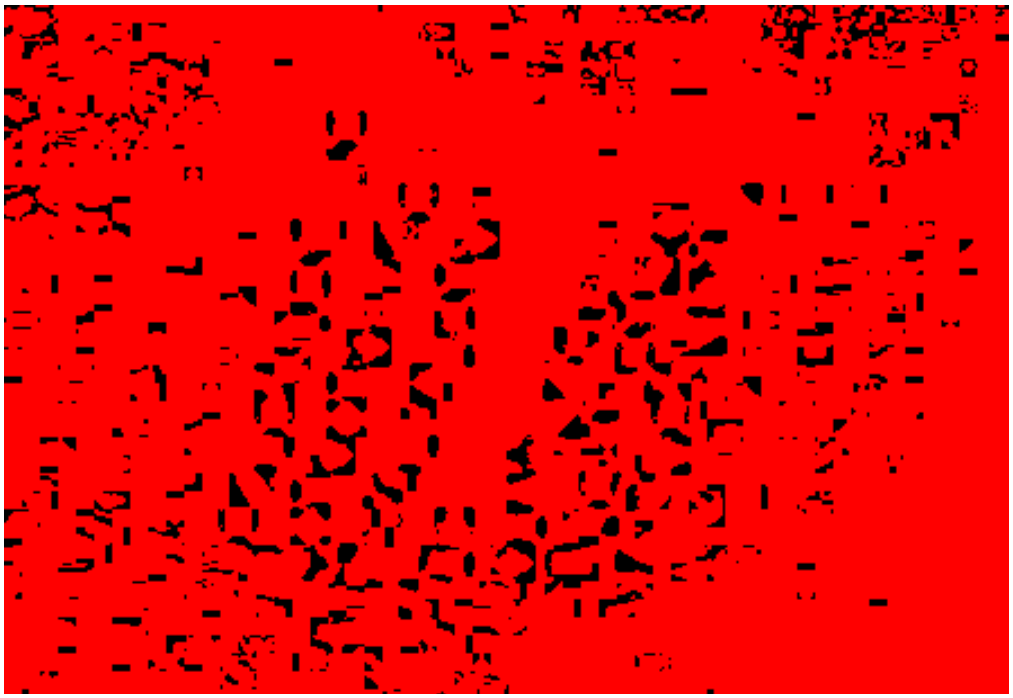


Figure 3-8: An example of a high precision grid as seen from Camera #2. The red dots indicate dots that were correctly identified and labeled. The blue dots represent the dot dropout in this image.

3.4.2.3 Estimation of 3D Projection

The 3D projections of the laser grid points are determined using the 2D pixel locations that were extracted in step using an implementation of the Multi-Camera Self-Calibration Toolbox in MATLAB [20]. This toolbox is designed to self-calibrate an array of cameras based on an input of 2D point correspondences from each camera. A key step in the self-calibration algorithm is determining the 3D projection of the calibration points. The present study utilizes this piece of the algorithm to estimate the projective depths of the laser dots based on the previously located 2D pixel locations from four camera viewpoints.

A detailed account of how the 3D projection algorithms work can be found in Svoboda et al. [20]. Only a brief introduction will be given in this paper.

The method utilizes the pinhole camera model extended to many points, (3):

$$W_s = \begin{bmatrix} \lambda_1^1 \begin{bmatrix} u_1^1 \\ v_1^1 \\ 1 \end{bmatrix} & \cdots & \lambda_n^1 \begin{bmatrix} u_n^1 \\ v_n^1 \\ 1 \end{bmatrix} \\ \vdots & \vdots & \vdots \\ \lambda_1^m \begin{bmatrix} u_1^m \\ v_1^m \\ 1 \end{bmatrix} & \cdots & \lambda_n^m \begin{bmatrix} u_n^m \\ v_n^m \\ 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} P^1 \\ \vdots \\ P^m \end{bmatrix}_{3m \times 4} \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}_{4 \times n} \tag{3}$$

where u and v are the pixel coordinates, λ is the projective depth of the point, P is the camera projection matrix, and X is the 3D real world coordinates. The self-calibration algorithms seek to determine the projective depths, λ, of each point so that (3) can be solved via Euclidean Stratification, yielding the camera projection matrix for the array of cameras.

The projective depths are estimated using a technique developed by Sturm et al. [20] [25]. First, one point from the whole data set is selected and the projective depth is set to 1. Then, image pairs, from the same time step but from different cameras, are iteratively selected according to commonly detected points. The epipolar geometry is robustly computed via the RANSAC-7 algorithm [20] [26] (see Figure 3-9). The projective depths of the rest of the points are estimated relative to this initial point using:

$$\lambda_p^i = \frac{\left(e^{ic} \times u_p^i\right) \bullet \left(F^{ic} u_p^c\right)}{\left\| e^{ic} \times u_p^i \right\|^2} \lambda_p^c \tag{4}$$

where e is the epipole, u is the point pixel location, and F is the fundamental matrix.

For this method, any point that is seen in at least two cameras can be projected into 3D coordinates. Having additional camera viewpoints is an advantage for at least two reasons. First, the likelihood that a given point will be seen in two cameras is increased. Second, the projective depth estimation will be more robust to outliers and provide a better 3D projection [20] [26]. The

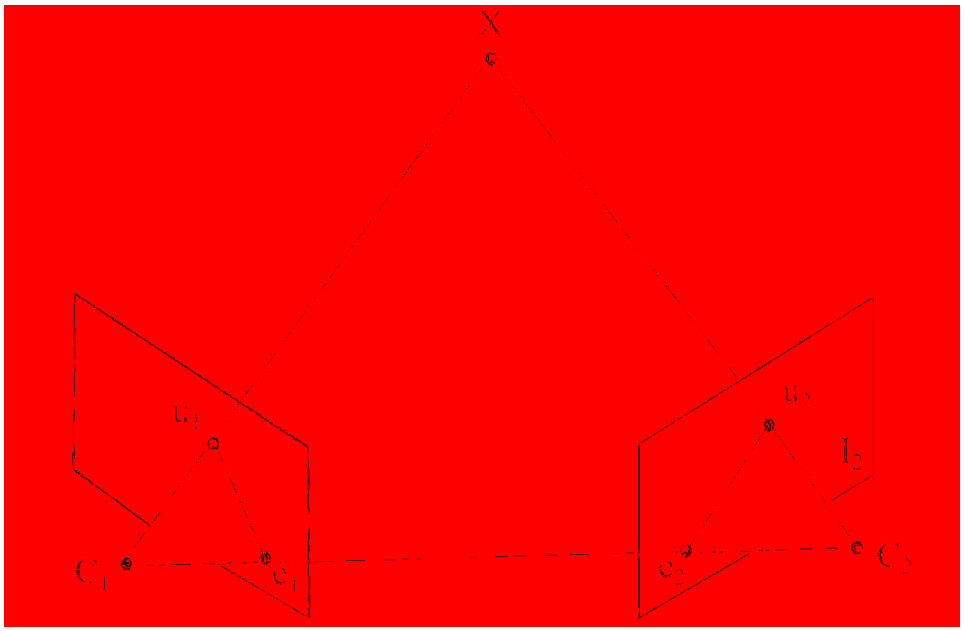


Figure 3-9: Epipolar geometry representation, where e is the epipole, u is the point pixel location, C is the camera center, and X is the object

algorithm utilizes a robust method of detecting and eliminating potential outliers, accurately projecting all of the points into 3D space, and calculating the 2D reprojection error for each camera.

The results are projected into an arbitrary coordinate system on an arbitrary scale. Some care has to be taken to align the results so that each time step is in the same frame of reference and can be compared directly. To do this a checkerboard calibration grid is imaged in the field of view of all four cameras before or after the experiment while all cameras are in their fixed locations. The corner points are located using auto correlation and are passed into the 3D projection algorithm along with the laser dot grid points. The resulting 3D projection can then be aligned to the calibration plane to obtain a known orientation and scale for the 3D projection. For the results presented herein a black and white checkerboard was used for the alignment (See Figure 3-10).

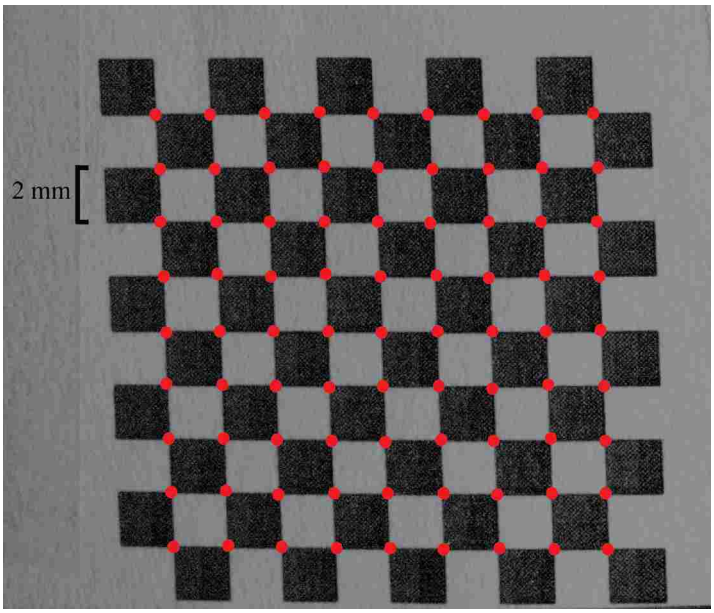


Figure 3-10: A 10 x 10 checkerboard calibration grid with 2 mm squares. Corner points are found using auto-correlation and used to align the 3D projection to a known scale and orientation.

3.5 Results and Discussion

3.5.1 Algorithm Performance

It was observed that some of the camera perspectives made clear identification of dots very challenging. In particular, cameras #1 and #4 were problematic. To solve these problems two problematic columns were omitted from each of these cameras. Every point in these omitted columns were still visible in the other three cameras. This method eliminated many false positives and misidentified points and still provided a valid 3D projection. As long as the points were identified in a minimum of two cameras, they could be appropriately projected into 3D coordinates.

The automatic 2D dot-locating piece of the algorithm was successful in finding an average of 91% of the dots with zero false positives, points that were labeled as dots when no dot was actually present. Table 1 contains a list of the average number of points found in each camera. This average was calculated by looking at a sequence of eight images, representing one

cycle of vocal fold oscillation. For the images from Camera #1 the algorithm had the lowest percentage of points found at 82%. For Camera #2 and #3 it had a percentage of 92 and 89, respectively, and for Camera #4 the algorithm was able to identify 100% of the points correctly.

The user input parameters for this experiment were intended to give the highest number of points found with the lowest number of false positives. Thus 9% of points were missed, mainly due to thresholding and morphology. The points were likely initially dim or initially small in size relative to the other dots in the array. This is where an alternative threshold relaxation method could improve performance of the algorithm, as described above. However, even with this threshold adjustment it is likely that 100% dot location could only be achieved through added user interaction, implemented herein. In our method the user is able to manually select points that were missed by the automatic detection and even correct false positives if they occur. With the added user interaction 100% of visible points can be found and passed on to the 3D projection.

Table 3-1: Summary of number of points found for all four cameras, the numbers represent an average over one phase of vocal folds oscillation (about 8 time steps). Search points represent the maximum number of points that were searched for in each image. Points found represents the average number of points found in each image. False positives represents the average number of points that were detected when no point was present in that location. The % found represents the average % of points that were found by each camera. On average, 91% of points were found in all cameras with zero false positives.

	Search points	Points found	False positives	% found
Camera 1	45	37	0	82%
Camera 2	63	58	0	92%
Camera 3	63	56	0	89%
Camera 4	38	38	0	100%
Total	209	189	0	91%

3.5.2 Error Estimation

The accuracy of the projection is evident from the 2D reprojection error. To calculate the 2D reprojection error the 3D points are projected back to 2D pixel coordinates and compared to the original 2D pixel position from which the 3D point was estimated. Figure 3-11 contains a chart that shows the 2D reprojection error for all four cameras when projecting 3D points from 12 consecutive time steps. The average 2D reprojection error for each camera ranges from 0.37 pixels to 0.49 pixels, with an overall average 2D reprojection error of only 0.45 pixels. The average resolution for the four camera viewpoints was 22 pixels/mm. A 2D reprojection error of 0.45 translates into an error of 20 μm. This level of 3D projection accuracy is similar to that seen by Luegmair et al. [15] who reported errors from 15-25 μm. The 3D projection accuracy of the current technique is a dramatic improvement when compared to depth-kymography, which is subject to errors of 50 μm [17].

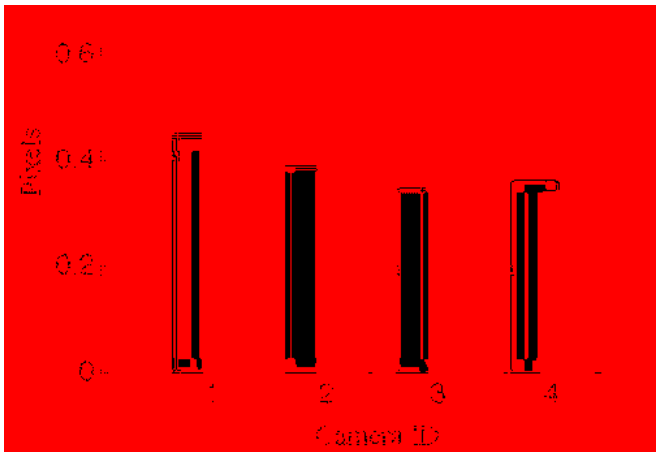


Figure 3-11: This bar chart shows the average 2D reprojection error of all four cameras at each time step. The error is averaged over 70 time steps. The red represents the mean reprojection error and the blue represents the standard deviation.

3.5.3 Experimental Results

In Figure 3-12 ten frames from Camera #2 are presented representing one full cycle of an excised human vocal fold oscillation, captured with the above-mentioned timing scheme. Next to each of the raw images is a corresponding 3D reconstruction. The 3D surface reconstruction was created using bi-cubic interpolation between the 3D projection points. The colors on the reconstruction represent the vertical elevation of the vocal folds for a given point as indicated by the colorbar.

The 3D reconstruction appears to mimic the expected behavior of the vocal folds. The vocal folds begin in a fully closed position as evidenced by a large valley region of low elevation. As air flows toward the vocal folds pressure begins to build and the vocal folds are forced apart. The 3D reconstruction shows a gradual increase in elevation as pressure builds upstream of the vocal fold and a more extreme increase in elevation as the vocal folds fully open to release the pressure. The vocal folds gradually return to their original position. It is clear from looking at the raw images that the right vocal fold has more dramatic motion than the left vocal fold. This phenomenon is confirmed by the 3D reconstruction. Most of the vocal fold action occurs on the side of the reconstruction corresponding to the right vocal fold.

In Figure 3-13 the 3D vocal fold reconstruction at t = 3.01 ms is considered in more detail. This time step corresponds to the closing phase of the vocal fold cycle. Three cross sections of the reconstruction are taken at ¼, ½, and ¾ of the overall y depth, S1, S2, and S3 respectively. A general tendency toward the center is observed for all three lines with a stronger tendency in areas where the vocal folds are nearly closed, S3 and S2. The area through which S1 passes is elevated due to the vocal folds still being in an open position.

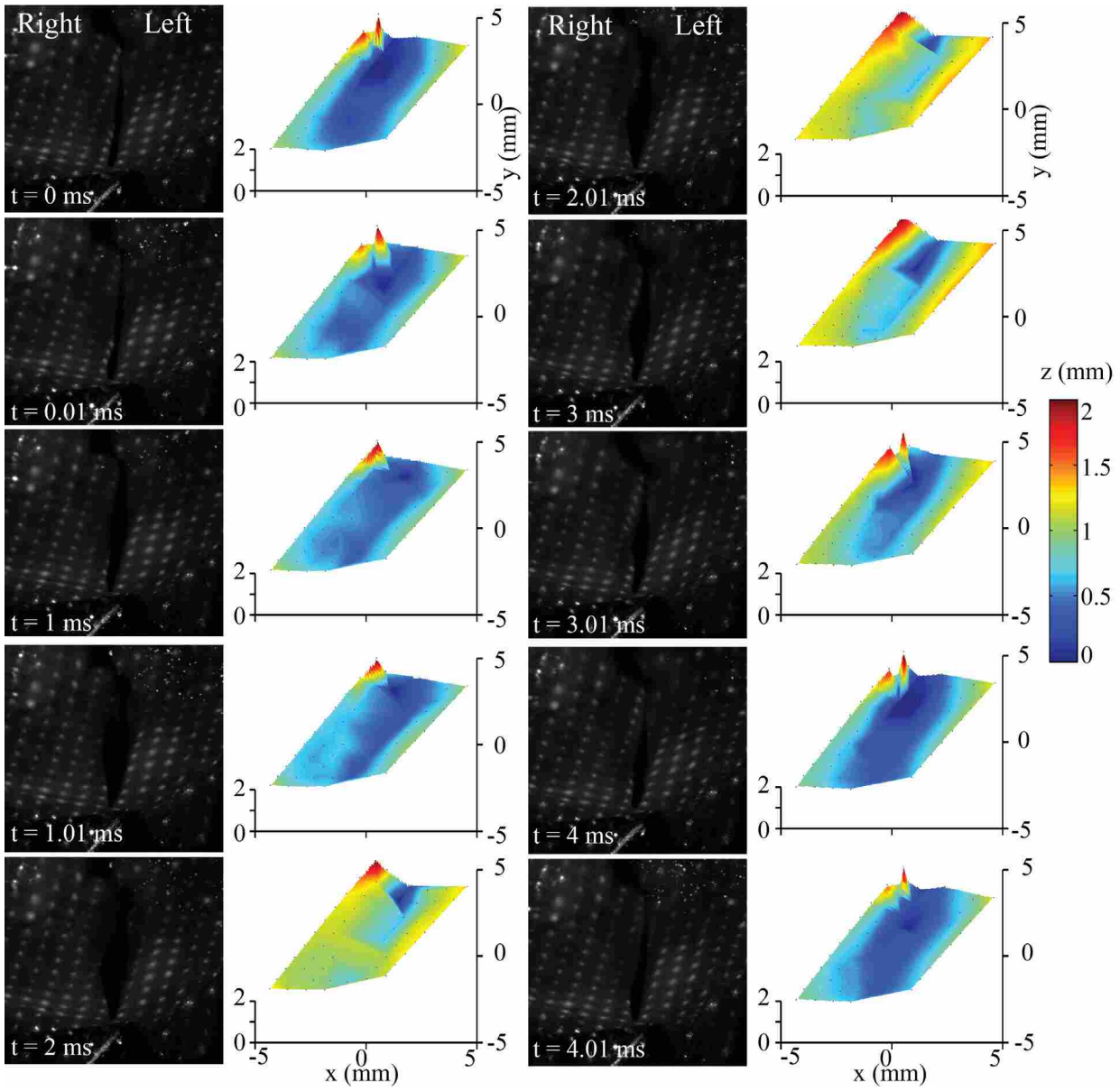


Figure 3-12: ***Left*** **- Ten raw images from Camera #2, representing one full cycle of vocal fold oscillation at approximately 250 Hz. The vocal folds begin in the fully closed position. As air pressure upstream of the folds builds they are force open until the pressure is released (t = 2 – 2.01 ms) and the vocal folds then return to their adducted position.** ***Right*** **– the 3D superior surface reconstruction created by bi-cubic interpolation between the 3D projection points.**

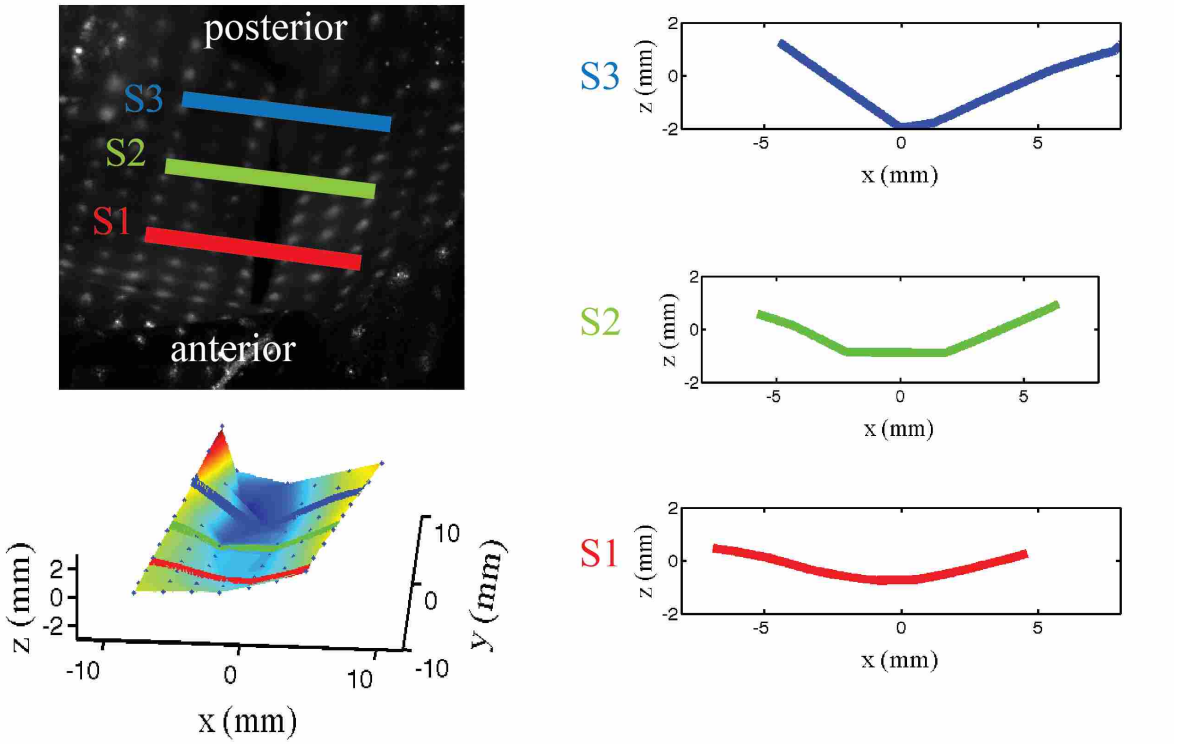


Figure 3-13: A closer examination of t = 3.01 ms from Fig 11. The raw image for this time step is marked with three lines, S1, S2, and S3 (top-left). The 3D reconstruction is shown (from Figure 3-11), marked with the three lines of interest (bottom-left). A 2D cross-section view of the three lines indicated by S1, S2 and S3. All three lines have a tendency toward the center. S3 had the strongest tendency toward the center, corresponding to the vocal fold nearly close at that point at seen in the raw image.

When the vocal folds vibrate, the tissue of the vocal folds is displaced with a wave-like motion. This wave is called the "mucosal wave". The mucosal wave is dependent on the structure of the vocal folds. Thus abnormalities in the vocal folds can be detected via abnormalities in the mucosal wave [17]. Reconstruction of the superior surface motion of the vocal folds allows for visualization of the mucosal wave. The mucosal wave is visible in this 3D reconstruction when viewed as a depth-kymograph (see Figure 3-14). A depth-kymograph measures the z-depth of one horizontal line across the vocal folds in time [17] [18] [15]. The depth-kymograph in Figure 3-14 was constructed across the line, S2, through eight consecutive cycles of oscillation. Eight peaks and troughs are visible in the kymograph corresponding to each cycle

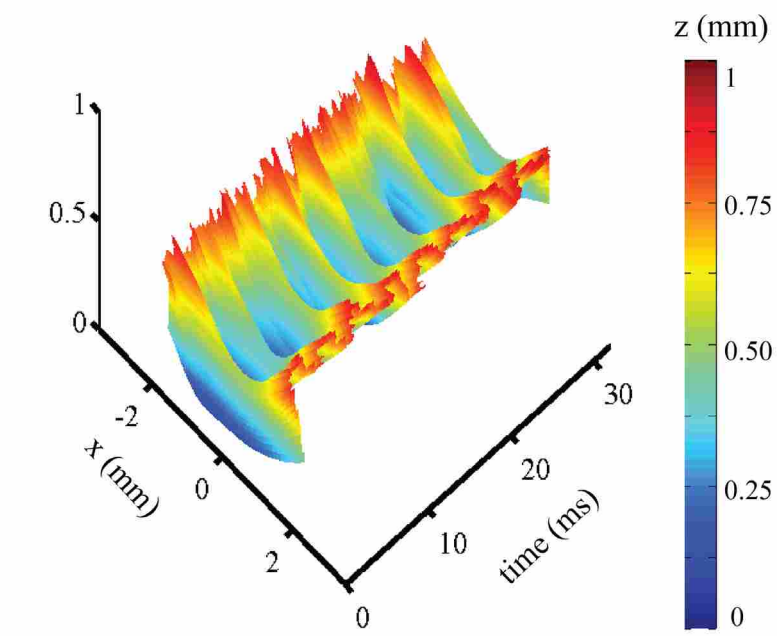


Figure 3-14: A depth-kymograph taken at line S2, showing a 3D view of vocal fold vibration over 8 cycles of oscillation. On the x-axis, zero corresponds to the mid-sagittal plane. Negative values correspond to the right vocal fold.

of oscillation. The crest of the traveling mucosal wave is visible in the green and yellow region [17] [15]. A threshold was set of 0.5 mm was set to locate the crest of the mucosal wave and can be seen for the right vocal fold in S2 in Figure 3-14. The frequency for the mucousal wave was calculated using the MATLAB FFT function. The frequency at S2 for both the left and right vocal folds was 254 Hz. This corresponds with the approximate period that is visible in the kymograph (T = 4 ms). Other mucosal wave properties were calculated at the S2 profile including: average lateral propagation velocity, horizontal amplitude, and vertical amplitude. The numerical values for these can be found in Table 3-2.

A distinct wave-like pattern is visible in the 2D kymograph view for all three lines, S1, S2, and S3, with depth indicated by color (Figure 3-14). The crest of the traveling mucosal wave causes this wave pattern [17] [15]. The crest of the mucosal wave is indicated on the kymograph

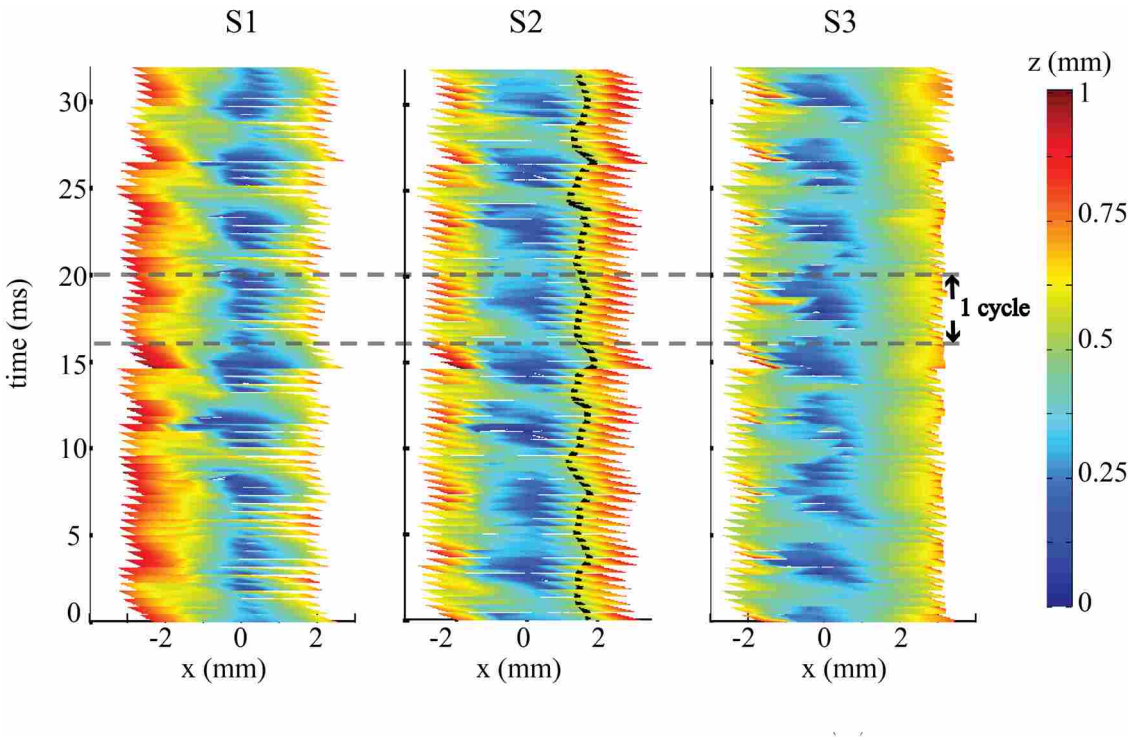


Figure 3-15: 2D view of a depth-kymographs for the three lines S1, S2, and S3, as indicated in Fig .12, for the reconstruction of the excised human vocal folds. The kymographs were constructed by tracking eight consecutive cycles of oscillation. The crest of the mucosal wave is visible in all three kymographs with varying degrees of asymmetry. The approximate position of the mucosal wave crest is traced with a dashed black line on the kymograph for S2.

Table 3-2: Calculated mucosal wave properties for the S2 cross-section. Frequency, average velocity, vertical amplitude, and horizontal amplitude are listed for both left and right vocal folds.

	Right Vocal Fold	Left Vocal Fold
Frequency (Hz)	254	254
Average velocity (m/s)	0.29	0.25
Max. vertical amplitude (mm)	0.49	0.50
Max. horizontal amplitude (mm)	1.0	1.0

for S2 by a dashed black line. The pattern is consistent with other depth-kymograph depictions of the mucosal wave [17] [15].

Asymmetry in the mucosal wave of the right and left vocal folds is evident in terms of both wave amplitude and phase. The kymograph for S3 particularly displays this asymmetry.

The crests of the right vocal fold's mucosal wave seem to fall within the troughs of the left vocal fold's mucosal wave. This observation reiterates the earlier visual observation of asymmetric oscillation.

Multiple sinkholes are visible (blue) in all three kymographs corresponding to approximately 2-3 ms into each cycle. From Figure 3-11, at 2 ms, it is observed that the vocal folds are in their fully open position. This behavior was also observed by Luegmair et al. [15].

3.6 Conclusions

A novel method for tracking the superior surface motion of a vocal fold model using multiple camera viewpoints was developed. This new method utilizes a continuous grid of laser dots, four high-speed camera viewpoints, and a computer vision algorithm to track the surface motion with minimal user interaction. The algorithm is easily adaptable to additional cameras.

The method utilizes an existing fluid imaging experimental setup to gather motion data at the same moment in time as fluid flow data. This coupling of flow and motion data has the potential to help gain a better understanding of the FSI during vocal fold vibration. However, because the images are also used as SAPIV data, the resolution of the oscillation cycle was limited to 1000 Hz, or for the specific presented here 4-5 image pairs per cycle. The technique could easily be used at higher frame rates and would yield a higher resolution tracking of the vocal fold motion.

The motion tracking algorithms effectively preprocess, locate 2D dot positions, and estimate 3D projections for a grid of points on the vocal fold surface. With an initial guess, based on user input, the automatic 2D dot location is able to find an average of 91% of the points with zero false positives found. The user can then identify missed points and return 100% of visible dots. The 3D projection then accurately estimates the 3D position for each of those points. The

3D projection functions with an average 2D reprojection error of 0.45 pixels or 20 μm, indicating a good 3D projection with sub pixel accuracy.

The 3D reconstruction of the superior surface motion of the excised vocal folds closely modeled the expected behavior of real vocal folds. Observations of the motion were consistent with those that have been found using other current 3D motion tracking techniques [15]. The mucosal wave was clearly visible across three different cross sections of the 3D reconstruction using depth-kymography.

This new technique for 3D surface reconstruction is an improvement when compared to the technique presented by George et al. [17] in that a full 3D reconstruction is obtained instead of just a point or a line across the vocal folds. With the full 3D reconstruction multiple profiles can be examined for the same moment in time. Also, the current technique has an improved 3D projection accuracy, 20 μm reprojection error, compared to 50 μm error for depth-kymography.

The current technique also has several advantages relative to the structured light technique presented by Luegmair et al. [15]. One key difference is the simplicity of the calibration for the multiple viewpoint technique. The cameras don't actually have to be calibrated but a calibration target is simply used as a reference to align the 3D projection to world coordinates by imaging the grid and extracting the corner points. The structured light approach requires a separate calibration for the camera and for the laser projection system; the calibration technique is described extensively in [15]. Another advantage of the multiple viewpoint approach is the increase in reprojection accuracy with the increase of camera number. Having four camera viewpoints allows for a high projective accuracy compared to a stereo pair (20 μm error compared to 25 μm error). This accuracy increases with the number of cameras used, reaching a maximum between 10-15 cameras [22]. Finally, the 3D projection algorithm

used for the current technique is robust to outliers. Thus, little to no post processing is required to view the 3D superior surface reconstruction. The structured light, stereo technique is more sensitive to noise and requires additional user post processing to remove outliers and smooth the vocal fold surface.

One drawback of the current technique is the cost of the experimental equipment. The technique requires the use of at least 3 high-speed cameras as well as the laser projection system. Whereas only one camera and the laser projection system are required for the structured light approach described by Luegmair et al. [15]. Another limitation of this technique is the up front user input. With multiple camera viewpoints, the amount of user input is multiplied by each camera added to the setup. This fact highlights the importance of increasing the automated accuracy of the computer algorithms used for this technique. Improvement of the automatic point location is a necessary area of future research.

In the future, this technique will be coupled with SAPIV to obtain simultaneous vocal fold motion and flow data. The effects and interactions of the fluid flow and the motion can then be investigated in hopes that a greater understanding of the fluid structure interactions of vocal fold vibrations can be gained.

Other potential future work includes a higher resolution reconstruction of the vocal fold oscillation cycle. If the experimental setup is designed to only track vocal fold motion, a higher camera frame rate can be used to gather the data.

Like the work presented by Luegmair et al. [15], this technique has potential to be miniaturized and utilized for in vivo vocal fold motion tracking. A lenslet array could be used to provide four unique viewpoints of the vocal folds in conjunction with a laser grid projection

system. This system could provide a tool for medical professionals to use in the diagnosis and treatment of vocal folds related disorders.

3.7 Acknowledgements

We would like to thank NSF grant CMMI #1126862 for funding the equipment and development of the synthetic aperture technique at BYU, and NIH/NIDCD grant R01DC009616 for funding SLT, and JRN and data relating to the vocal fold experiments and the University of Erlangen Graduate School in Advanced Optical Technologies (SAOT) for partial support of SLT.

4 WHOLE FIELD 3D CHARACTERIZATION OF THE GLOTTAL JET USING SYNTHETIC APERTURE PARTICLE IMAGE VELOCIMETRY

4.1 Contributing Authors and Affiliations

Joseph R. Nielson, D. Jesse Daily, Scott L. Thomson, Tadd T. Truscott
Department of Mechanical Engineering, Brigham Young University, Provo, UT 84602

Georg Luegmair, Michael Döllinger
Department of Phoniatrics & Pediatric Audiology, University Hospital Erlangen Medical School, Erlangen, Germany

4.2 Abstract

Synthetic aperture particle image velocimetry is used with an excised human vocal fold model to study the airflow over the vocal folds during voice production. A whole field, time-resolved, 3D description of the flow is presented over multiple cycles of vocal fold oscillations. The 3D flow data is synced in time with a 3D reconstruction of the superior surface of the vocal folds, as well as, the subglottal pressure wave.

4.3 Introduction

Many recent studies have focused on characterizing the fluid structure interactions (FSI) that occur during vocal fold oscillation. This oscillation occurs as air from the lungs flows past the vocal folds. In preparation for speech, muscles bring the vocal folds together, adducted in a closed position. As air flows towards the vocal folds, pressure builds in the subglottal region until the vocal folds are forced open and air flows through the glottis. The area of the glottis

increases, the velocity of the air increases, and the pressure decreases, allowing the vocal folds to close again. This process continues as air continues to flow through the larynx (See Figure 4-1). The air that passes between the vocal folds, the glottal jet, has been the focus of recent studies because it is the primary source of sound during speech. A better understanding of the glottal jet and its interactions with the vocal folds will yield a better understanding of voice production. A full characterization of the FSI of vocal fold vibration can benefit the development of computational and analytical vocal fold models. The end goal is the improvement of treatment and prevention of vocal fold related voice disorders [2].

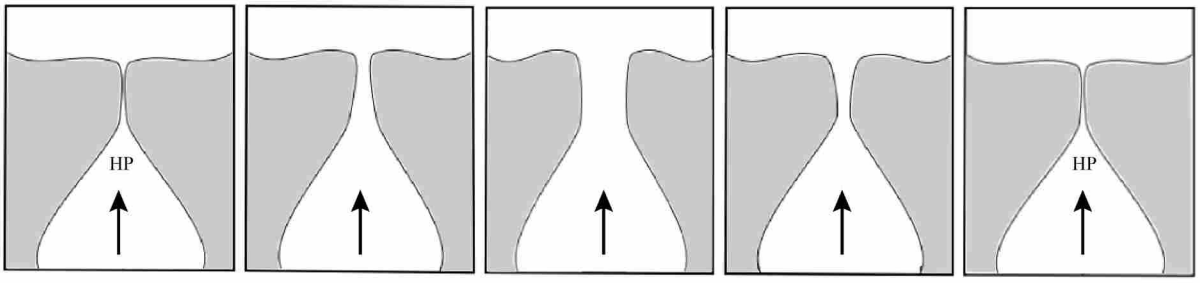


Figure 4-1: Schematic of a cycle of vocal fold flow induced vibrations. The vocal folds are shown in each image sequence as gray structures. The cycle starts at the left with the vocal folds in their adducted position. As high pressure builds in the subglottal region the vocal folds begin to open. As the air forces the vocal folds apart, *middle,* the pressure decreases and the velocity increases. Finally, as the vocal folds return to their closed position the pressure increases again, *right*.

In order to characterize the flow dynamics important for speech production, voice researchers have sought a better understanding of the glottal jet. Visualizing the glottal jet *in vivo* is very challenging due to the vocal folds location deep within the larynx. Thus different *in vitro* methods have been developed for investigations of the glottal jet. These include experimental studies using synthetic (static, driven, and self-oscillating) and excised models, as well as computational studies.

Many studies have used particle image velocimetry (PIV) as a method for visualizing the glottal jet dynamics. Two-dimensional PIV has been used extensively with various synthetic and excised vocal fold models. Valuable information about glottal jet dynamics has been gained from these 2D studies and confirmed results from computational models, including: jet axis switching [2] [8] [3] [1], flow separation [2] [3] [1] [4] [29] [30] [31], vortices [2] [8] [3] [1] [4], jet flapping [2] [3] [3] [8], and vena contracta [3] [8].

Due to the highly three-dimensional nature of the glottal jet, recent studies have focused on resolving the flow field in three dimensions. Triep et al. [9] used traditional 2D PIV at various slices of a 3D volume downstream of the vocal folds. A cam driven synthetic vocal fold model was used with water as the working fluid. The time resolved flow field was reconstructed with each 2D slice phase averaged. This study, in addition to confirming results from previous 2D studies, highlighted the 3D and unsteady nature of the glottal jet and stressed the need for a whole field 3D study of the jet. Krebs et al. [10] used stereoscopic PIV, to resolve a third component in each slice of the 3D volume. Again, each slice was phase averaged to reconstruct the 3D volume. The study examined, in more detail, the 3D nature of axis switching and noted that the switching does not occur at a fixed distance from the vocal folds and is connected to the flow separation. Despite these recent studies of the 3D flow field there do not appear to be any whole-field, time-resolved, 3D glottal jet descriptions in the voice research literature.

In the present study synthetic aperture particle image velocimetry (SAPIV) was used with excised human vocal folds to reconstruct the glottal jet flow fields. SAPIV is a technique to visualize flow fields in three dimensions. An array of synchronized high speed CCD cameras is used to image the particle field of interest. The images are then digitally refocused using synthetic aperture image processing algorithms. The result is a complete focal stack of the region

of interest. For a given plane in the focal stack, particles in that plane will appear sharply in focus (high intensity) and particles out of plane will appear out of focus (low intensity). These images are thresholded so that only particles in a given plane are visible. From this point traditional 3D PIV algorithms can be used to calculate the velocity field of the flow. Having multiple camera viewpoints allows for 3D resolution of flow fields with high seeding densities and partial occlusions [22]. The result of the SAPIV experiment is a whole-field, time-resolved, 3D characterization of the glottal jet over multiple consecutive cycles of oscillation [23]. For a detailed description of the methodology of SAPIV see [22].

Another subject that has been a focus of study for an increased understanding of the FSI of vocal fold oscillation is the motion of the vocal folds. The superior surface of the vocal folds can easily be seen in vivo with the use of an endoscope. This technique is currently used in the diagnosis and treatment of vocal fold related disorders. However, getting quantitative data is not easy because it is uncomfortable and unnatural for the patient and the optical access is limited. Thus, just as is the case with glottal jet studies, experimental studies are often performed *in vitro* [12] 7] [14] [15].

The technique outlined by Luegmair et al. [15] is used herein to reconstruct the 3D superior surface motion of the vocal folds simultaneously with the SAPIV flow data. The result is a 3D characterization of the vocal fold shape and motion at the same moment in time as the 3D flow field characterization.

4.4 Methods

4.4.1 Experimental setup

An excised human vocal fold from a 81 year old female was prepared by removing excess cartilage and tissue surrounding the vocal folds and mounted on a 16 mm diameter

stainless steel pipe (trachea tube) and clamped to ensure a tight seal (see Figure 4-2). Two screws were used to secure and stabilize the larynx in the mounting. The vocal folds were tensioned by hanging a 20 g weight in the front and a 10 g weight in the back. The trachea tube was connected to a compressed air reservoir and the flow rate was adjusted to the minimum needed to induce vibration in the vocal folds (4 L/min), controlled by a custom LabVIEW program.

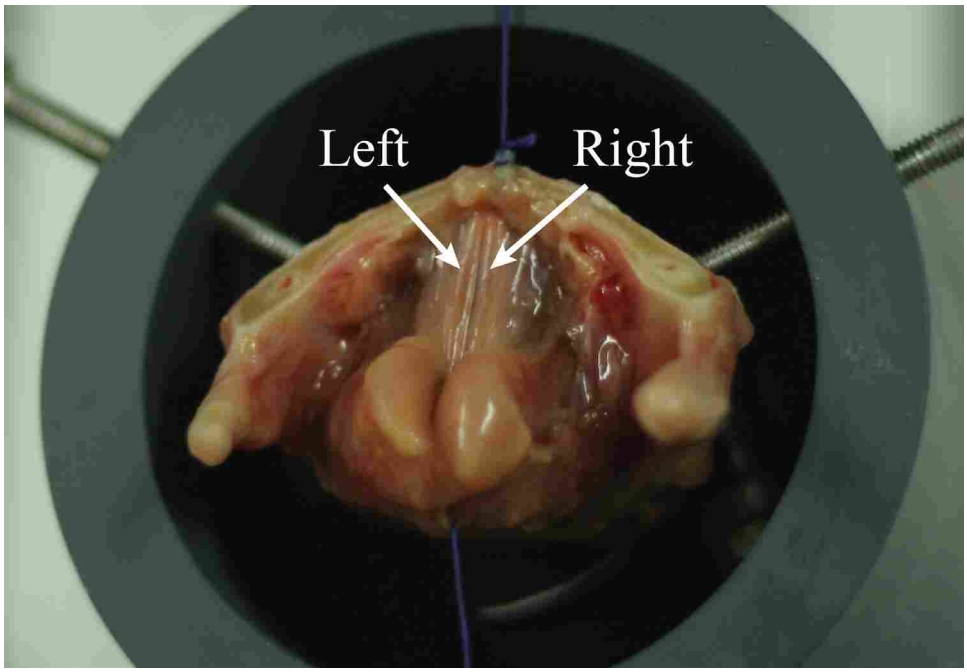


Figure 4-2: Excised human vocal folds used in this study. The vocal folds are adducted and held in place via sutures. Weights are attached to the front and back of the model to apply tension. The model is stabilized and held in place by two screws.

The physical dimensions of the vocal folds were measured prior to the experiments. The length of the vocal folds was 12 mm. (see Figure 4-3). In their adducted position the vocal folds have a small gap between them. At its widest point the gap measured approximately 1 mm (see Figure 4-3). A slight bowing of the vocal folds caused the gap, with the right fold having a slightly more significant bow than the left fold.

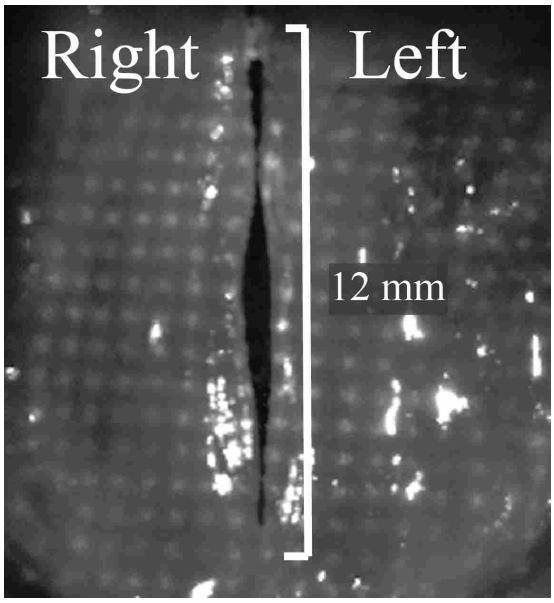


Figure 4-3: Superior surface of the excised vocal folds. For this particular set of vocal folds, in their adducted position with no airflow, they have a small gap between them caused by a slight bowing of the vocal folds.

A pressure probe (MPXV7002) was connected to the trachea tube with a 49 cm tube, 12.5 cm upstream of the vocal folds, to measure the subglottal pressure. This pressure wave measurement was used to synchronize the motion and flow data.

The airstream was seeded with hollow polymer microspheres for the particle image velocimetry measurements. The particles had a mean diameter of 40 μm and density of 25 kg/m3 (Expancel, 461 DET 40 d25). Flow measurements of the particle field were taken in the region immediately above the surface of the excised vocal folds.

The particles were illuminated by a double-pulsed Darwin-Duo laser system (Quantronix, Nd:YLF, 1000 Hz). The delay between the laser pulses was set to 10 μs. A beam expander (Edmund Optics, 8x) was attached to the Darwin-Duo laser to illuminate a volume ranging from the vocal fold surface to approximately 30 mm above the surface. The first 1-2 mm above the vocal folds surface was not illuminated by the laser because the cartilage surrounding the vocal

folds obstructed the laser beam. The laser beam was directed toward the vocal fold model oriented perpendicular to the sagittal plane (Figure 4-4).

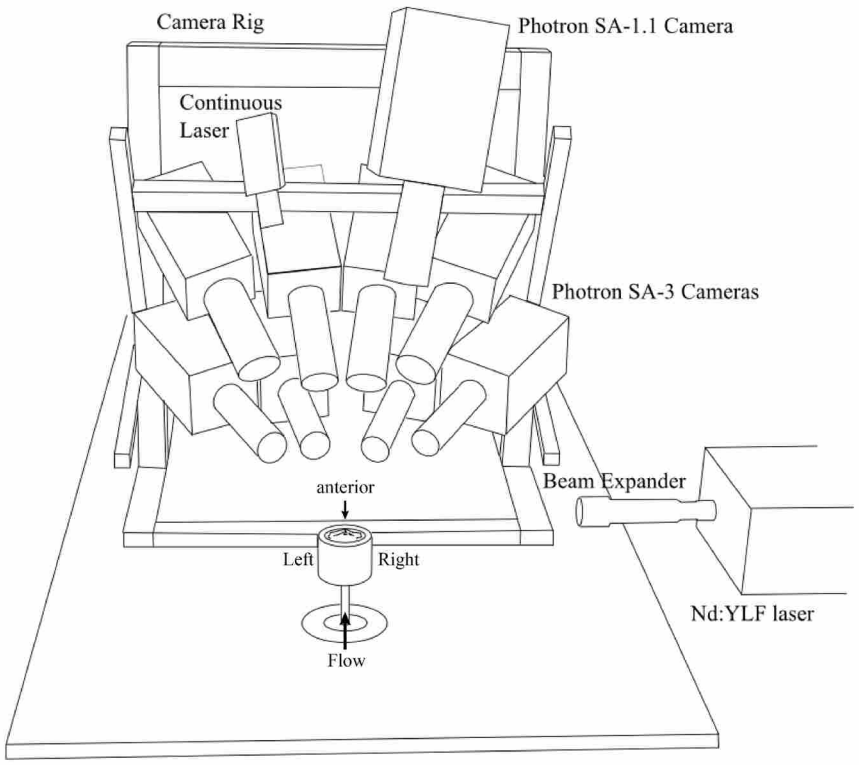


Figure 4-4: Diagram of the SAPIV setup. Eight high-speed cameras (Photron SA-3) are focused on the volume above the vocal fold model. Another camera is mounted above for motion tracking. A high-speed Nd:YLF laser, on the right, is used to illuminate the particles in the glottal jet.

The SAPIV measurements utilized eight FASTCAM SA3 high-speed cameras (Photron, 2000 fps, 640 x 640 pixels). The cameras were mounted on a custom camera rig with four cameras on the top row and four cameras on the bottom row (see Figure 4-4). The cameras were aimed and focused on the region extending 20 mm above the vocal fold superior surface [23].

The motion of the superior surface of the vocal folds was detected using optical stereo triangulation. For this purpose, a ninth high-speed camera (Photron SA-1.1, 5000 fps, 1024 x 112 pixels) and laser projector system were added to the setup. The laser system consisted of a

continuous laser (528 nm) and a micro lens array, which projected the laser beam as a grid of dots (see Figure 4-3). Both the camera and the laser were positioned superior to the vocal fold surface and slightly off to the frontal plane of the larynx (as seen in [15]). The method for motion tracking was based upon the work presented by Luegmair et al. [15] and a brief summary of the technique can be found in section 4.4.2.2.

In order to capture motion and flow data simultaneously all nine high-speed cameras were synced using a master pulse generator signal (BNC, Model 575). A timing scheme was designed such that the double-pulsed PIV laser (1000 Hz) was not visible in the motion tracking images. The motion-tracking camera captured images at 5000 fps. In order to avoid the PIV laser, the shutter speed was set to 1/12000 that caused a slight delay between the motion data and the flow measurements (13 μs).

4.4.2 Data Processing

4.4.2.1 SAPIV

The cameras were calibrated using a process based on the multi-camera self-calibration method presented in [20]. A checkerboard target was placed within the view of all eight cameras and images were taken at 4 different planes throughout the volume of interest. The checkerboard intersection points were found using an auto-correlation method. Using these points, the cameras were self-calibrated using epipolar geometry relationships. The result of the calibration was a camera projection matrix with intrinsic and extrinsic parameters for each camera.

It has been well established that there is a need for preprocessing for volumetric PIV [22]. The preprocessing is able to reduce the effect of differences between laser pulse intensity and to remove background noise. The specific preprocessing technique used was based on the approach outlined by Belden et al. [22]. The approach consists of the following steps:

(i) Subtract sliding minimum (10 pixel window)

(ii) Blur the image by convolution with 3x3 Gaussian kernel

(iii) Equalize the image histograms to the histogram of the image with the highest contrast

(iv) Increase the contrast by trimming the bottom and top 0.1% of intensity values

(v) Subtract sliding minimum (10 pixel window)

The SAPIV data processing was completed using the procedure outlined by Belden et al. [22] and Daily et al. [23]. The synthetic aperture refocusing algorithms utilized the calibrated camera projection matrix to combine the images from all eight cameras into a focal stack. The focal stack consisted of 128 planes, dividing 18 mm of volume depth into increments of 0.14 mm. One image plane from the focal stack is shown in Figure 4-5. Particles that lie on that particular plane appear sharply in focus (high intensity) while particles that do not lie on that plane are out of focus and low intensity. To extract only the particles that lie on a given plane the focal stack images were intensity thresholded. The threshold was set at >35/255 for images corresponding to the first laser pulse and >40/255 for images corresponding to the second laser pulse. The result was a focal stack containing only the particles sharply in focus for each plane (see Figure 4-5).

The focal stack of the particle field was processed using a 3D version of the matPIV algorithms [22] [32]. The 3D matPIV algorithms determined the 3D velocity fields. The 3D PIV interrogation window size was set to [64, 64, 32] for the first pass and [32, 32, 16] for the second and third passes, with 50% overlap.

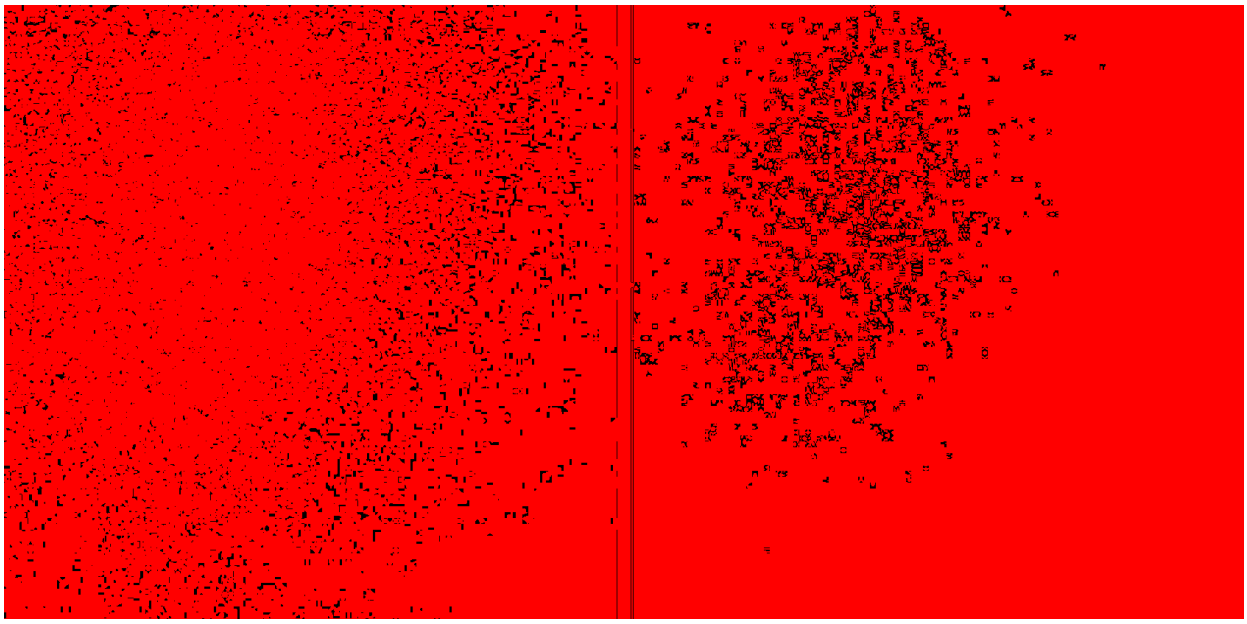


Figure 4-5: Left – One refocused image from a focal stack of 128 images. Warping images from eight high-speed cameras onto a single plane created the refocused image. The particle field on that plane appears as high intensity points in the image. Low intensity points are out of focus particles located on other focal planes. Right – The same image with an intensity threshold of 35/255 applied to the image. Only high intensity points remain, representing particles on that focal plane.

4.4.2.2 3D Motion Tracking

The 3D motion of the vocal folds' superior surface was tracked using structured lighting [15]. The optical reconstruction setup was calibrated by the process described by Luegmair et al. [15]. This process consists of two steps: First, the camera calibration method described by Zhang [33] is applied to the camera. This method utilizes multiple images of a planar calibration target, taken at different orientations and exploits the homography to obtain a calibration. Second, the laser projector's position and opening angle are identified using a similar method, exploiting the homography of a planar calibration target in multiple images.

The 3D surface reconstruction consists of two steps. First, the pixel coordinates for each laser dot in the image are identified. With these pixel locations a trajectory can be calculated from stereoscopic correspondence. Second, the trajectories need to be related to corresponding laser rays from the projector. For this purpose, the frames of the high-speed footage were

processed semi-automatically, allowing the segmentation of the laser dots' entire trajectories. The assignment to a laser ray is done manually, allowing for the reconstruction of the superior vocal fold surface.

4.4.2.3 Subglottal pressure

The pressure data was recorded with a custom LabVIEW program at a rate of 35 kHz. A master trigger signal was recorded as well to indicate when the SAPIV and motion data capture began. Using this trigger signal, the section of the pressure wave that corresponds with the SAPIV and 3D motion data was extracted.

The pressure probe was connected by a 49 cm tube to the trachea tube, 12.5 cm upstream of the vocal folds. To obtain a pressure reading for the subglottal region a delay was calculated using the following equation:

$$D = \frac{d}{c} \tag{1}$$

where D is the delay, d is the distance between the pressure probe and the vocal folds, and c is the speed of sound in air (343.2 m/s). A distance of 61.5 cm corresponds to a delay of 1.79 ms. This delay was considered when syncing the subglottal pressure with the flow and motion data (Figures 4.6 and 4.7).

4.5 Results and Discussion

The flow data was sampled at 1000 Hz, thus the glottal jet is reconstructed in time with 1 ms between PIV image pairs (See Figure 4-6). The subglottal pressure wave is plotted for two consecutive cycles of interest, with the times at which an SAPIV reading was taken marked with a red dot. A discrete Fourier transform of the pressure signal over 100 cycles, via the fast Fourier

transform function in MATLAB, revealed a pressure wave frequency of 350 Hz. A time resolution of 3 samples per cycle is shown for two consecutive cycles of oscillation. For both cycles, the samples are taken when the vocal folds are in: the closed phase (minimum glottis area), at opening phase (increasing glottis area), and closing phase (decreasing glottis area). This can be observed in the high-speed images of the vocal folds (see Figure 4-6).

The 3D glottal jet for an excised human vocal fold model is reconstructed over the course of two cycles in Figure 4-6. The data is presented with synchronized superior surface motion and subglottal pressure. As mentioned above, the high-speed images are actually delayed from the flow measurements by approximately 13 μs. For an oscillation frequency of 350 Hz. 13 μs represents 0.45% of a full cycle. Thus we the motion and flow data will be treated herein as if they occurred at the same moment in time.

The pressure signal at t = 0 ms is in the rising peak of its cycle; this corresponds to the vocal folds in their closed phase. When the vocal folds are opening (t = 1 ms) the pressure is decreasing and it begins to increase again as the folds are closing (t = 2 ms). This behavior matches with intuition of what would be expected from the pressure driven vibrations.

The 3D superior surface motion of the vocal folds is displayed in Figure 4-6 as a depth map profile for three lines S1, S2, and S3. The general shape of the vocal folds is clearly visible in each of the three cross sections. When the vocal folds are in their closed phase (t = 0 ms, t = 3 ms) the 2D profiles are relatively flat with a slight narrowing of S3. There is a very noticeable difference when the folds are opening (t = 1 ms, t = 4 ms) because both the right and left vocal folds have elevated and a large gap, with a flat bottom (glottis), appears between the two folds. As the vocal folds move into the closing portion of the phase (t = 2 ms, t = 5 ms), the folds reach their highest vertical elevation and a very steep gap remains between the folds. Across these two

cycles of oscillation S1, S2, and S3 have very similar behavior. S3 has the largest overall vertical amplitudes and has a more dramatic transition from the folds into the glottis.

2D slices of the 3D velocity stack are used to examine the reconstructed glottal jet. First three coronal slices are considered. Slices are taken about the mid-coronal plane and 2.25 mm on either side of the mid-coronal plane, indicated in the top-left of the figure by S1, S2, and S3. A series of six velocity vector fields for each slice are plotted along with a contour color map. The orange in each plot represents the approximate position of the vocal folds.

The glottal jet evolution is clearly visible in both cycles. Little to no jet is visible at t = 0 ms, a well establish, high velocity (30-40 m/s) jet appears in the second time step, and a diminishing jet in the third. The velocity fields reveals a similar jet at the corresponding phases for each cycle (t = 0 ms with t = 3 ms, and t = 1 ms with t = 4 ms, t = 2 ms with t = 5 ms.).

In general the velocity vectors are aligned with the downstream axis. Across two cycles of oscillation there does not appear to be any lateral deflection of the jet. The general shape of the jet matches with what has been reported in other 3D studies with the jet beginning narrow in the lateral direction and expanding out as it travels downstream [9] [10].

Across all of the velocity fields the jet velocities are mostly in the range of 10-20 m/s, with maximum velocities reaching 30-40 m/s. These velocities are similar to what has been reported in other glottal jet studies. This velocity range also matches closely with the expected theoretical instantaneous velocity from Bernoulli's equation:

$$V = \sqrt{\frac{2*P}{\rho}} \tag{2}$$

where P is the pressure drop and ρ is the density of air. Using a pressure drop of 0.85 kPa (from Figure 4-7) and a density of 1.2 kg/m^3, the predicted velocity is 37.6 m/s.

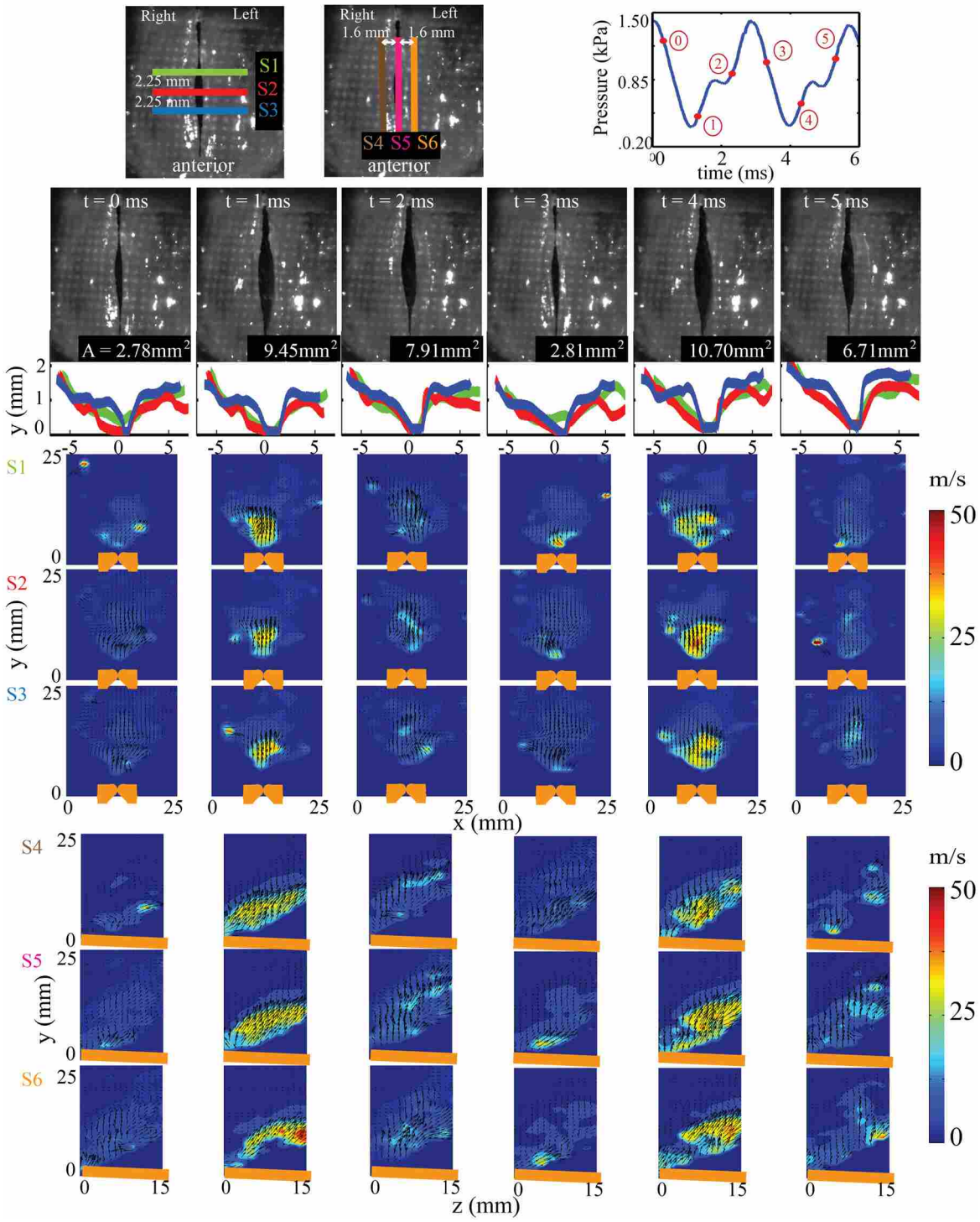


Figure 4-6: The 3D reconstruction of the glottal jet is viewed at three 2D cross sections in the coronal direction and three 2D cross sections in the sagittal direction (*upper left*). The evolution of the jet is visible and is correlated to the phases of vocal fold oscillation according to high-speed images of the vocal folds taken at the same moment in time (*upper right*). The 3D motion data is plotted as three surface profile lines in 2D, corresponding to S1, S2, and S3 (*upper left*). The subglottal pressure wave is plotted for the two cycles of interest and the pressure for each corresponding time step is plotted (red dots). The approximate position of the vocal folds is indicated in the bottom of each plot in orange.

This value is close to the range seen at t = 1 ms and t = 4 ms in the yellow and orange regions (Figure 4-6). The velocity fields are similar at each of the three coronal cross-sections, for a given time step, both in terms of jet shape and velocity magnitudes.

Next, three sagittal slices are considered. Slices are taken about the mid-sagittal plane and 1.6 mm on either side of the mid-sagittal plane, indicated in the top-left of the figure by S4, S5, and S6. Again, a series of six velocity vector fields is plotted for each of the slices. The orange area at the bottom of each plot indicates the approximate position of the vocal folds.

The *bottom right* corners of the sagittal plane cross sections are void of velocity vectors. This is due, in part, to the challenge of utilizing an excised human vocal fold model. The laser light was partially blocked by cartilage on the larynx and thus a small piece of the jet is omitted in that region.

The jet evolution is visible just as in the coronal planes. The sequence begins (t = 0 ms, t = 3 ms) with little to no flow when the vocal folds are in their closed phase, reaches a maximum when the folds are opening (t = 1 ms, t = 4 ms), and begins to disperse as the vocal folds close again (t = 2 ms, t = 5 ms).

The direction of flow in the sagittal cross sections is upward in the positive y direction with an incline toward the positive z direction, nearly 45 degrees at t = 1 ms and t = 4 ms. This incline is likely due in part to the way the vocal folds were oriented within the larynx. Often in excised larynges, as was the case with this model, the vocals can be recessed into the larynx and even oriented slightly askew.

The same velocity magnitudes, as seen in the coronal cross sections, views are observed in the sagittal view with maximum velocities occurring when the vocal folds are near their fully

open position. The maximum velocities occur at locations longitudinally near the middle of the vocal folds.

The behavior of the glottal jet appears to be very similar when comparing the two cycles. Keeping in mind that there is a slight phase shift between cycles, in general the velocity magnitudes and directions are consistent. Jet evolution across the different cross sections also appears to be consistent between the two cycles.

A more complete examination of the glottal jet evolution is obtained by piecing together a full cycle of oscillation from multiple cycles of time resolved data. Figure 4-7 shows one full cycle of oscillation, consisting of 16 data points reconstructed from 8 different data cycles. Each row represents one moment in time with 0.19 ms between each time step. Each row contains a plot of the subglottal pressure wave with a red dot indicating the current pressure, a high-speed image of the vocal fold's superior surface, a plot from the 3D motion tracking showing three depth profiles, and two slices of the glottal jet velocity stack (mid-coronal and mid-sagittal). In the velocity field plots the approximate position of the vocal folds is indicated in orange.

The high-speed images show the vocal fold behavior over the full cycle. The area of the glottis is calculated by tracing the 2D outline of the glottis. The area for each time step is indicated at the bottom-right of each high-speed image. The folds begin in their adducted position (t = 0 ms); only a small gap ($A < 3mm^2$) between the vocal folds is visible due to imperfections in these particular vocal folds. As pressure beneath the vocal folds increases the folds begin to open. At t = 1.33 ms the vocal folds reach their point of largest area, this corresponds to the vocal folds being completely open. With the pressure relieved the folds then begin to return to their equilibrium position. At the end of the cycle the vocal folds are in the same position that they started in (t = 2.85 ms).

More specific information about the vocal fold motion is obtained from the depth profiles plotted in Figure 4-7. The vocal folds begin in a relatively flat position (t = 0 ms) and are uniform across all three cuts (S1, S2, and S3). Immediately after the oscillation cycle begins movement of both the left and right vocal folds is visible (t = 0.19 ms, t = 0.38 ms). The folds separate and a relatively large gap, with a flat bottom, appears between them. Over the course of the cycle the right vocal fold appears to have more dramatic motion than the left vocal fold. Both vocal folds reach the maximum vertical amplitude when the vocal folds are fully open (t = 1.33 ms).

The evolution of the glottal jet is visible in 2D cross sections at the mid coronal and mid sagittal planes. As the vocal folds begin to open the glottal jet begins to emerge from between the vocal folds (t = 0.38 ms) and is fully evolved around the point when the vocal folds are fully open (t = 1.33 ms). Throughout the cycle the jet remains aligned with the downstream axis from the mid-coronal view. The evolution and disintegration of the glottal jet appears to occur smoothly. The jet steadily increases in size and then decreases at relatively the same rate. There is not an abrupt disturbance of the jet as the folds close. This is in line with what was observed by Triep et al. [9].

In the sagittal plane cross sections (S1, S2, and S3) the velocity vectors are directed upwards (positive y) and also slightly forwards (positive z). The jet appears to become more narrow in the longitudinal direction as it progresses further downstream of the vocal folds.

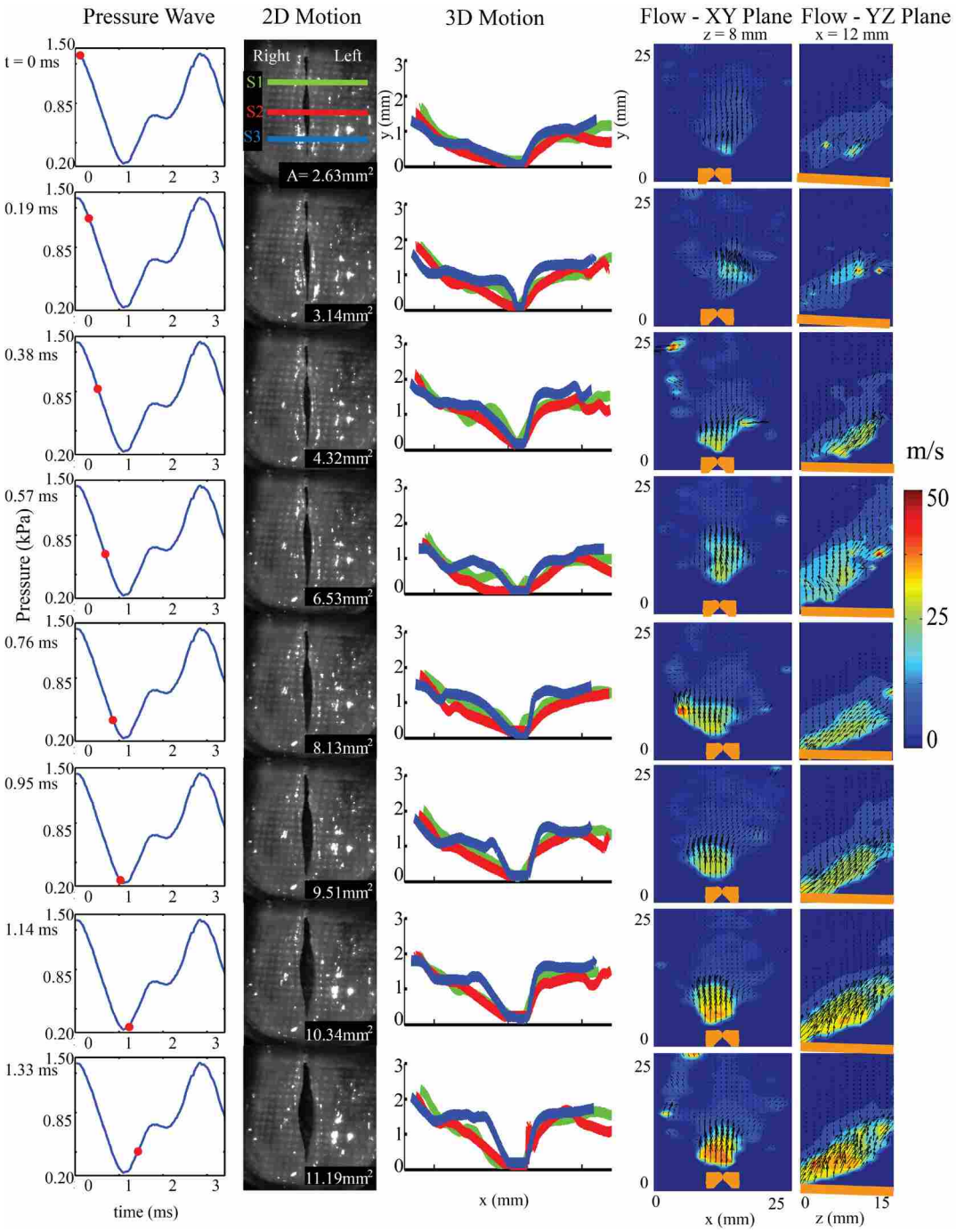


Figure 4-7a: A compilation plot showing synced data, from left to right: subglottal pressure, high speed camera image of 2D vocal fold motion, three sagittal cross section depth maps, and glottal jet data from slices of the 3D velocity stack in two directions. Each row represents one moment in time. 8 time steps are assembled to show half of a complete cycle of oscillation. The subglottal pressure wave is plotted in each time step with a red dot indicating where that moment in time falls on the pressure wave. The 2D vocal fold motion is represented by raw images captured by the camera that was oriented superior to the vocal folds. The 3D motion is depicted as three profile lines indicated in upper-left. Two velocity field plots are shown, left - one cut through the XY plane located approximately in the center of the jet, z = 9 mm. Right - one cut through the YZ plane approximately in the center of the glottal jet, x = 10 mm.

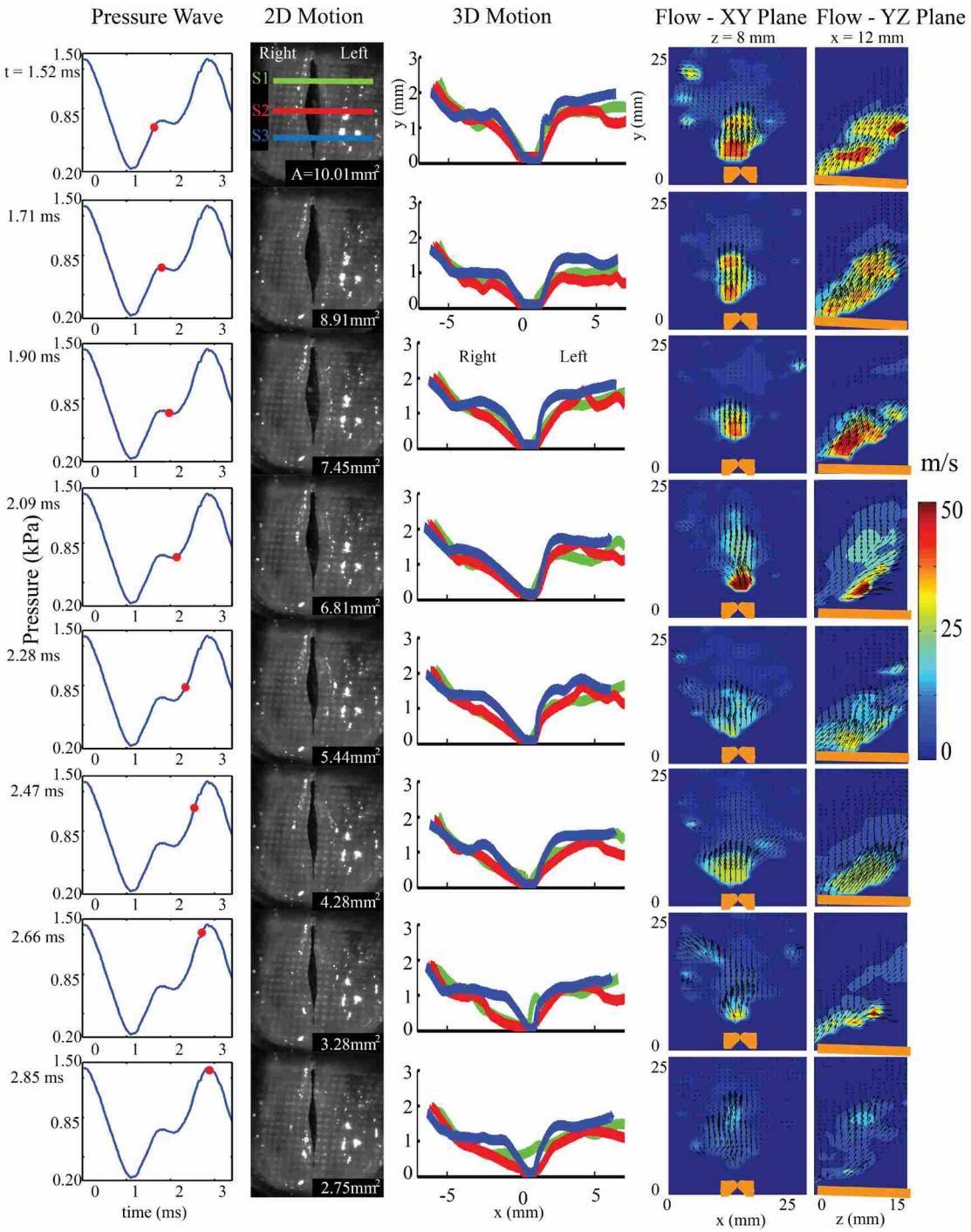


Figure 4-7b: A continuation of the plot from Figure 4-8a. Each row represents one moment in time. 8 time steps are assembled to show the second half of a complete cycle of oscillation. The subglottal pressure wave is plotted in each time step with a red dot indicating where that moment in time falls on the pressure wave. The 2D vocal fold motion is represented by raw images captured by the camera that was oriented superior to the vocal folds. The 3D motion is depicted as three profile lines indicated in upper-left. Two velocity field plots are shown, left - one cut through the XY plane located approximately in the center of the jet, z = 9 mm. Right - one cut through the YZ plane approximately in the center of the glottal jet, x = 10 mm.

The velocity of the jet continues to accelerate beyond the point where the vocal folds have the largest glottal area (t = 1.33 ms). Even as the vocal folds are closing, the jet continues to accelerate and reaches its highest velocity around t = 1.51 ms or 5 = 1.71 ms. This is consistent with observations made by Triep et al [9], where it was noticed that the maximum volume flux occurs after the maximum opening of the glottis. This is not obvious from looking only at the XY view of the glottal jet. From the XY perspective the maximum velocities seem to appear around t = 1.33 ms, where the folds are fully open. However, with a whole field 3D reconstruction, additional information from the YZ plane indicates that high velocities continue through t = 1.71 ms where a large iso-contour of high velocity is visible.

4.6 Conclusions

A whole field, time-resolved 3D glottal jet was reconstructed using SAPIV on excised human vocal folds. Data was presented as time resolved for two cycles of oscillation consisting of 3 samples per cycle. A higher resolution data set, with 16 samples, was assembled to further investigate the glottal jet evolution. The unsteady 3D nature of the glottal jet that was observed is consistent with current voice research literature findings [9] [10].

The glottal jet velocity magnitudes mostly ranged between 10-20 m/s but reached a maximum of 30-40 m/s as the vocal folds reached their fully open position and began to close. The measured maximum velocity magnitude corresponds closely to the theoretical prediction of 37.6 m/s. Both the velocity magnitudes and this cyclic pattern of velocity increase are similar to results presented using other 3D flow reconstruction techniques [10].

The sequence in Figure 4-8 provided a valuable view of a full cycle of oscillation. The subglottal pressure, 2D motion, 3D motion, and 3D flow data matched well with expected vocal

fold behavior. The charted evolution of the glottal jet clearly showed a smooth evolution and disintegration of the flow field as cited in other recent findings [9].

The time resolved results do have some noisy areas in the velocity fields likely due to particle density in that section of the volume. However, some noisy is to be expected from vectors calculated at one moment in time. In many previous vocal fold 2D and 3D PIV studies hundreds of vectors are averaged to obtain a less noisy view of the glottal jet. In contrast, this study presented a snapshot of a unique unsteady glottal jet, rather than a general representation of 500 jets averaged into one. Both representations provide a valuable contribution to the further understanding of this complex phenomenon.

Despite having only one unique jet in each plot the SAPIV data was still effective in characterizing the evolution of the glottal jet. That speaks to the power of this fluid imaging technique and warrants its further use in investigating vocal fold fluid flow phenomena.

This study used a pair of healthy human vocal folds. Future studies will use SAPIV to characterize the 3D flow for various vocal fold conditions: vocal folds with abnormalities (polyps, scar tissue, etc.), asymmetric tensioning, and variable flow rates. These experiments can be repeated on many different types of vocal fold models in addition to the excised models. Utilizing SAPIV with vocal folds that vibrate at lower frequencies will increase the resolution of the flow data for each cycle and improve the time resolved results.

Time averaged results can also play an important role in understanding the 3D glottal jet. Taking phase locked SAPIV measurements to obtain a whole-field 3D glottal jet and averaging over hundreds of time steps would really clean up the data and allow for corroboration between this data and the 2D observations of the glottal jet from other authors.

Future studies that focus on coupling the motion of vocal folds with the glottal flow can help improve the understanding of these complicated interactions. Visualizing these phenomena and their interactions will help researchers and clinicians better understand the physics of speech production, with the ultimate aim of helping improve the clinical approach to treating patients with voice disorders.

4.7 Acknowledgements

We would like to thank NSF grant CMMI #1126862 and BYU for funding the equipment and development of the synthetic aperture technique. The NIH/NIDCD grant R01DC009616 for funding TTT, SLT, DJD and JRN and data relating to the vocal fold experiments and the University of Erlangen Graduate School in Advanced Optical Technologies (SAOT) for partial support of SLT. Furthermore, this work was work was supported by DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) grant no. FOR894/2 ”Stroemungsphysikalische Grundlagen der Menschlichen Stimmgebung”.

5 CONCLUSIONS AND FUTURE WORK

Voice quality is strongly linked to quality of life. In an effort to improve medical diagnosis and treatment of voice disorders, voice researchers have sought a complete understanding of the physics behind voice production. The work presented in this thesis expands two main areas of vocal fold FSI research, motion tracking and flow field visualization.

5.1 Three-Dimensional Tracking of Vocal Fold Superior Surface Motion (Chapter 3)

A new technique for three-dimensional vocal fold motion tracking was developed using a projected grid of dots imaged simultaneously by four cameras. The 2D pixel locations of the dots were extracted from each image and projected into 3D coordinates using epipolar geometry relationships for the camera array. A fully 3D reconstruction of the superior surface of the vocal folds was obtained from these 3D projections for a given moment in time. The 3D motion for several cycles of oscillation was obtained by compiling 70 consecutive time-steps.

5.1.1 Conclusions

A novel technique for tracking the motion of the vocal folds using multiple camera viewpoints and limited user interaction was developed. Four high-speed cameras (2000 fps) recorded an excised vocal fold model vibrating at 250 Hz. Based on the images from these four cameras a fully 3D reconstruction of the superior surface of the vocal folds was achieved. The

3D reconstruction of 70 consecutive time steps was assembled to characterize the motion of the vocal folds over eight cycles of oscillation. The 3D reconstruction accurately modeled the observed behavior of vocal fold vibration with a clearly visible mucosal wave. The average reprojection error for this technique was on par with other contemporary techniques (~20 μm).

5.1.2 Future Work

Future work for this project should focus in four areas. First, the computer vision algorithms need further development. Suggestions were made in Chapter 3 regarding improved thresholding algorithms that revisit missed points with a relaxed threshold after the first pass of dot location. This enhancement would increase the already 91% of points that are found by the computer automatically. Any other MATLAB fixes that limit user interaction will increase the viability of this technique.

Second, this technique is ideally suited to synchronization with fluid imaging measurements. The data presented herein were in fact also PIV data. This would allow motion and flow measurements to be taken at exactly the same moment in time all using the same experimental setup.

Third, additional experiments should be performed using high camera frame rates. The data for this technique does not have to be taken at the same time as PIV data. Valuable insight into the motion of the vocal folds could be gained with an experiment purely focused on utilizing this technique to understand the 3D motion of the vocal folds. The motion data could also be synchronized with subglottal pressure or acoustic data for further study of these important interactions in vocal fold vibration.

Finally, due to the non-invasive nature of this motion tracking technique, it is ideally suited to be miniaturized and used *in vivo*. Four unique viewpoints could be obtained using one

camera and a lenslet array with a laser dot grid applied to the vocal folds. This technique could provide physicians with a valuable tool to diagnose and treat those who suffer from voice disorders.

5.2 Whole-Field 3D Characterization of the Glottal Jet Using Synthetic Aperture Particle Image Velocimetry (Chapter 4)

Synthetic aperture particle image velocimetry was used with an excised human vocal fold model to study the airflow over the vocal folds during voice production. The 3D flow data was synced in time with motion and pressure data to obtain a greater understanding of vocal fold FSI.

5.2.1 Conclusions

A whole field, time resolved, three-dimensional reconstruction of the vocal fold fluid flow was obtained using synthetic aperture particle image velocimetry. Simultaneous 3D flow fields, subglottal pressure waves, and superior surface motion were presented for 2 consecutive cycles of oscillation. The reconstructed vocal fold fluid flow and motion measurements correlated exactly with expected behavior. A higher resolution view of one full cycle of oscillation consisting of 16 data points was compiled from 8 time resolved cycles via pressure data. The result was a full three-dimensional characterization of the evolution and disintegration of the glottal jet.

5.2.2 Future Work

Future studies will use SAPIV to characterize the 3D flow for varied vocal fold conditions: vocal folds with abnormalities (polyps, scar tissue, etc.), asymmetric vs. symmetric tensioning, and variable flow rates. These experiments can also be repeated on many different

types of vocal fold models. Working with excised vocal folds is a big challenge in a lot of ways. A more controlled experimental environment would ensure high quality SAPIV results. Utilizing SAPIV with vocal folds that vibrate at lower frequencies will increase the resolution of the flow data for each cycle and improve the time resolved results.

Time averaged results can also play an important role in understanding the 3D glottal jet. Taking phase locked SAPIV measurements to obtain a whole-field 3D glottal jet and averaging over hundreds of time steps would really clean up the data and allow for comparisons between other 2D observations of the glottal jet from other authors.

Another focus of research will be to completely synchronize flow, motion, pressure, and acoustic data. As mentioned above, the motion tracking technique present herein is ideal for simultaneous flow and motion measurements and would prove to be a useful tool when used in conjunction with SAPIV. Future studies that focus on coupling the motion of vocal folds, the acoustic signal, and the glottal flow can help improve the understanding of these complicated interactions. Visualizing these phenomena and their interactions will help researchers and clinicians better understand the physics of speech production, with the ultimate aim of helping improve the clinical approach to treating patients with voice disorders.

REFERENCES

[1] Triep, M., Brücker, C., and Schröder, W. High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds. Exp. Fluids. 39(2): 232–245 (2005).

[2] Drechsel, J.S., Thomson, S.L. Influence of supraglottal structures on the glottal jet exiting a two-layer synthetic, self-oscillating vocal fold model. J. Acoust. Soc. Am. 123(6): 4434-4445 (2007).

[3] Khosla, S., Muruguppan, S., Gutmark, E., Scherer, R. Vortical flow field during phonation in an excised canine larynx model. Ann. Otol. Rhinol. Laryngol. 116: 217–228 (2007).

[4] Krane, M., Barry, M., and Wei, T. Unsteady behavior of flow in a scaled-up vocal folds model. J. Acoust. Soc. Am. 122(6): 3659-3670 (2007).

[5] Pickup, B. A., & Thomson, S. L. Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models. J. Biomech. 42(14): 2219-2225 (2009).

[6] Thomson, S.L., Mongeau, L., Frankel, S.H. Aerodynamic transfer of energy to the vocal folds. J. Acoust. Soc. Am. 118(3): 1689-1700 (2005).

[7] Murray, P. R., Thomson, S. L. Synthetic, Multi-Layer, Self-Oscillating Vocal Fold Model Fabrication. J. Vis. Exp. 58: e3498 (2011).

[8] Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds. J. Acoust. Soc. Am. 121(2): 1102–1118 (2007).

[9] Triep, M., Brücker, C. Three-dimensional nature of the glottal jet. J. Acoust. Soc. Am. 127(3): 1537–1547 (2010).

[10] Krebs, F., Silva, F., Sciamarella, D., Artana, G. A three-dimensional study of the glottal jet. Exp. Fluids. (2011).

[11] Döllinger, M. The next step in voice assessment: High-speed digital endoscopy and objective evaluation, Current Bioinformatics, vol. 4(11): 101–111), (2009).

[12] Döllinger, M., Berry, D.A., Berke, G.S. Medial surface dynamics of an in vivo canine vocal fold during phonation. J. Acoust. Soc. Am. 117(5): (2005).

[13] Murray, P. Flow-induced responses of normal, bowed and augmented synthetic vocal fold models. Master's thesis, Brigham Young University (2011).

[14] Mantha, S., Mongeau, L., Siegmund, T. ASME conference proceedings. pp. 77–78 (2005).

[15] Luegmair, G., Kniesburges, S., Zimmermann, M., Sutor, A., Eysholdt, U., Döllinger. M. Optical reconstruction of high-speed surface dynamics in an uncontrollable environment. IEEE Transactions on Medical Imaging. 29(12): (2010).

[16] Riede, T., Lingle, S., Hunter, E.J., Titze, I.R. Cervids with different vocal behavior demonstrate different viscoelastic properties of their vocal folds, J. Morphol. 1(271): 1–11, (2010).

[17] George, N.A. de Mul, F.F.M., Qiu, Q., Rakhorst, G., Schutte. H.K. Depth-kymography: High-speed calibrated 3-D imaging of human vocal fold vibration dynamics, Phys. Med. Biol., vol. 53(10): 2667–2675, (2008).

[18] Svec, J. G., Schutte, H.K. Videokymography: High-speed line scanning of vocal fold vibration, J. Voice, 10(2): 201–205, (1996).

[19] Li, J., Chellappa, R. "Structure from planar motion," IEEE Trans. Image Process., 15(11): 3466–3477 (2006).

[20] Svoboda, T., Martinec, M., Pajdla, T. A convenient multi-camera self-calibration for virtual environments. PRESENCE: Teleoperators and Virtual Environments. 14 (4): 407-422 (2005).

[21] Gedeon, D., Rupitsch, S., Luegmair, G., Lerch, R., Döllinger, M. A Novel Setup for the Controlled Adduction of Vocal Folds", in: Proceedings of the 13-th Mechatronics Forum International Conference. Linz, Austria, Trauner, Linz (2012).

[22] Belden, J., Truscott, T. Axiak, M. Techet, A. H. "Three-dimensional synthetic aperture particle imaging velocimetry," Measurement Science and Technology 21:125403, (2010).

[23] Daily, J., Nielson, J., Truscott, T., Thomson, S. Three-dimensional whole-field measurement of a glottal jet using synthetic aperture particle image velocimetry. (in preparation).

[24] Wieneke, B. Volume self-calibration for 3D particle image velocimetry. Exp. Fluids. 45: 549–56 (2008).

[25] Szeliski, R. Computer Vision: Algorithms and Applications, Springer, Berlin, Germany, 1st edition, (2011).

[26] Sturm, P., Triggs. B. A factorization based algorithm for multi-image projective structure and motion. In European Conference on Computer Vision, volume 2: 709–720, (1996).

[27] Hartley, R., Zisserman, A. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK, (2000).

[28] Krausert, C.R., Olszewski, A.E., Taylor, L.N., McMurray, J.S., Dailey, S.H., Jiang, J.J. Mucosal Wave Measurement and Visualization Techniques. Journal of Voice. 25(4): 395-405 (2011).

[29] Becker S., Kniesburges S., Muller S., Delgado A., Link G., Kaltenbacher M., Dollinger M. Flow-structure-acoustic interaction in a human voice model. J Acoust Soc Am 125(3):1351–1361 (2009).

[30] Erath B.D., Plesniak M.W.,An investigation of asymmetric flow features in a scaled-up driven model of the human vocal folds. Exp Fluids 49(1):131–146 (2010).

[31] Sidlof P., Doare O., Cadot O., Chaigne A. Measurement of flow separation in a human vocal folds model. Exp Fluids 51:123–136 (2011).

APPENDIX A. MATLAB CODE FOR DOT TRACKING ALGORITHMS

This appendix contains the MATLAB code for dot tracking algorithms described in Chapter 3. The main program run by the user is main.m. This main program allows for user input and then calls the function fitgrid_v2.m, which takes the user input and interpolates a first guess grid. The next two functions, firstdotlocate.m and dotlocate_v2.m, are very similar, the former is used for only the first time step and the latter is used for all subsequent time steps. The data is reformatted using svobodaformat.m, to prepare the data for the 3D reprojection code. The MATLAB code for the Svoboda Multi-Camera Self-Calibration is available from Tomas Svoboda's website: http://cmp.felk.cvut.cz/~svoboda/SelfCal/. Finally, a short script is included, checkdist.m, which allows the user to determine the expected motion of a dot from one frame to the next. This information is valuable for determining the minimum size of the interrogation window.

A.1 main.m

```
% Created by Joey Nielson - 04/2012
%
% Dot tracking from multiple perspectives

numcams = 4;
begin_time = 197;
end_time = 208;
img = cell(1,4);
% Row and column locations for initial points
col = [1,2,3,4,5,6,7];
row = [2,5,12;2,5,12;2,5,12;10,11,12;3,5,11;3,5,11;3,5,11];
% X = zeros(4,21); %Used when initial points are selected
% Y = zeros(4,21);
```

```
for i=1:numcams
    img{i} = imread(['camera_' num2str(i) '/processed_images/C00'...
        num2str(i) 'H001S0001000' num2str(begin_time) '.tif']);
%     imshow(img{i}); %Used when initial points are selected
%     [X(i,:), Y(i,:)] = ginput(21);
end

%% Fit initial grid
[lgridx,lgridy] = fitgrid_v2(X,Y,numcams,col,row,img);

lgridx(:,1:9,4) = 0;
lgridy(:,1:9,4) = 0;

%% Find dot locations
[lgridx,lgridy,ngridx,ngridy] = firstdotlocate(X,Y,numcams,col,row,img);
points = svobodaformat(lgridx,lgridy,numcams,begin_time);

% Move to next images
for j = 1:(end_time-begin_time)
    for i=1:numcams
        img{i} = imread(['camera_' num2str(i) '/processed_images/C00'...
            num2str(i) 'H001S0001000' num2str(begin_time+j) '.tif']);
    end

    [lgridx,lgridy,ngridx,ngridy] = dotlocate_v2(lgridx,lgridy,X,Y,...
        numcams,col,row,img);
    points = svobodaformat(lgridx,lgridy,numcams,begin_time+j);
end
```

A.2 fitgrid_v2.m

```
function [lgridx,lgridy] = fitgrid_v2(X,Y,numcams,col,row,img)
% Created by Joey Nielson - 03/06/12
%
% This function fits a dot grid to the vocal folds for when all columns are
% used
% input variables: X, Y points selected from images
%                  numcams = number of cameras
%                  row, col - initial points picked
%                  img - struct of images
% output variables: lgridx, lgridy - grid points

lgridx = zeros(numcams,max(max(row)),max(max(col)));
lgridy = zeros(numcams,max(max(row)),max(max(col)));

for cam = 1:numcams
    count = 1;
    readnum = 1;
    % Interpolate points in 3 chosen columns

    for i=1:length(col) % Iterate for each column
        column = col(i);
        rows = row(i);
        first = 1;

        for j=1:2 % Iterate for each section of the line (2 default)
            pt1 = [Y(cam,readnum),X(cam,readnum)];
            if first == 1
```

```
                lgridy(cam,rows,column) = pt1(1);
                lgridx(cam,rows,column) = pt1(2);
                rows = rows + 1;
                first = 0;
            end
            pt2 = [Y(cam,readnum+1),X(cam,readnum+1)];

            drow = pt2(1)-pt1(1);
            dcol = pt2(2)-pt1(2);
            m = drow/dcol;
            dif = row(count,j+1)-row(count,j);

            pert_row = drow/dif;
            new_row = pt1(1);

            % Iterate for each interpolated point
            for k = 1:dif
                new_row = new_row+pert_row;
                new_col = (1/m)*(new_row-pt2(1))+pt2(2);
                lgridy(cam,rows,column) = new_row;
                lgridx(cam,rows,column) = new_col;
                rows = rows + 1;
            end
            readnum = readnum+1;
        end
        count = count+1;
        readnum = readnum+1;
    end
end

%Zero missing points
xzero = find(lgridx<0);
yzero = find(lgridy<0);
lgridx(xzero) = 0;
lgridx(yzero) = 0;
lgridy(xzero) = 0;
lgridy(yzero) = 0;

%% Plotting
%This will plot each image with the fitted grid of dots
% for j = 1:numcams
% imshow(img{j}); hold on
% for i = 1:7
%     plot(lgridx(j,:,i),lgridy(j,:,i),'.r');
%     plot(X(j,:),Y(j,:),'.g');
% end
% pause;
% end
```

A.3 firstdotlocate.m

```
function [lgridx,lgridy,ngridx,ngridy] = dotlocate_v2(X,Y,numcams,...
    col,row,img)
% Created by Joey Nielson - 04/04/12
%
% This function locates exact dot locations from an initial guess grid
% input variables: X,Y - selected points
%                  numcams = number of cameras
%                  row, col - initial points picked
%                  img - struct of images
% output variables: lgridx, lgridy - corrected laser grid points
```

```
%                   ngridx, ngridy - grid points to use as next guess

tym=0;
txm=0;
winrad = 6;
[lgridx,lgridy] = fitgrid_v2(X,Y,numcams,col,row);
lgridx(:,1:9,4) = 0;
lgridy(:,1:9,4) = 0;
bimg = cell(1,4);

tempgridx = lgridx;
tempgridy = lgridy;

% Iterate for each camera
for cam=1:numcams
    tempimg = conv2(im2double(img{cam}),1/16*[1,2,1;2,4,2;1,2,1],'same');
    T = 10/255;
    bimg = im2bw(img{cam},T);
    bimg = imerode(bimg,ones(3));
    bimg = imdilate(bimg,ones(4));
    if cam == 1
        lgridx(cam,:,5:6) = 0;
        lgridy(cam,:,5:6) = 0;
    end
    if cam == 4
            lgridx(cam,:,2:4) = 0;
            lgridy(cam,:,2:4) = 0;
    end
    for cols=1:max(col) % Iterate for each column
        for rows=1:max(max(row)) % Iterate for each row
            xcol = round(lgridx(cam,rows,cols)); % Get point
            yrow = round(lgridy(cam,rows,cols));

            if xcol ~=0 % Check if point is a zero point
                % Crop image to window around point
                bimg_crop = bimg((yrow-winrad):(yrow+winrad),...
                    (xcol-winrad):(xcol+winrad));
                timg = img{cam}((yrow-winrad):(yrow+winrad),...
                    (xcol-winrad):(xcol+winrad));
                timg = conv2(im2double(timg),1/16*[1,2,1;2,4,2;1,2,1],...
                    'same');

                % Determine area and centroid for each region
                clear areas
                areas = regionprops(bimg_crop, 'Area');
                clear cent
                cent = regionprops(bimg_crop,timg,'WeightedCentroid');

                if size(areas,1) ~= 0

                    if size(areas,1) == 1 % If there is only one region
                        cpt = cent(1).WeightedCentroid;
                        xm = cpt(1);
                        ym = cpt(2);
                    else
                        clear score
                        score = zeros(size(areas,1),4);

                        %Score areas
                        winning = 1;
                        losing = 1;
                        toparea = areas(1).Area;
                        botarea = areas(1).Area;
```

```
for t=1:size(areas,1)
    if areas(t).Area >= toparea
        winning = t;
    else if areas(t).Area <= botarea
            botarea = areas(t).Area;
            losing = t;
        end
    end
    xd = [winrad,winrad;cent(t).WeightedCentroid];
    dist(t) = pdist(xd,'euclidean');
end
score(:,1) =  (size(areas,1)+1)/2;
score(winning,1) = size(areas,1);
score(losing,1) = 1;

%Score proximity to initial guess
high = max(dist);
low = min(dist);
for t=1:size(areas,1)
    if dist(t) == high
        score(t,2) = 1;
    else if dist(t) == low
            score(t,2) = size(areas,1);
        else
            score(t,2) = (size(areas,1)+1)/2;
        end
    end
end

%Score slope from initial guess
for t=1:size(areas,1)
    cpt = cent(t).WeightedCentroid;
    if rows == 1
        m = (lgridy(cam,rows+1,cols)-...
            lgridy(cam,rows,cols))/...
            (lgridx(cam,rows+1,cols)-...
            lgridx(cam,rows,cols));
        m2 = (lgridy(cam,rows+1,cols)-...
            (xcol-winrad)+cpt(1)-1)/...
            (lgridx(cam,rows+1,cols)-...
            (yrow-winrad)+cpt(2)-1);
    else if rows == 12 ||...
                lgridy(cam,rows+1,cols) == 0
            m = (lgridy(cam,rows,cols)-...
                lgridy(cam,rows-1,cols))/...
                (lgridx(cam,rows,cols)-...
                lgridx(cam,rows-1,cols));
            m2 = ((xcol-winrad)+cpt(1)-1-...
                lgridy(cam,rows-1,cols))/...
                ((yrow-winrad)+cpt(2)-1-...
                lgridx(cam,rows-1,cols));
        else
            m = (lgridy(cam,rows+1,cols)-...
                lgridy(cam,rows,cols))/...
                (lgridx(cam,rows+1,cols)-...
                lgridx(cam,rows,cols));
            m2 = (lgridy(cam,rows+1,cols)-...
                (xcol-winrad)+cpt(1)-1)/...
                (lgridx(cam,rows+1,cols)-...
                (yrow-winrad)+cpt(2)-1);
        end
    end
    dif(t) = abs(m-m2);
end
```

```
[~,winning] = min(dif);
[~,losing] = max(dif);
score(:,3) = (size(areas,1)+1)/2;
score(winning,3) =  size(areas,1);
score(losing,3) = 1;
clear dif

%Score distance to next point of initial guess
for t=1:size(areas,1)
    cpt = cent(t).WeightedCentroid;
    if rows == 1
        p = [lgridy(cam,rows+1,cols),...
            lgridx(cam,rows+1,cols);...
            lgridy(cam,rows,cols),...
            lgridx(cam,rows,cols)];
        d = pdist(p,'euclidean');
        p2 = [lgridy(cam,rows+1,cols),...
            lgridx(cam,rows+1,cols);...
            (xcol-winrad)+cpt(1)-1,...
            (yrow-winrad)+cpt(2)-1];
        d2 = pdist(p2,'euclidean');
    else if rows == 12 ||...
                lgridy(cam,rows+1,cols) == 0
            p = [lgridy(cam,rows-1,cols),...
                lgridx(cam,rows-1,cols);...
                lgridy(cam,rows,cols),...
                lgridx(cam,rows,cols)];
            d = pdist(p,'euclidean');
            p2 = [lgridy(cam,rows-1,cols),...
                lgridx(cam,rows-1,cols);...
                (xcol-winrad)+cpt(1)-1,...
                (yrow-winrad)+cpt(2)-1];
            d2 = pdist(p2,'euclidean');
        else
            p = [lgridy(cam,rows+1,cols),...
                lgridx(cam,rows+1,cols);...
                lgridy(cam,rows,cols),...
                lgridx(cam,rows,cols)];
            d = pdist(p,'euclidean');
            p2 = [lgridy(cam,rows+1,cols),...
                lgridx(cam,rows+1,cols);...
                (xcol-winrad)+cpt(1)-1,...
                (yrow-winrad)+cpt(2)-1];
            d2 = pdist(p2,'euclidean');
        end
    end
    dif(t) = abs(d-d2);
end
[~,winning] = min(dif);
[~,losing] = max(dif);
score(:,4) = (size(areas,1)+1)/2;
score(winning,4) =  size(areas,1);
score(losing,4) = 1;

%Check if dot is already occupied
for t=1:size(areas,1)
    cpt = cent(t).WeightedCentroid;
    if rows ~=1
        chd = [(xcol-winrad)+cpt(1)-1,...
            (yrow-winrad)+cpt(2)-1;...
            lgridx(cam,rows-1,cols),...
            lgridy(cam,rows-1,cols)];
        chdist = pdist(chd,'euclidean');
        if chdist < 10
```

```
                                score(t,:) = 0;
                            end
                        end
                    end

                    %Final Score
                    winning = 1;
                    topscore = score(1,1)+3*score(1,2)+score(1,3)...
                        +score(1,4);
                    for t=1:size(areas,1)
                        pscore = score(t,1)+3*score(t,2)+score(t,3)...
                            +score(t,4);
                        if pscore >= topscore
                            winning = t;
                        end
                    end
                    cpt = cent(winning).WeightedCentroid;
                    xm = cpt(1);
                    ym = cpt(2);
                end
                % Determine new row and column
                nrow = (yrow-winrad)+ym-1;
                ncol = (xcol-winrad)+xm-1;

                % Set point to new row and column
                lgridx(cam,rows,cols) = ncol;
                lgridy(cam,rows,cols) = nrow;
            else
                lgridx(cam,rows,cols) = 0;
                lgridy(cam,rows,cols) = 0;
            end
            clear dist dif
        end
    end
end
end

%% New Grid
xpts = find(lgridx==0);
ypts = find(lgridy==0);

ngridx = lgridx;
ngridy = lgridy;
ngridx(xpts) = tempgridx(xpts);
ngridy(ypts) = tempgridy(ypts);

%% Plotting
% This will plot each image with the fitted grid of dots
winrad = 6;
figure;

for j = 1:numcams
    imshow(img{j}); hold on
    for i = 1:7
        plot(lgridx(j,:,i),lgridy(j,:,i),'.r');
    end

    reply = input('Missed points? [Y]/[N]: ', 's');
    if isempty(reply)
        reply = 'Y';
    end

    while reply == 'Y';
```

```
        mcol = input('Column number: ');
        mrow = input('Row number: ');
        [tx,ty] = ginput(1);

        img_crop = img{j}((ty-winrad):(ty+winrad),(tx-winrad):(tx+winrad));
        T = graythresh(img_crop);
        bimg_crop = im2bw(img_crop,T);
        timg = img{j}((ty-winrad):(ty+winrad),(tx-winrad):(tx+winrad));
        timg = conv2(im2double(timg),1/16*[1,2,1;2,4,2;1,2,1],'same');

        clear cent dist
        cent = regionprops(bimg_crop,timg,'WeightedCentroid');

        for t=1:size(cent,1)
            xd = [winrad,winrad;cent(t).WeightedCentroid];
            dist(t) = pdist(xd,'euclidean');
        end

        %Score proximity to initial guess
        low = min(dist);
        min_num = find(dist == low);

        mpt = cent(min_num).WeightedCentroid;
        xm = mpt(1);
        ym = mpt(2);

        nrow = (ty-winrad)+ym-1;
        ncol = (tx-winrad)+xm-1;

        % Set point to new row and column
        lgridx(j,mrow,mcol) = ncol;
        lgridy(j,mrow,mcol) = nrow;

        reply = input('Missed points? [Y]/[N]: ', 's');
        if isempty(reply)
            reply = 'Y';
        end
    end
    for i = 1:7
        plot(lgridx(j,:,i),lgridy(j,:,i),'.g');
    end
    pause;
end
```

A.4 dotlocate_v2.m

```
function [lgridx,lgridy,ngridx,ngridy] = dotlocate_v2(lgridx,lgridy,X,...
     Y,numcams,col,row,img)
% Created by Joey Nielson - 04/04/12
%
% This function locates exact dot locations from an initial guess grid
% input variables: lgridx, lgridy - initial guess locations
%                  X,Y - selected points
%                  numcams = number of cameras
%                  row, col - initial points picked
%                  img - struct of images
% output variables: lgridx, lgridy - corrected laser grid points
%                  ngridx, ngridy - grid points to use as next guess
```

```
tym=0;
txm=0;
winrad = 6;
lgridx(:,1:9,4) = 0;
lgridy(:,1:9,4) = 0;
bimg = cell(1,4);

tempgridx = lgridx;
tempgridy = lgridy;

% Iterate for each camera
for cam=1:numcams
    T = 10/255;
    bimg = im2bw(img{cam},T);
    bimg = imerode(bimg,ones(3));
    bimg = imdilate(bimg,ones(4));
%     figure(1); imshow(bimg); pause;
    if cam == 1
        lgridx(cam,:,5:6) = 0;
        lgridy(cam,:,5:6) = 0;
    end
    if cam == 4
            lgridx(cam,:,2:4) = 0;
            lgridy(cam,:,2:4) = 0;
    end
    for cols=1:max(col) % Iterate for each column
        for rows=1:max(max(row)) % Iterate for each row
            xcol = round(lgridx(cam,rows,cols)); % Get point
            yrow = round(lgridy(cam,rows,cols));

            if xcol ~=0 && ~isnan(xcol)% Check if point is a zero point
                % Crop image to window around point
                bimg_crop = bimg((yrow-winrad):(yrow+winrad),...
                    (xcol-winrad):(xcol+winrad));
                timg = img{cam}((yrow-winrad):(yrow+winrad),...
                    (xcol-winrad):(xcol+winrad));
                timg = conv2(im2double(timg),1/16*...
                    [1,2,1;2,4,2;1,2,1],'same');

                % Determine area and centroid for each region
                clear areas
                areas = regionprops(bimg_crop, 'Area');
                clear cent
                cent = regionprops(bimg_crop,timg,'WeightedCentroid');

                if size(areas,1) ~= 0

                    if size(areas,1) == 1 % If there is only one region
                        cpt = cent(1).WeightedCentroid;
                        xm = cpt(1);
                        ym = cpt(2);
                    else
                        clear score
                        score = zeros(size(areas,1),4);

                        %Score areas
                        winning = 1;
                        losing = 1;
                        toparea = areas(1).Area;
                        botarea = areas(1).Area;
                        for t=1:size(areas,1)
                            if areas(t).Area >= toparea
                                winning = t;
```

```
    else if areas(t).Area <= botarea
            botarea = areas(t).Area;
            losing = t;
        end
    end
    xd = [winrad,winrad;cent(t).WeightedCentroid];
    dist(t) = pdist(xd,'euclidean');
end
score(:,1) =  (size(areas,1)+1)/2;
score(winning,1) = size(areas,1);
score(losing,1) = 1;

%Score proximity to initial guess
high = max(dist);
low = min(dist);
for t=1:size(areas,1)
    if dist(t) == high
        score(t,2) = 1;
    else if dist(t) == low
            score(t,2) = size(areas,1);
        else
            score(t,2) = (size(areas,1)+1)/2;
        end
    end
end

%Score slope from initial guess
for t=1:size(areas,1)
    cpt = cent(t).WeightedCentroid;
    if rows == 1
        m = (lgridy(cam,rows+1,cols)-...
            lgridy(cam,rows,cols))/...
            (lgridx(cam,rows+1,cols)-...
            lgridx(cam,rows,cols));
        m2 = (lgridy(cam,rows+1,cols)-...
            (xcol-winrad)+cpt(1)-1)/...
            (lgridx(cam,rows+1,cols)-...
            (yrow-winrad)+cpt(2)-1);
    else if rows == 12 ||...
                lgridy(cam,rows+1,cols) == 0
            m = (lgridy(cam,rows,cols)-...
                lgridy(cam,rows-1,cols))/...
                (lgridx(cam,rows,cols)-...
                lgridx(cam,rows-1,cols));
            m2 = ((xcol-winrad)+cpt(1)-1-...
                lgridy(cam,rows-1,cols))/...
                ((yrow-winrad)+cpt(2)-1-...
                lgridx(cam,rows-1,cols));
        else
            m = (lgridy(cam,rows+1,cols)-...
                lgridy(cam,rows,cols))/...
                (lgridx(cam,rows+1,cols)-...
                lgridx(cam,rows,cols));
            m2 = (lgridy(cam,rows+1,cols)-...
                (xcol-winrad)+cpt(1)-1)/...
                (lgridx(cam,rows+1,cols)-...
                (yrow-winrad)+cpt(2)-1);
        end
    end
    dif(t) = abs(m-m2);
end
[~,winning] = min(dif);
[~,losing] = max(dif);
score(:,3) = (size(areas,1)+1)/2;
```

```
score(winning,3) =  size(areas,1);
score(losing,3) = 1;
clear dif

%Score distance to next point of initial guess
for t=1:size(areas,1)
    cpt = cent(t).WeightedCentroid;
    if rows == 1
        p = [lgridy(cam,rows+1,cols),...
            lgridx(cam,rows+1,cols);...
            lgridy(cam,rows,cols),...
            lgridx(cam,rows,cols)];
        d = pdist(p,'euclidean');
        p2 = [lgridy(cam,rows+1,cols),...
            lgridx(cam,rows+1,cols);...
            (xcol-winrad)+cpt(1)-1,...
            (yrow-winrad)+cpt(2)-1];
        d2 = pdist(p2,'euclidean');
    else if rows == 12 || ...
                lgridy(cam,rows+1,cols) == 0
            p = [lgridy(cam,rows-1,cols),...
                lgridx(cam,rows-1,cols);...
                lgridy(cam,rows,cols),...
                lgridx(cam,rows,cols)];
            d = pdist(p,'euclidean');
            p2 = [lgridy(cam,rows-1,cols),...
                lgridx(cam,rows-1,cols);...
                (xcol-winrad)+cpt(1)-1,...
                (yrow-winrad)+cpt(2)-1];
            d2 = pdist(p2,'euclidean');
        else
            p = [lgridy(cam,rows+1,cols),...
                lgridx(cam,rows+1,cols);...
                lgridy(cam,rows,cols),...
                lgridx(cam,rows,cols)];
            d = pdist(p,'euclidean');
            p2 = [lgridy(cam,rows+1,cols),...
                lgridx(cam,rows+1,cols);...
                (xcol-winrad)+cpt(1)-1,...
                (yrow-winrad)+cpt(2)-1];
            d2 = pdist(p2,'euclidean');
        end
    end
    dif(t) = abs(d-d2);
end
[~,winning] = min(dif);
[~,losing] = max(dif);
score(:,4) = (size(areas,1)+1)/2;
score(winning,4) =  size(areas,1);
score(losing,4) = 1;

%Check if dot is already occupied
for t=1:size(areas,1)
    cpt = cent(t).WeightedCentroid;
    if rows ~=1
        chd = [(xcol-winrad)+cpt(1)-1,...
            (yrow-winrad)+cpt(2)-1;...
            lgridx(cam,rows-1,cols),...
            lgridy(cam,rows-1,cols)];
        chdist = pdist(chd,'euclidean');
        if chdist < 10
            score(t,:) = 0;
        end
    end
```

```
                        end

                        %Final Score
                        winning = 1;
                        topscore = score(1,1)+3*score(1,2)+score(1,3)+...
                            score(1,4);
                        for t=1:size(areas,1)
                            pscore = score(t,1)+3*score(t,2)+score(t,3)+...
                                score(t,4);
                            if pscore >= topscore
                                winning = t;
                            end
                        end
                        cpt = cent(winning).WeightedCentroid;
                        xm = cpt(1);
                        ym = cpt(2);
                    end
                    % Determine new row and column
                    nrow = (yrow-winrad)+ym-1;
                    ncol = (xcol-winrad)+xm-1;

                    % Set point to new row and column
                    lgridx(cam,rows,cols) = ncol;
                    lgridy(cam,rows,cols) = nrow;
                else
                    lgridx(cam,rows,cols) = 0;
                    lgridy(cam,rows,cols) = 0;
                end
                clear dist dif
            end
        end
    end
end

%% New Grid
xpts = find(lgridx==0);
ypts = find(lgridy==0);

ngridx = lgridx;
ngridy = lgridy;
ngridx(xpts) = tempgridx(xpts);
ngridy(ypts) = tempgridy(ypts);

%% Plotting
% This will plot each image with the fitted grid of dots
winrad = 6;
figure;
for j = 1:numcams
    imshow(img{j}); hold on
    for i = 1:7
        plot(lgridx(j,:,i),lgridy(j,:,i),'.r');
    end

    reply = input('Missed points? [Y]/[N]: ', 's');
    if isempty(reply)
        reply = 'Y';
    end

    while reply == 'Y';
        mcol = input('Column number: ');
        mrow = input('Row number: ');
        [tx,ty] = ginput(1);
```

```
        img_crop = img{j}((ty-winrad):(ty+winrad),(tx-winrad):(tx+winrad));
        T = graythresh(img_crop);
        bimg_crop = im2bw(img_crop,T);
        timg = img{j}((ty-winrad):(ty+winrad),(tx-winrad):(tx+winrad));
        timg = conv2(im2double(timg),1/16*[1,2,1;2,4,2;1,2,1],'same');

        clear cent dist
        cent = regionprops(bimg_crop,timg,'WeightedCentroid');

        for t=1:size(cent,1)
            xd = [winrad,winrad;cent(t).WeightedCentroid];
            dist(t) = pdist(xd,'euclidean');
        end

        %Score proximity to initial guess
        low = min(dist);
        min_num = find(dist == low);
        mpt = cent(min_num).WeightedCentroid;
        xm = mpt(1);
        ym = mpt(2);
        nrow = (ty-winrad)+ym-1;
        ncol = (tx-winrad)+xm-1;

        % Set point to new row and column
        lgridx(j,mrow,mcol) = ncol;
        lgridy(j,mrow,mcol) = nrow;

        reply = input('Missed points? [Y]/[N]: ', 's');
        if isempty(reply)
            reply = 'Y';
        end
    end
    for i = 1:7
        plot(lgridx(j,:,i),lgridy(j,:,i),'.g');
    end
    pause;
end
```

A.5 svobodaformat.m

```
function [points] = svobodaformat(lgridx,lgridy,numcams,time)
% Created by Joey Nielson - 04/04/12
%
% Reformats the 2D dot locations for each camera into a format that can be
% used by the Svoboda 3D projection code.
% input variables: lgridx, lgridy = laser grid points
%                  numcams = number of cameras
%                  time = time step number
% output variables: points = a (3*#cam)x(#points) matrix of all point pixel
%                   locations

points = zeros(numcams,size(lgridx,2)*size(lgridx,3));

for i =1:numcams
    count = 1;
    for k = 1:size(lgridx,3)
        for j = 1:size(lgridx,2)
            points(i*3-2,count) = lgridx(i,j,k);
            points(i*3-1,count) = lgridy(i,j,k);
            if lgridy(i,j,k) == 0;
```

```
                points(i*3,count) = NaN;
            else
                points(i*3,count) = 1;
            end
            count = count +1;
        end
    end
end

nn = find(points==0);
points(nn) = NaN;
save  (['./results2/time' num2str(time) '/points2.dat'], 'points',...
    '-ASCII');
IdMat = ones(size(points,1)/3,size(points,2));
save  (['./results2/time' num2str(time) '/IdMat2.dat'], 'IdMat','-ASCII');

cam = 1;
for i=1:3:size(points,1)
    for j = 1:size(points,2)
        if isnan(points(i,j)) == 1
            IdMat(cam,j) = 0;
        end
    end
    cam = cam +1;
end

calib_points = load(['./calib_plane/points.dat']);
npoints = calib_points(1:12,1:81);

for i=1:3:size(points,1)
    points(i:i+1,:) = points(i:i+1,:) + 192;
end

points = [npoints points];
IdMat = [ones(numcams,81) IdMat];
Res = repmat([1024,1024],numcams,1);
mkdir(['./results2/time' num2str(time)]);
save  (['./results2/time' num2str(time) '/points.dat'], 'points','-ASCII');
save  (['./results2/time' num2str(time) '/IdMat.dat'], 'IdMat','-ASCII');
save  (['./results2/time' num2str(time) '/Res.dat'], 'Res','-ASCII');
save  (['./results2/time' num2str(time) '/Res2.dat'], 'Res','-ASCII');
```

A.6 checkdist.m

```
% Created by Joey Nielson - 04/2012
%
% This script allows the user to select three points from consective images
% and measure how much a given point moves from time to time.

x2 = zeros(1,3);
y2 = zeros(1,3);
for time=197:199
    img{time} = imread(['camera_1/processed_images/C001H001S0001000'
num2str(time) '.tif']);
    imshow(img{time});
    [x2(1,time-197+1), y2(1,time-197+1)] = ginput(1);
end

p = [x2(1,3),y2(1,3);x2(1,2),y2(1,2)];
d = pdist(p,'euclidean');
```