Semi-automatic Tracking of Hyolaryngeal Coordinates in Videofluoroscopic Swallow Studies

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Rajkiran Natarajan

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

Many research questions in dysphagia research require frame-by-frame annotation of anatomical landmarks visible in videofluorographs as part of the research workflow, which can be a tedious and error prone process. Such annotation is done manually using image analysis tools, is error prone, and characterized by poor rater reliability. In this thesis, a computer-assisted workflow that uses a point tracking technique based on the Kanade-Lucas-Tomasi tracker to semi-automate the annotation process, is developed and evaluated. Techniques to semi-automate the annotation process have been explored but none have had their research value demonstrated. To demonstrate the research value of a workflow based on point tracking in enhancing the annotation process, the developed workflow was used to perform an enhanced version of the recently published Coordinate Mapping swallowing study annotation technique to determine several swallowing parameters. Evaluation was done on eight swallow studies obtained from a variety of clinical sources and showed that the workflow produced annotation results with clinically insignificant spatial errors. The workflow has the potential to significantly enhance research processes that require frame-by-frame annotation of anatomical landmarks in videofluorographs as part of their data preparation steps, by reducing the total time required to annotate clinical cases.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Ian K. Stavness, without whose guidance and expertise, this thesis would not have been possible. His unconditional assistance throughout my enrollment in the graduate program ensured that the student experience leading up to this final triumph was marked evenly with highlights of success. His expertise in the mechanics of swallowing was absolutely necessary. And I cannot possibly overstate his enthusiasm for his research specialty, which was instilled in me and propelled our many collaborative works. To summarize his tireless research spirit succinctly, I would like to recall a line from Dante Alighieri's Divine Comedy: *con l'ali snelle ... del gran disio*; borne on the swiftly beating wings of great desire.

Special thanks go to our clinical collaborator Dr. William Pearson Jr. at the Department of Anatomy and Cell Biology at Georgia Regents University, who maintained the inter-disciplinary cohesion of our research. It is a great stroke of fortune to have had one of the leading researchers of oropharyngeal dysphagia oversee my research, and I am contented to see his esteemed research benefit from this thesis.

Thanks must be given to those who went out of their way for the provision of high quality videofluorographs which were necessary for the evaluation of this thesis' subject: Dr. Nancy Solomon at the Walter Reed Military Medical Center, Dr. Angela Dietsch at the University of Nebraska at Lincoln, Dr. Bonnie Martin-Harris at the Medical University of South Carolina, and again Dr. William Pearson Jr. Research was generously funded by NSERC, CIHR, and the U.S. Army Medical Research and Development Program (W81XWH-12-2-0021, PI: Solomon).

Finally, I cannot forget to thank the thesis committee for volunteering to review my thesis and the technical staff of the various computational facilities that I relied on at the University of Saskatchewan.

Preface

This dissertation is an original intellectual product of the author. The work presented hereafter is the result of an inter-disciplinary research effort between the Department of Computer Science at the University of Saskatchewan and Dr. William Pearson Jr. from the Department of Anatomy and Cell Biology at Georgia Regents University, aimed at improving computer-aided annotation of videofluorographs from videofluoroscopic swallow studies.

Ethics approval for all experimental data and methods was obtained from the University of Saskatchewan's Biomedical Research Ethics Board (ethics approval number 357205-14).

I would like to gratefully acknowledge the videofluoroscopic data provided by my collaborators: Dr. William Pearson Jr. (Georgia Regents University), Dr. Bonnie Martin-Harris (Medical University of South Carolina), and Dr. Nancy Solomon (Walter Reed Military Medical Center; US Army Medical Research and Development Program, W81XWH-12-2-0021, PI: Solomon).

The anatomy illustrations used in section 2 are adapted from various medical publications to whom credit is given in the captions of those figures.

Portions of research reported in this thesis have been published as *Natarajan RN, Stavness I, Pearson Jr. W. Semi-automatic tracking of hyolaryngeal coordinates in videofluoroscopic studies. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization (2015): 1-11* I was the lead investigator and was entirely responsible for concept formulation, data collection, and manuscript preparation, and partly responsible for data analysis. Dr. I. K. Stavness and Dr. William Pearson Jr. were co-investigators who contributed the remainder of the data analysis effort required to complete the manuscript. They were also the supervisors of the project and were involved from conception until manuscript completion.

Contents

List of Tables

List of Figures

List of Abbreviations

VF	VideofluoroscopeVideofluorograph
VFSS	Videofluoroscopic Swallow Study
RMSE	Root-mean-square-error

Chapter 1

Introduction

The gold standard for evaluating swallowing health in clinical and research settings is the Videofluoroscopic Swallow Study (VFSS). The VFSS is an X-ray video of the upper airway of the patient taken while he/she swallows. In clinical settings, diagnoses made from VFSS are generally based on perceptual inferences about swallowing health from what is apparent in the image details of the video. Diagnoses are usually concerned with determining the severity of a patient's dysphagia, a clinical term used to describe the condition of having impaired swallowing function. In a typical diagnosis procedure, a patient will be made to swallow food while situated in front of a videofluoroscope (VF) instrument, the VF video is prepared, and perceptual judgements about swallowing function are then made from what is apparent in the recorded imagery. For example, if food is seen to enter the airway after swallows, then an observation of aspiration may be made and the patient diagnosed with dysphagia. Figure 1.1 shows a sequence of 4 frames taken from a VFSS in temporally increasing order, showing the movements of various anatomical structures during the course of a swallow.

The problem with such subjective dysphagia diagnoses is that they are characterized by low inter- and intra-rater reliability, making it unlikely that a particular clinician will apply the same standards of diagnosis across different clinical cases. In research settings, evaluations are usually done quantitatively to facilitate the scientific process. However, answering some research questions often requires analysis of significant numbers of VFSSs. The most tedious kinds of quantitative analyses involve determining the locations of anatomical landmarks in one or more frames of a VFSS and using them to determine swallowing parameters that are of interest; hereafter, the task of recording/determining coordinates is referred to simply as annotation.

Thus far, quantitative VFSS annotation and analysis has mostly been done manually with the aid of image

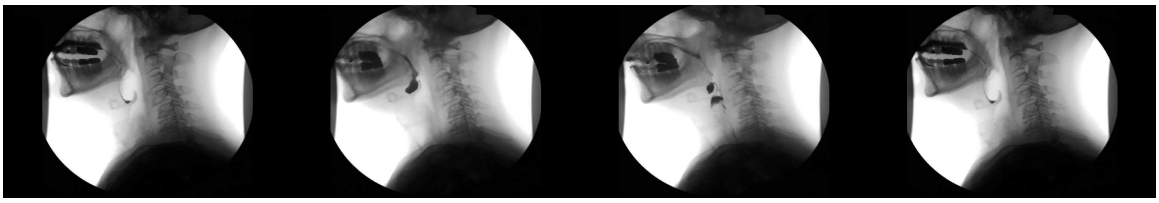

Figure 1.1: A montage of frames taken from a VFSS shown in temporally increasing order. Pixel intensities indicate radio-density of objects within the instrument's detector's field-of-view, hence, skeletal features with high radio-density appear darker than less dense features like soft-tissue and air.

markup and annotation tools like the National Institute of Health's ImageJ® tool. Manual markup with such tools often involves using point and click GUI controls to record the locations of anatomical landmarks. This is feasible for studies where the number of VFSSs needing to be analyzed is small. However, for studies requiring analysis of large numbers of VFSSs, this can be impractical. Further, some of these techniques are characterized by sub-optimal rater reliability. Thus, research that requires analysis of large numbers of VFSSs is expensive in terms of research time and is characterized by problems associated with rater reliability.

In response to these problems, the research community has explored computerized techniques to partly automate the analysis process. These techniques often automate some part of the annotation process, automatically determine a few swallowing parameters, and focus on requiring minimal user input. However, many of these techniques are designed to automate the annotation process and allow determination of one or a few swallowing parameters and are too narrowly focused on enhancing a particular research process for a research question. It is generally desirable to be able to annotate arbitrary landmarks in VFSSs as this would enhance research processes that can determine several swallowing parameters at once. Such a technique has not yet been developed and evaluated.

1.1 Problem

The problem addressed in this thesis is that currently no technique exists that has been evaluated for its research value in enhancing research processes that require frame-by-frame landmark annotation of VFSSs. Such a technique must overcome these problems associated with manual annotation:

1. Tediousness: Manual annotation is time consuming. A typical VFSS capturing a single swallow can have between 30-50 frames. The number of frames and the number of landmarks needing to be annotated depend on the research question being asked but typically vary between 3 landmarks in 2 frames for determining one swallowing parameter, to up to 10 landmarks in all of a VFSS's video frames for performing multivariate analysis and morphometry. Annotating a single VFSS can take an experienced clinician between 20 to 40 minutes.

2. Accuracy: One of the factors affecting accuracy of manual annotation is the experience of the analyst. Even the most experienced analyst is fallible and can introduce significant analysis errors. Research quality depends on accurate experimental work.

3. Rater reliability: Rater reliability is a problem when many different clinical cases need to be processed. There are no published studies on rater reliability of research procedures that require frame-by-frame annotation of landmarks, however, it is believed to be marginal. Inter-rater reliability determines consistency between results produced by different analysts for analysis of the same clinical specimen. Lack of rater reliability affects replicability of clinical research and hence research quality. Computer techniques that are deterministic can generally produce accurate and consistent analysis, reducing sources

of accuracy and reliability errors. Intra-rater reliability determines how consistently a rater applies evaluation standards across different clinical specimens. Use of deterministic computer techniques will likely help improve rater reliability of manual analysis. This thesis approaches the problem of intra-rater reliability.

1.2 Motivation

Research based on quantitative analysis of swallowing with VFSSs is generally laborious. A growing number of research questions concerned with the oropharyngeal phase of swallowing are being asked and answered by multivariate analysis of hyolaryngeal kinematics across large numbers of patient samples. Pisegna et al. (2015) used Coordinate Mapping (described in section 4.1) to determine swallowing mechanism differences between patients experiencing post-stroke aspiration and normal subjects. Pearson et al. (2015) also used Coordinate Mapping to determine mechanical differences associated with a recently developed quantitative scoring scheme for dysphagia severity assessment. Dietsch et al. (2015) used Coordinate Mapping to explore mechanical differences in swallowing between control groups and patients with post-trauma aspiration. The number of clinical cases that need to be analyzed will vary based on the statistical-significance requirements of an individual study. Upwards of 10 VFSSs may need to be analyzed to constitute a significant study sample. Some emerging clinical protocols like the MBSImP (Martin-Harris et al., 2008) require analysis of 12 VFSSs per patient. Some large-scale studies like Dietsch et al. (2015) for example relied on manual analysis of 219 VFSSs.

The development and evaluation of a computer technique that enhances research processes by reducing annotation time and producing results that are accurate would enhance research quality and be a significant contribution to the dysphagia research community.

1.3 Solution

To address the problems associated with manual annotation, several techniques have been developed by the research community. Recent techniques use well-established computer-vision algorithms to semi-automate manually intensive parts of analysis. As mentioned before, these techniques focus on enhancing the process of determining one or a few swallowing parameters and very few of them have general utility in helping answer a broad range of research questions that require frame-by-frame analysis of anatomical landmark coordinates - a key step in research questions that ask a wide variety of questions related to hyolaryngeal kinematics.

One solution that has been modestly explored is the Kanade-Lucas-Tomasi feature tracker to track the positions of a landmark from frame to frame using Harris corners as initialization points in a starting frame. This technique works well for tracking features in short video sequences, which most VFSSs are. A workflow that allows for the semi-automatic annotation of several landmarks in a sitting with the help of a GUI is one way to implement such a solution. Importance should be placed on minimal introduction of user input.

In this thesis, a computer-assisted workflow that semi-automates the process of landmark annotation was developed with the goal of enhancing the process of VFSS landmark annotation. The workflow was evaluated by employing it to perform annotation based on a recently published clinical technique, Coordinate Mapping, that allows for several swallowing measurements (also referred to in the context of swallowing research as kinematic variables) to be made efficiently. The effectiveness and research value of the workflow was determined by studying the accuracy of annotations made using the tool.

1.4 Evaluation

The developed workflow was evaluated using 8 VFSSs obtained from a variety of clinical sources to determine several swallowing kinematics using the Coordinate Mapping technique. As per Coordinate Mapping, evaluation required annotation of 6 anatomical landmarks in every frame of each VFSS. Workflow and ground truth results were compared to determine the following characteristics of the workflow:

1. Anatomical landmark annotation accuracy: How close on average, geometrically, are workflow determined coordinates and ground truth?

2. Kinematic variables determination accuracy: Do determined anatomical landmark coordinates determine kinematic variables accurately? Are the kinematics correlated with the ground truth?

3. Sensitivity analysis: How different are annotation results when small deviations are introduced in how the workflow is used between repeated trials on the same VFSS?

1.5 Contributions

The main contributions of this thesis are:

1. Design and implementation of a computer-assisted workflow for the semi-automatic annotation of anatomical landmarks in VFSS imagery.

2. Evaluation of the accuracy of the workflow with a recently published protocol -Coordinate Mapping- for determining swallowing parameters from annotated VFSS imagery.

3. Evaluation of the intra-rater sensitivity of the workflow to determine the similarity of annotation results produced between repeated uses of the workflow on the same VFSS by the same user.

4. Deployment of the workflow in a research setting in which it was used to annotate VFSSs for a multivariate analysis study of swallowing improvements after swallowing-related exercise.

1.6 Overview of thesis

The dissertation elaborates the research effort described in this chapter. Chapter 2 provides clinical background on swallowing and swallowing disorders at a level sufficient to understand the clinical aspects of this thesis. Chapter 3 surveys previous work related to computer-assisted diagnosis and/or annotation using VFSSs. Chapter 4 describes the design and implementation of the tool used to implement the proposed workflow. The Coordinate Mapping technique which was used as a case study for demonstrating the research potential of the workflow is also introduced. Chapter 5 describes the methodology adopted to evaluate the workflow's accuracy as well as the results of the evaluation. The results and their implications are also discussed. Chapter 7 summarizes the work presented and discusses possibilities for extending the findings of this thesis.

Chapter 2

Background

Because this thesis is concerned with the use of computer-aided-diagnosis tools to enhance dysphagia diagnosis, some background on the anatomy of the human oropharyngeal region, swallowing function, and treatment of swallowing disorders is necessary. A primer on the medical details relevant to this thesis work is provided to help non-clinical readers understand the clinical aspects of this thesis.

2.1 Basic oropharyngeal anatomy

Many skeletal, soft-tissue structures, and nerves are involved in swallowing. The scope of swallowing anatomy in this thesis is limited to a small subset of those anatomical elements, excluding nerves. These structures are located in three regions of the pharynx: the nasopharynx, oropharynx, and the laryngopharynx. Figure 2.1 shows lateral and posterior views of the upper airway in which these regions are delineated. This swallowing-related anatomical elements that this thesis is concerned with are located mainly in the oropharynx and laryngopharynx.

A detailed explanation of how these are involved in swallowing is given in the subsequent section. The main anatomical elements whose structures and functions the non-clinical reader is likely to be unfamiliar with and will be referenced later in the thesis are:

1. Bolus: The bolus refers to mechanically digested food that is intended to be swallowed. Although not an anatomical structure, it is a prominent artifact in VFSSs and is a common control in dysphagia research.

2. Epiglottis: The epiglottis is a leaf-shaped flap made of elastic cartilaginous tissue and is attached to the entrance of the larynx. When not involved in swallowing, it is pointed at an oblique upwards angle, pointing towards the spine. During swallowing, the epiglottis is forced to fold over the vocal chords to assist in the prevention of bolus particles entering the airway.

3. Pharynx: The pharynx constitutes the part of the throat that is immediately behind the nasal cavity and above the esophagus and larynx. It is made up of several muscle bundles.

4. Larynx: The larynx is a hollow tube-shaped organ made of cartilaginous tissue that is composed of several anatomical structures. Its primary roles are to provide a conduit for respiration and speech.

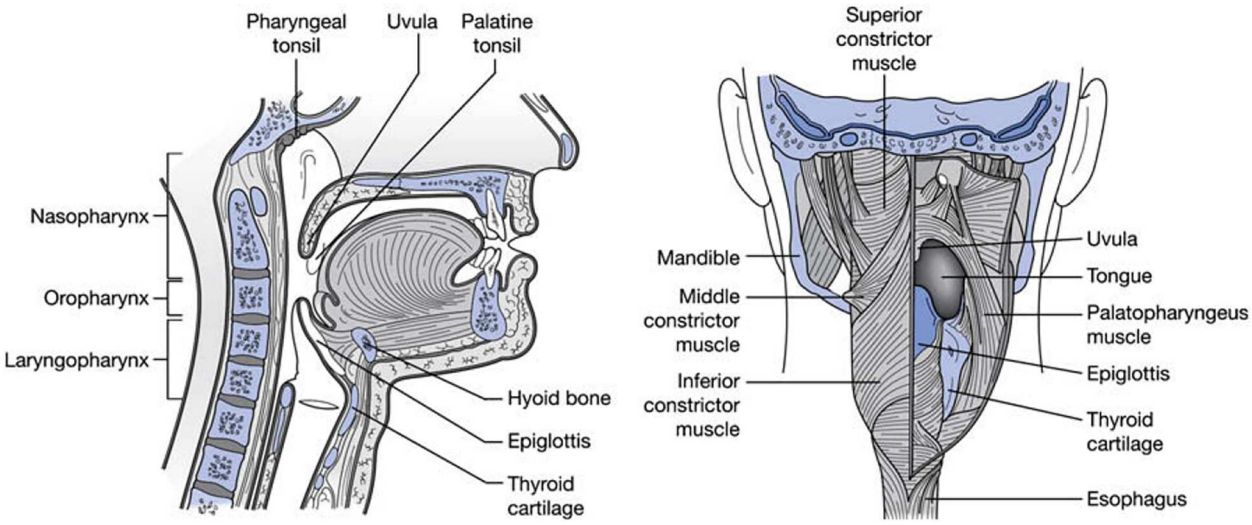

Figure 2.1: Anatomy of the upper airway in the lateral (A) and posterior (B) views (Matsuo and Palmer, 2008)

Figure 2.3 shows a drawing of the larynx.

5. Hyoid bone: The hyoid bone is a horseshoe shaped bone that at rest, lies just behind and under the chin at the point of the midline of the neck. During swallowing, it moves superiorly (upwards) and anteriorly (forwards). This correctness of this trajectory is sometimes used as an indicator of swallowing problems. It is a floating bone and in VFSSs appears as a radio-opaque "L"-shaped object just under the mandible. Figure 2.2 shows an anatomical drawing of the hyoid bone

6. Cricoid cartilage: The cricoid cartilage is a ring-like cartilaginous structure that surrounds the trachea and is located near the middle and center of the neck.

7. Hyolaryngeal complex: The hyoid bone, the larynx, the thyroid cartilage, and the cricoid cartilage are inter-connected by various tendons and muscles. During swallowing, this complex elevates superiorly and anteriorly to assist in closing the airway and opening the upper-esophageal sphincter to allow food to enter the digestive system by way of the esophagus.

8. Muscles attached to the hyolaryngeal complex: Several muscles connect to various elements of the hyolaryngeal complex and also between them. A subset of these muscles are responsible for the individual movements of the hyolaryngeal structures. Another subset of these muscles is responsible for the superior and anterior movement of the hyolaryngeal complex during swallowing. This subset of muscles is of concern to this thesis only insofar as the movement of the entire hyolaryngeal complex is concerned.

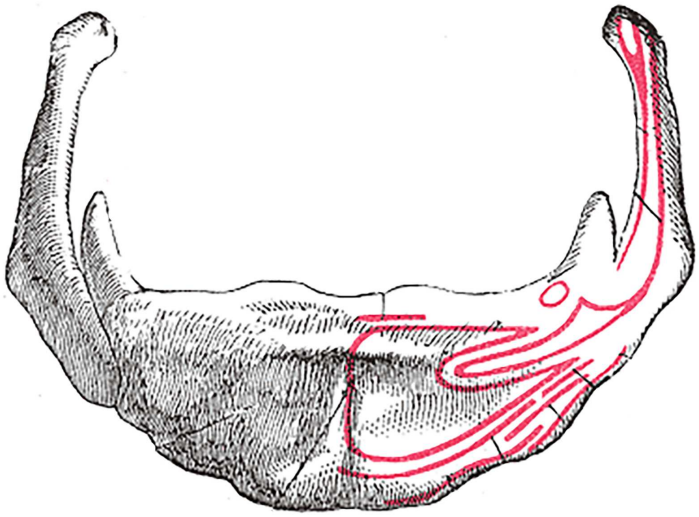

Figure 2.2: An anatomical drawing of the hyoid bone. Red contours on one side show attachment points for various muscles (Gray, 1918)

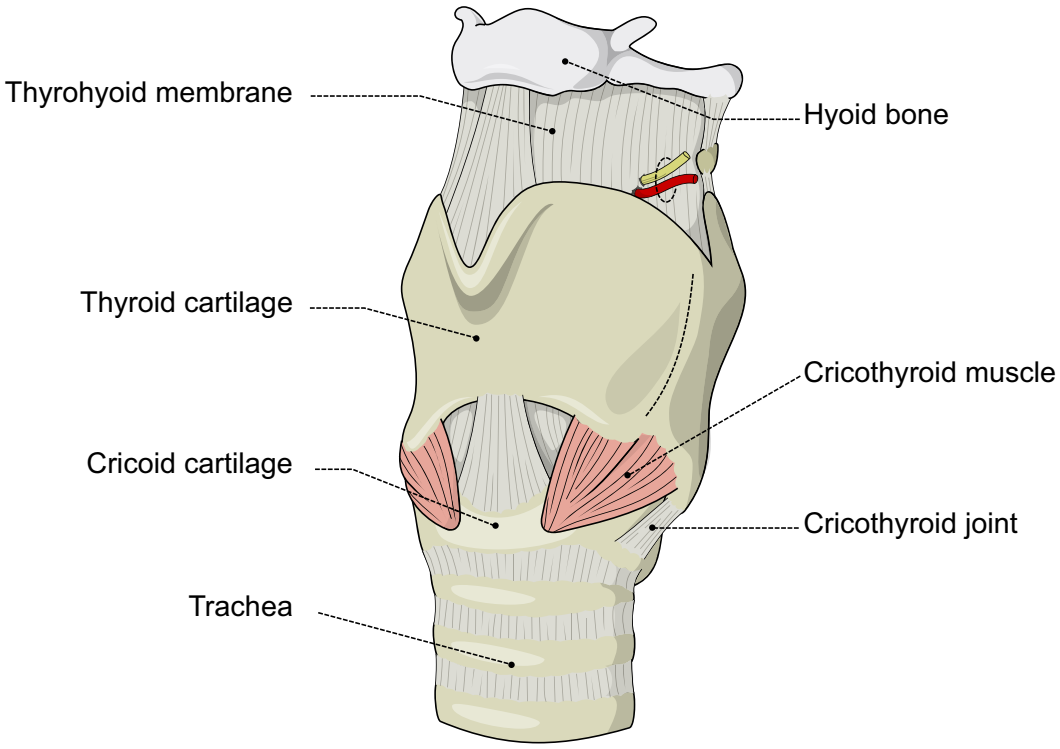

Figure 2.3: An anatomical drawing of the larynx (Remesz, 2008)

2.2 Swallowing

Swallowing is a complex physiological process that involves the coordination of many muscles to control the movement of several oropharyngeal structures with the goal of converting a respiratory conduit into a digestive one. Synchronized movements are required to prevent aspiration (entrance of food or liquid boluses into the airway). Paradigmatic models exist to explain swallowing of food and liquid boluses; except for differences arising from bolus consistency, the models are similar. To facilitate introduction of swallowing function at a level sufficient for this thesis, a simplified summary of the Four Stage Model for drinking and swallowing liquid is presented, based on the detailed analysis of Matsuo and Palmer (2008). The four stages, following intake of the liquid bolus into the oral cavity, are:

1. Oral preparatory stage: in which the bolus is held firmly in place in the anterior part of the floor of the mouth by the tongue, against the area of the hard palate posterior to the upper dental arch. At this stage, the oral cavity is sealed by the tongue and soft palate to prevent the bolus from entering the oropharynx.

2. Oral propulsive stage: in which the tongue tip rises against the hard palate, pressing against it progressively from the front of the hard palate to the back, causing the bolus to be propelled into the back of the oral cavity. The oral preparatory and oral propulsive stages can last up to a few seconds.

3. Pharyngeal stage: in which the bolus enters the pharynx. At this stage, the soft palate elevates to press against the nasopharynx, to prevent bolus regurgitation into the nasal cavity. The pharyngeal constrictor muscles contract from top to bottom, squeezing the bolus downward into the pharynx. Noteworthy in this stage is that the epiglottis tilts backwards to seal the laryngeal vestibule, closing off the airway. Also in this stage, the hyoid bone and larynx (and other structures associated with the hyolaryngeal complex) are pulled upwards and then forwards, which among other things, causes a relaxed upper esophageal sphincter to open, allowing the bolus to enter.

4. Esophageal stage: in which the bolus has entered the esophagus and into the digestive system.

2.3 Dysphagia

Dysphagia is a term used to indicate a state of diminished swallowing function. The causes of dysphagia are numerous and include, but are not limited to, damage to upper airway anatomy caused by trauma, side-effects of oropharyngeal cancer treatment, systemic neurological diseases like Parkinson's disease or stroke, and aging (Roden and Altman, 2013). For example, treatment of oropharyngeal cancers with intensity modulated radiation therapy can cause weakening of muscles involved in elevation of the hyolaryngeal complex. Stroke can lead to difficulty initiating swallowing. Absorption of explosion induced shock waves by the throat area can damage structures involved in swallowing (Cherney et al., 2010).

The sequelae of untreated dysphagia include, but are not limited to, increased deterioration of swallowing ability, aspiration, and apnea. It has been reported that dysphagia affects between 300,000 - 600, 000 patients yearly (Sura et al., 2012). Dysphagia incidences are more commonly reported by studies investigating dysphagia as a sequel to various injuries and risk-factors. Steele et al. (1997) estimates that 68% of elderly nursing home residents have dysphagia. Yearly, up to 200,000 new dysphagia cases are estimated as a result of stroke (Martino et al., 2000).

Oropharyngeal dysphagia is dysphagia that occurs from abnormalities originating in the oral and pharyngeal stages of swallowing. The techniques in this thesis focus on oropharyngeal dysphagia. An accepted clinical indicator of oropharyngeal dysphagia is the apparent abnormal movement of the hyolaryngeal complex from VFSS (Obeidin et al., 2014).

2.4 Evaluation of swallowing function

As mentioned in section 1, the VFSS is the gold standard for the evaluation of swallowing function in clinical and research settings. One particular protocol, the modified barium swallow study (MBSS), which employs videofluorographic media for visualization, is the gold standard for evaluating aspiration. In an MBSS, the patient's upper airway region is situated within the field of view of a videofluoroscope instrument. Then, while the instrument is recording, the patient is made to swallow a food or liquid bolus that has been mixed with barium sulfate, an X-ray contrast agent. Generally, perceptive judgements about the movement of the radio-opaque bolus and swallowing anatomy are used to infer swallowing health.

The major shortcoming of the MBSS, and qualitative VFSS analysis in general, is that its interpretation by clinicians is largely subjective. Several rater reliability studies using qualitative/categorically scored VFSS analysis have been done, however the methodologies used in these studies is heterogeneous with respect to the aspects of swallowing the participants were asked to evaluate, making it difficult to draw a single conclusion about rater reliability. These realities are reflected in the recent survey of rater reliability studies done by Baijens et al. (2013), which also suggests that in general, human VFSS evaluation is not thought to be reliable.

Rater reliability problems mean that clinicians cannot be expected to reliably apply the same standards to diagnose different clinical cases. Mis-diagnosis has the undesirable consequence of false negatives, in which case the patient's dysphagia goes unnoticed, or false positives, in which case a non-dysphagic patient may be sent down unnecessary and expensive treatment courses.

One solution to the subjectivity problem is to develop metrics that quantify swallowing parameters. Oral and pharyngeal transit times are two such metrics. Other quantitative measurements recently developed include the Penetration-Aspiration scale (Robbins et al., 1999), hyoid excursion (Kim and McCullough, 2008), and laryngeal elevation (Obeidin et al., 2014). Recently, Martin-Harris et al. (2008) developed the MBSImp, a multi-point assessment protocol each with a numeric scale to categorize clinical observations, as a

way of quantifying swallowing health comprehensively and practically. The protocol prescribes recording and assessment of 12 different swallows of boluses of various thicknesses. Using sets of categorical scores, 17 aspects of swallowing involved in the oral, pharyngeal, and esophageal phases of swallowing are evaluated in these 12 swallows. For example, one aspect, anterior hyoid excursion, is scored as 0, 1, or 2, for the observations of complete anterior movement, partial anterior movement, and no anterior movement, respectively. The motivation of the protocol is to increase rater reliability in dysphagia diagnosis by creating well-defined scoring categories to disambiguate clinical observations by neatly categorizing possible observations. Clinical evaluation of the protocol has shown favourable rater reliability.

Many of these measures are widely used and reported in research literature. However, determining them often requires intensive use of manual image manipulation tools or frame-by-frame analysis with video software. The techniques of Kim and McCullough (2008) and Obeidin et al. (2014) for example required manual coordinate marking of anatomical landmarks necessary for their calculations. Such analysis is feasible when the study involves a few clinical cases, but may become a practical limitation for research that requires analysis of significant numbers of clinical cases; research quality is progressively enhanced as the number of cases analyzed increases. This is one of the motivations for developing computer-aided-diagnosis tools (these are surveyed in section 3) that enhance research quality and was also the motivation for this thesis.

Chapter 3

Related Work

The work presented in this thesis fits in the category of computer-aided-analysis of swallowing. The motivation for this work was to enhance manual VFSS analysis by automating it with computer assistance. Previous works with similar motivations are surveyed in this chapter to help situate the contributions of this thesis.

In the early 1990s, the shortcomings of manual VFSS analysis stimulated research in enhancing VFSS analysis using image processing and computer vision techniques. Because of the heterogeneity in swallowing parameters adopted by clinicians to evaluate dysphagia, different attempts focused on enhancing analysis of different swallowing parameters. From a computer-aided-diagnosis perspective, these techniques can be classified into 2 categories: manual and semi-automatic.

3.1 Manual determination of swallowing parameters

The earliest attempts at enhancing VFSS analysis were manual techniques that used image manipulation tools to help annotate or markup some subset of a VFSS's frames and determine a few swallowing parameters. Humphreys et al. (1987) appears to have demonstrated the first use of early image manipulation tools to manually annotate the locations of anatomical landmarks in each frame of a digitized VFSS for use in clinical analysis. Crary et al. (1994) extended this form of analysis to demonstrate the determination of advanced swallowing parameters like pharyngeal contraction, laryngeal elevation, and laryngeal closure. Ludlow et al. (2007) used frame-by-frame annotation of landmarks in each frame of a VFSS to determine superior and anterior displacement of the hyoid to test hypotheses concerning the relationship between surface electrical stimulation around the airway region, and swallowing. Similarly, Burnett et al. (2003) used manual annotation of the thyroid prominence (Adam's apple bone) in VFSSs along with electromyography information to determine which muscle pair elevates the larynx during swallowing. More recently, Ceccarelli et al. (2011) demonstrated how manual annotation of anatomical landmarks in key VFSS frames can be used to calculate oral and pharyngeal transit times.

What is noteworthy about these techniques is that they do not employ any computer vision based techniques to enhance the annotation process, but rely on the user making entirely manual annotations to determine swallowing parameters through the use of macros in spreadsheet software (like Microsoft Excel)

using landmark coordinates obtained from professional image annotation software like the National Institute of Health's ImageJ ® or Peak Performance Technologies' Motus 2000 ® software.

3.2 Semi-automatic determination of swallowing parameters

More recent attempts at enhancing VFSS analysis employ computer vision techniques to semi-automate some part of a tedious annotation process and automatically determine several swallowing parameters. Chen et al. (2001) demonstrated an almost fully-automatic technique that uses feature matching between frames to track the locations of fiducial markers placed on the tongue; the markers are detected automatically and tracked without need for user input. Provided VF instrument settings are carefully chosen to provide contrast of radiopaque objects, fiducial markers can be made to show up as prominent features in VFSSs. The feasibility of using fiducial markers to track anatomical structures is limited to use in the oral cavity, as placement within the body requires surgery. Thus the utility of this technique is limited to studies in which the movements of fiducial markers is of interest.

Kellen et al. (2010) demonstrated a technique to track the movement of the hyoid bone on a frame-by-frame basis by registering the intensity of the pixels representing the hyoid bone between frames to determine its translation. The user is required to use an image painting tool to paint over the location of the hyoid bone in the first frame of the VFSS to initialize the program with the locations of the pixels representing the hyoid bone. Intensity-based registration is then done in subsequent frames to infer the translation of the hyoid bone. This allowed semi-automatic determination of hyoid excursion; decreased hyoid bone excursion is considered to be symptomatic of oropharyngeal dysphagia.

Aung et al. (2009) and Aung et al. (2010b) demonstrated semi-automatic techniques to determine oral and pharyngeal transit times. The user is required to initiate the algorithm by specifying two lines that delineate the boundaries between the oral cavity and the pharynx, and, at the start of the cricopharyngeal region. Spatio-temporal information of the bolus crossing these two lines is recorded. Transit times are automatically inferred by locating peaks in pixel intensity which correspond to the bolus crossing these two lines. Image registration between frames is used to determine rigid transformations that indicate head and neck movement are used to compensate for head and neck movement observed in VFSSs.

Aung et al. (2010a) demonstrated another technique based on Active Shape Modeling to track various skeletal, cartilaginous, and soft tissue structures visible in VFSS. The user is required to annotate 16 points in the first VFSS frame, allowing initialization of the Active Shape Modeling algorithm with 8 lines to indicate the boundaries of those structures and track them over subsequent VFSS frames. This technique is noteworthy because it could potentially be used to track the positions of any anatomical feature that contrasts itself against surrounding anatomy in digitized VFSS.

Noorwali (2013) demonstrated the use of the Kanade-Lucas tracker to track the anterior edges of the hyoid bone and epiglottis to infer timing of key events in swallowing involving their movements. However the

effectiveness of this method was not fully evaluated and it is not possible to know how well this technique works with VFSS images taken from the general population.

Hossain et al. (2014) demonstrated the use of template matching with a Haar classifier to track the positions of the hyoid bone and several vertebrae allowing for a variety of swallowing parameters involving only the hyoid to be inferred automatically. The user is required to draw a square over the location of the vertebrae in the first frame of the VFSS as a hint to their tracking algorithm so that the search area for template matching is localized. One disadvantage of such template matching techniques is that they require a training set of images to train the template matcher. This is generally feasible for anatomical features with high contrast like skeletal structures, but may be difficult for cartilaginous features which often do not present consistently in VFSSs and which exhibit significant variation between VFSS because of the lack of uniformity of instrument settings used to capture individual VFSSs.

The survey of recent techniques suggests that these techniques are specialized to determine one or a few swallowing parameters. In contrast, the technique presented in this paper can be used to answer a variety of research questions. In fact, the Coordinate Mapping technique which is used as the case study can determine up to 9 different swallowing parameters that are used in research of the hyolaryngeal complex. The technique presented in this thesis is a contribution to the body of research in semi-automated computer-aided-diagnoses of swallowing function that use computer vision techniques. It represents an advancement in this regard as its clinical research value has been demonstrated by using Coordinate Mapping as a case study.

As the VFSS is expected to remain the gold standard and the need to make objective swallowing measurements will rise, research in computer-aided-diagnosis of swallowing is an ongoing field. The technique in this thesis can be used to enhance any research process that requires annotation of landmarks in VFSSs in one or more frames. For example, theoretical swallowing research is often done with animal subjects in which fiducials are implanted on various points of swallowing anatomy and tracked over several frames. The number of frames requiring annotation can be upwards of 400, especially if VFSSs are recorded at higher frame rates. A workflow based on the KLT tracker can be used to annotate fiducials in each frame efficiently. Recently, Gerkema et al. (2015) demonstrated the use of image registration to eliminate the effect of head movement in VFSS recordings. Their technique requires the annotation of up to 30 landmarks in 3 VFSS frames to perform image warping in order to help clinicians account for head movement when making measurements. They found that increasing the number of landmarks increased accuracy of their technique. The technique described in this thesis could be used to annotate those landmarks semi-automatically and thus minimize manual effort and guard against risks of manual annotation.

Chapter 4

Design and Implementation

This section discusses the implementation details of the developed workflow, theory behind some of the core computer-vision techniques used, and the Coordinate Mapping technique used as the case study.

4.1 Coordinate Mapping

The pharyngeal phase of swallowing is a complicated synchrony of over twenty muscles, skeletal elements, and nervous pathways, working together to transfer the bolus into the esophagus while keeping the airway protected. Prior to the bolus entering the esophagus, elevation of the hyolaryngeal complex occurs and works to convert a respiratory conduit into a digestive tract (Obeidin et al., 2014).

Coordinate Mapping is a quantitative technique which allows for easy and reliable characterization of hyolaryngeal movement, an important aspect of dysphagia research. Prior to the development of Coordinate Mapping, multiple measurements made from a single VFSS were required to obtain one swallowing parameter. Coordinate Mapping allows for multiple parameters to be determined using one set of coordinates. The technique developed by Obeidin et al. (2014) and the modifications made to that technique for use in this study are described.

Prior techniques developed to measure swallowing parameters used a variety of patient-specific reference systems designed to correct for head and neck movements in the calculation of swallowing parameters and a variety of scalars, which are used to scale swallowing parameters when no calibration references -used to convert measurements made in pixel space to physical space- are visible in VFSS imagery. Because of this heterogeneity, measurements made using those techniques are incompatible with each other. Coordinate Mapping enforces some uniformity in swallowing measurements so that they can be compared. One further advantage of Coordinate Mapping is that it allows swallowing measurements to be made that reflect structure-function relationships in swallowing anatomy.

The Coordinate Mapping technique requires that the clinician scroll through a VFSS's frames and note the frame numbers that correspond to the hyoid being in positions of rest and maximal superior-anterior excursion. In each of these two frames, the locations of 9 anatomical landmarks are to be annotated:

1. L1: The intersection of the inferior line of the body of the mandible and the symphyseal outline of the mandible.

2. L2: The posterior nasal spine.

3. L3: The anterior edge of the C1 vertebra.

4. L4: The anterior-inferior edge of the C2 vertebra.

5. L5: The anterior-inferior edge of the C4 vertebra.

6. L6: The anterior-inferior edge of the hyoid bone.

7. L7: The inferior air column of the hypopharynx proximal to the upper esophageal sphincter.

8. L8: The posterior-inferior margin of the cricoid cartilage at the tracheal air column.

9. L9: The anterior-inferior margin of the cricoid cartilage at the tracheal air column.

Once these landmarks have been annotated, their coordinates can then be used to calculate a variety of swallowing measurements involving the hyolaryngeal complex. L4 is only used in calculations if no calibration references is visible in the image; in that case, all measurements are scaled to the distance between the L2 and L5 landmarks. Obeidin et al. (2014) demonstrates the calculation of 8 such measurements:

1. K1: Anterior hyoid movement with respect to vertebrae: Calculates the anterior movement of the hyoid (L6) in reference to a line approximating the vertebrae (L3 and L5) (Kim and McCullough, 2008)

2. K2: Superior hyoid movement with respect to vertebrae: Calculates the superior movement of the hyoid (L6) in reference to a line approximating the vertebrae (L3 and L5) (Kim and McCullough, 2008)

3. K3: Hyoid excursion with respect to mandible: Calculates the displacement of the hyoid (L6) towards a line approximating the mylohyoid line of the mandible (L1 and L3), approximating the function of the suprahyoid muscles (Martin-Harris et al., 2008)

4. K4: Hyoid excursion with respect to vertebrae: Calculates the movement of the hyoid (L6) in reference to a line approximating the vertebrae (L3 and L5) by combining the individual vector components of hyoid movement (Leonard et al., 2000)

5. K5: Superior laryngeal movement with respect to vertebrae: Calculates the direction of the larynx (L9) in reference to a line approximating the vertebrae (L3 and L5) (Logemann et al., 2000)

6. K6: Hyolaryngeal approximation: Calculates an approximation of the hyoid (L6) and larynx (L9) (Leonard et al., 2000)

7. K7: Laryngeal elevation: Calculates the displacement of the posterior larynx (L8) towards C1 (L3), giving an approximation of the action of the stylopharyngeus (Obeidin et al., 2014)

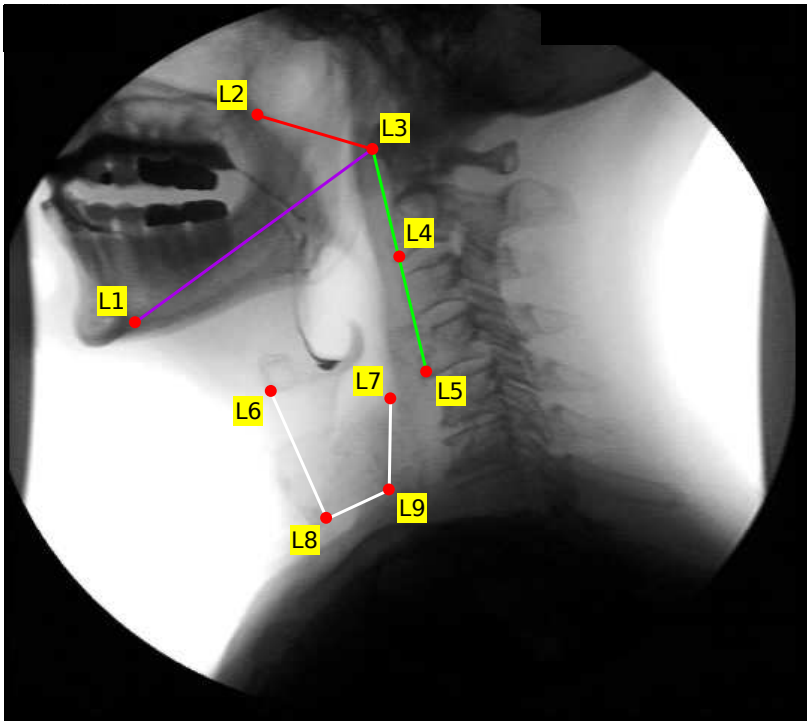

Figure 4.1: Locations of L1-L9 in a VFSS frame. The red line is an approximation of the hard palate, the magenta line of the mylohyoid line, the green line of the vertebral approximation, and the white line of the hyolaryngeal complex

8. K8: Pharyngeal shortening: Calculates the displacement of the upper esophageal sphincter (L7) towards the hard palate (L2), giving an approximation of the action of the palatopharyngeus (Obeidin et al., 2014)

Measurements made using landmark coordinates based on these two frames are considered extrema, because the measurements correspond to the rest and extreme positions of the hyolaryngeal complex. However, the same parameters can also be calculated for every frame of the VFSS in which the hyolaryngeal complex is in transit, to and from the rest position, resulting in a set of frame-by-frame kinematics. In this modification of Coordinate Mapping, the first key frame is always the frame in which the hyoid is at rest and kinematics are calculated by for each subsequent frame using the rest positions in the first frame as the reference. The ability to determine coordinates of landmarks and kinematics for each frame in transitwould allow for enhanced research analyses like multi-variate analysis of swallowing function in the pharyngeal stage and visualization and inverse modeling of pharyngeal swallowing in biomechanics simulation packages.

A typical VFSS containing only the frames for a swallow sequence can have between 30-50 frames. Annotating the 9 landmarks of Coordinate Mapping in all frames typically requires about 45 minutes. Research often requires the annotation of significant numbers of VFSSs and the time required to manually prepare data for a complete study can span a few days. As mentioned before, the manual data collection process is prone to accuracy and reliability problems.

4.2 Implementation of workflow

The KLT tracking algorithm as applied to tracking objects between successive images in a sequence operates under two assumptions: i) the object being tracked moves small distances between successive frames and ii) the pixel intensities representing the region of interest does not change. Very few applications of the algorithm respect these assumptions, but even under less-than-ideal circumstances, the algorithm is known to be robust enough to produce reliable tracking of features over short sequences of images. The oropharyngeal phase of swallowing recorded in most VFSSs are comprised of short sequences of images in which, between subsequent frames: i) pixel intensities representing anatomical features generally do not change greatly and ii) most anatomical features do not translate great distances in image space. Hence, the KLT tracking algorithm is a natural choice for the problem of tracking anatomical features in VFSS media.

4.2.1 Kanade-Lucas-Tomasi (KLT) Tracker

The KLT tracker is the synthesis of ideas presented in Lucas et al. (1981), Tomasi and Kanade (1991), and Shi and Tomasi (1994), applied to the problem of tracking feature points in image sequences. Lucas et al. (1981) presents an efficient technique for image alignment, the Lucas-Kanade (LK) image alignment, whereas Tomasi and Kanade (1991) and Shi and Tomasi (1994) explore applying the technique to the problem of feature tracking in image sequences. Any feature tracker that relies on the use of the LK algorithm to track the displacement of features between subsequent frames in an image sequence is considered a KLT tracker. The following overview of LK alignment and the KLT tracker is based on the derivation provided by Baker and Matthews (2004).

The goal of the image alignment problem is to align a template image, $T(\boldsymbol{x})$, to a reference image, $I(\boldsymbol{x})$, where $\boldsymbol{x} = (x, y)^T$ is a vector of pixel coordinates. Let $W(\boldsymbol{x}; \boldsymbol{p})$, an image warp function, denote the set of allowed image transformations, where $\boldsymbol{p} = (p_1, ..., p_n)^T$ is a vector of parameters. $\boldsymbol{W}(\boldsymbol{x}; \boldsymbol{p})$ maps a pixel in one image space to a location in another image space. The number of parameters is determined by the types of transformations allowed. In the case of feature tracking in 2D images where only translations of T are allowed (optical flow) $\boldsymbol{p} = (p_1, p_2)^T$. The goal of LK image alignment is to find the $\boldsymbol{p}$ for which the sum-of-squares difference in pixel intensities between a warped image I and the template image T, which stated as an optimization problem is:

$$\min_{\boldsymbol{p}} \quad \sum_{\boldsymbol{x}} \left[I(\boldsymbol{W}(\boldsymbol{x}; \boldsymbol{p})) - T(\boldsymbol{x}) \right]^2 \tag{4.1}$$

where the sum is performed over all pixels x in $T(x)$. Equation 4.1 is a non-linear optimization problem. To facilitate its optimization, the assumption that a current estimate of $\boldsymbol{p}$ is known is made and is iteratively solved for increments made to $\boldsymbol{p}$:

$$\min_{\Delta \boldsymbol{p}} \quad \sum_{\boldsymbol{x}} \left[I(\boldsymbol{W}(\boldsymbol{x}; \boldsymbol{p} + \Delta \boldsymbol{p})) - T(\boldsymbol{x}) \right]^2 \tag{4.2}$$

Linearization of the goal's objective function is achieved by approximating $I(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}+\Delta\boldsymbol{p}))$ using its first order Taylor series approximation:

$$\min_{\Delta\boldsymbol{p}} \quad \sum_{\boldsymbol{x}} \left[I(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p})) + \nabla I \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} \Delta\boldsymbol{p} - T(\boldsymbol{x}) \right]^2 \tag{4.3}$$

where $\nabla I = \begin{bmatrix} I_x & I_y \end{bmatrix}^T$ is the gradient of I (directional changes in image intensity computed in the x and y image directions) evaluated at $W(\boldsymbol{x};\boldsymbol{p})$, and $\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}}$ is the Jacobian of the warp function.

The closed form solution for the minimum $\Delta\boldsymbol{p}$ is:

$$\Delta\boldsymbol{p} = H^{-1} \sum_{\boldsymbol{x}} \left[\nabla I \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} \right]^T [T(\boldsymbol{x}) - I(\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}))] \tag{4.4}$$

where H is the Hessian matrix:

$$H = \sum_{x} \left[\nabla I \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} \right]^T \left[\nabla I \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} \right] \tag{4.5}$$

The LK algorithm iteratively solves for $\Delta\boldsymbol{p}$, updating $\boldsymbol{p}$ at each step: $\boldsymbol{p} = \boldsymbol{p} + \Delta\boldsymbol{p}$, until $\|\Delta\boldsymbol{p}\| \leq \epsilon$, where ϵ is a user-defined threshold that is an input to the algorithm. The intermediate steps of the algorithm are not fully relevant here and the interested reader is referred to Baker and Matthews (2004) for details, however, it is worth noting that the steps of the algorithm to solve for the minimal $\Delta\boldsymbol{p}$ is in essence, Gauss-Newton gradient descent non-linear optimization. This is worth noting because the Gauss-Newton gradient descent non-linear optimization algorithm may converge poorly or not converge at all if the Hessian matrix (equation 4.5) is ill-conditioned (Björck, 1996). In the case where $\boldsymbol{W}$ is limited to displacements, that is, $\boldsymbol{p} = (p_1, p_2)^T$, the set of allowed warps is $\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}) = (x + p_1, y + p_2)^T$, and the Jacobian of such a $\boldsymbol{W}$ is

$$\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} = \begin{bmatrix} \frac{\partial \boldsymbol{W}_x}{\partial p_1} & \frac{\partial \boldsymbol{W}_x}{\partial p_2} \\ \frac{\partial \boldsymbol{W}_y}{\partial p_1} & \frac{\partial \boldsymbol{W}_y}{\partial p_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.6}$$

The Hessian for LK image alignment for such a $\boldsymbol{W}$ is:

$$H = \sum_{\boldsymbol{x}} \left[\nabla I \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} \right]^T \left[\nabla \frac{\partial \boldsymbol{W}}{\partial \boldsymbol{p}} \right] = \sum_{\boldsymbol{x}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sum_{\boldsymbol{x}} I_x I_x & \sum_{\boldsymbol{x}} I_y I_x \\ \sum_{\boldsymbol{x}} I_x I_y & \sum_{\boldsymbol{x}} I_y I_y \end{bmatrix} \tag{4.7}$$

(Shi and Tomasi, 1994) demonstrated that features that are optimized for use with LK image alignment are those for which the eigenvalues (λ_1, λ_2) of the Hessian (equation 4.7) are both large and approximately equal. The tie-in to image quality is that such features are corners: points in images that lie at the intersection of edges in image intensity. Suggested features to track are those for which $\lambda_1 \approx \lambda_2$ and $min(\lambda_1, \lambda_2) > \gamma$; γ is a user-defined input to this method of finding good features to track. Another corner detection method, the Harris corner detector (Harris and Stephens, 1988), determines corners similarly using the same principle based on the eigenvalues of the Hessian.

The basic idea of the KLT tracker applied to the task of tracking a point feature through image sequences is to use corner detection to determine an image feature in the first frame of the image sequence, obtain a patch around it, and use LK image alignment with a warp function $\boldsymbol{W}(\boldsymbol{x};\boldsymbol{p}) = (x+p_1, y+p_2)^T$, to determine its apparent displacement, (p_1, p_2), in the next frame, use an image patch around $(x+p_1, y+p_2)^T$ to calculate its displacement in the next frame, and so on.

As mentioned before, one of the assumptions of the KLT tracker is that feature points move small distances between adjacent frames. To make the algorithm robust against large feature displacements, modern implementations enhance the KLT tracking algorithm by performing LK alignment between adjacent frames using image pyramids, which are multiscale representations of the same image at different resolutions. In this technique, T and I are sub-sampled by powers of 2, resulting in a set of representations that are at lower resolutions than normal, where N is typically 3 or 4 in most implementations. For example, if $N = 2$, then T and I would each be sampled at half their resolutions, an estimate of $\boldsymbol{p}$ produced using them via LK alignment, and subsequently used as the initial guess for LK alignment using T and I at their original resolutions.

Because of its natural application to the problem of tracking anatomical landmarks between adjacent frames in VFSS Media, its robustness, and the availability of implementations, the KLT tracker was chosen to semi-automate the task of landmark annotation. Conceivably, Harris corners could be chosen around anatomical landmarks to be tracked, a small patch around the chosen landmark initialized as the template image (T), and its apparent displacement ($\boldsymbol{p}$) in the next VFSS frame (I) determined using LK alignment, and LK alignment performed again using the new landmark location (update T) and the subsequent VFSS frame (new I), and so on until an end frame of the series of frames in which the landmark being tracked is reached. The collection of $\boldsymbol{p}$'s could be used to determine the locations of the anatomical landmark in each frame. In this thesis, this principle of operation was implemented to effect semi-automatic annotation of anatomical landmarks.

The MATLAB computing platform and its associated toolboxes contained trusted implementations of the KLT tracker and corner detection algorithms (the Harris corner detector is used) and was thus chosen for implementing a GUI tool to facilitate landmark annotation with the KLT tracker.

4.2.2 Implementation of workflow GUI in MATLAB

The workflow was implemented in MATLAB 2014a (Mathworks Inc., Natick MA), using tools from the Image Processing and Computer Vision toolboxes, as a graphical user interface (GUI) tool. The key components used are an implementation of the Harris corner detector and an implementation of the pyramidal KLT tracker.

The GUI is the main interface for annotating landmarks in a VFSS. In a typical annotation session, the user loads a VFSS into the GUI, annotate landmarks, and save the session. A screenshot of the GUI is shown in figure 4.2. Because annotation of landmarks across VFSS frames requires the ability to manipulate images,

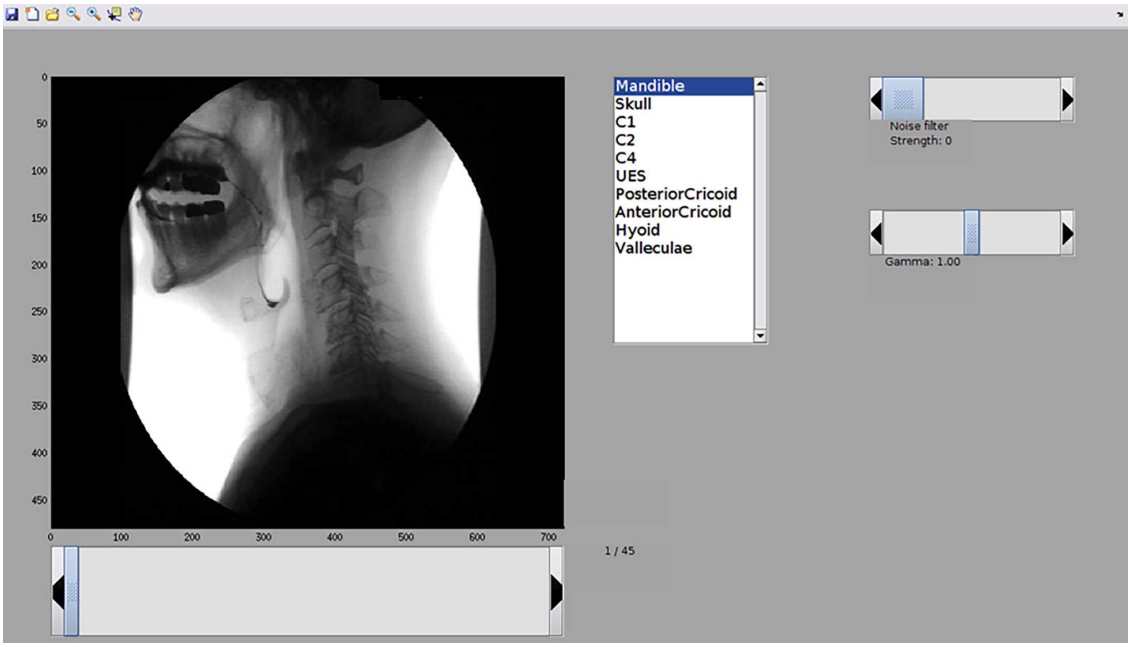

Figure 4.2: A screenshot of the graphical user interface for the semi-automatic landmark tracking tool. The entries in the list box control correspond to the landmarks L1-L9.

basic controls for zooming in and out, panning, adjusting the contrast, and filtering noise from the image are provided. Contrast adjustment is implemented with MATLAB's *imadjust* function. The user is provided with a slider control to adjust raise or lower contrast; internally the slider represents a range of floating point values between -1.0 and 1.0, which are then used as input to the imadjust functions gamma parameter which determines the relationship between the original image intensity and the contrast adjusted images intensity. The user is also provided with a slider to remove noise from images, which is implemented using the $medfilt2$ function which performs median filtering on the image; internally the slider's value is used to determine the width of the square operating window used in median filtering.

The workflow allows the user to annotate landmarks in a VFSS's frames in either semi-automatic or manual modes. Semi-automatic mode is intended to be used to annotate a landmark across several consecutive frames, whereas manual mode is intended to be used to make minor corrections to the annotation results of semi-automatic mode or to manually annotate a landmark.

The user invokes semi-automatic mode by pressing a pre-assigned key on the keyboard. Then the user selects the landmark to be tracked. The workflow then invokes the Harris feature detector to detect corners in a region-of-interest around this selection. The corner closest to the user's selection is selected and passed to KLT tracker implementation as the feature to track using LK image alignment. The workflow then steps the KLT tracker from the frame where the user invoked the KLT tracker to the subsequent frames, until the end of the VFSS is reached or the KLT tracker's minimized error falls below a predetermined threshold.

After the KLT tracker returns from its invocation, the user is allowed to scroll through the frames to verify the results of semi-automatic tracking. If the user disagrees with any subset of the frames that were

semi-automatically tracked, then they can be corrected in one of two ways: 1) invoking the semi-automatic tracker at the start frame where disagreement with the results starts, which is the recommended strategy when disagreement spans one or more groups of consecutive frames, or 2) by invoking the manual annotation tool, which allows the user to make corrections on a frame-by-frame basis, which is the recommended action to take when the annotations needing to be corrected are few and dispersed across several non-consecutive frames. The manual annotation tool is also invoked by pressing a pre-assigned key on the keyboard.

There are a number of input parameters to the Harris corner detector and the KLT tracker algorithms which determine how they behave in the implementation. These parameters are not adjustable by the user and appropriate values for these parameters were determined empirically by experimenting with VFSS not used in the evaluation of the workflow. The parameters used for the Harris corner detector are: $MinimumHarrisFeatureQuality = 0.01$, $Region - Of - Interest = [11, 11]$, $FilterSize = 5$. The parameters used for the KLT tracker algorithm are: $NumberofPyramidLevels = 3$, $BlockSize = [11, 11]$ (the size of the neighbourhood around the feature of interest within which the calculation of the gradient matrix in equation 4.7 is limited to).

Once all landmarks have been annotated in all frames of the VFSS, the workflow calculates the kinematic variables K1-K8 and allows the user to export the coordinates of L1-L9 in plain-text format so that it can be imported into tools downstream in the research process. For example, the coordinates can be loaded into MorphoJ (Klingenberg, 2011) as it has been done to perform multivariate analysis or in a biomechanics simulator like ArtiSynth (Fels et al., 2006) to visualize movement of L1-L9 using generic anatomical models.

Chapter 5

Evaluation

This section describes the data used to evaluate the developed workflow, the evaluation methodology, and presents the results of the evaluation.

5.1 Videofluoroscopic Data Used

To evaluate the workflow, 8 VFSSs were obtained from three clinical sources: one from the National Military Audiology and Speech Pathology Center at the Walter Reed National Military Medical Center, one from the department of otolaryngology at the Medical University of South Carolina, and 6 from the department of anatomy and cell biology at Georgia Regents University. Each VFSS was recorded with a different VF instrument and chosen for its high image quality after digitization. Permission to use these VFSSs as part of this thesis was obtained from the University of Saskatchewan's Biomedical Research Ethics Board (ethics approval number 357205-14). Because the Coordinate Mapping technique is only relevant to the subset of a VFSS recording in which oropharyngeal swallowing is observed, each original VFSS was clipped to contain only the frames containing this swallowing phase; this corresponds to the frames where the hyoid bone is seen to move from rest, to maximum excursion, and back to rest. After this clipping, VFSSs contained between 31 and 65 frames. Ground truth was prepared for each of the 8 VFSSs by manually annotating each of the landmarks in all frames of the VFSS. A clinician experienced in the use of Coordinate Mapping annotated 4 of the VFSSs while the author who was trained in the use of Coordinate Mapping by an experienced clinician annotated the other 4.

All measurements in the evaluation were made in pixels as calibration references were not available for the 8 VFSSs. While this means that each VFSS's inter-pixel distance represents a different physical distance, it is not uncommon to report such measurements in the clinical discipline in pixels. The resolutions and field-of-views of VFSSs used was high enough to ensure that physical distances corresponding to inter-pixel distances do not vary too greatly between VFSSs.

5.2 Evaluation methodologies and results

5.2.1 Landmark coordinate determination accuracy study

Description

To be a valid research tool, the workflow must have a level of accuracy that is clinically acceptable. The purpose of this study was to determine whether the accuracy of the annotations produced by the workflow is clinically acceptable.

A common experimental procedure was followed to annotate landmarks (L1-L6) in each of the 8 VFSSs. Each VFSS was annotated in a separate sitting and no VFSS was annotated across multiple sittings. A sitting is defined as opening the tool and loading a VFSS for annotation. Landmarks were annotated sequentially in ascending numerical order. For each landmark, the user scrolled to the first frame of the video. Then the automatic tracker tool was initialized and the ground truth location of that landmark for that frame was specified as the initialization point for the tracker. Then the user reviewed the tracking results. If the user disagreed with the tracking result at any frame, and if the number of consecutive frames across which there was disagreement was judged to be significant (greater than 5 frames), then the user invoked the automatic tracker again initializing it with the location of the ground truth coordinate for that landmark at that frame. If the number of consecutive frames was judged to be few, then the manual annotation tool was invoked and the consecutive frames were annotated manually in point and click fashion. A final review of the complete tracking result for the landmark was then done. This annotation strategy was then repeated for each of the other landmarks.

For each landmark in each VFSS, the root-mean-square-error (RMSE) between ground truth and the workflow's results was determined, resulting in a set of 6 RMSE numbers for each VFSS. The RMSE calculations were done separately for both the X and Y coordinates. These were then averaged across VFSSs on a per-landmark basis to determine the study-wise RMSE for the landmark.

It took roughly 5 minutes to annotate each VFSS and it was not necessary to re-initialize the automatic mode of the workflow more than 5 times for each VFSS.

Results

The results of this study are shown in figure 5.1. Error bars denote standard deviation of the distribution of errors with the half width of the war denoting a distance of one standard deviation from the mean. The results show that mean RMSE between landmarks was bounded above by 2 pixels. Errors between 1 to 4 pixels would generally be considered clinically insignificant (William Pearson Jr., personal communication, August 22, 2014) when taking measurements for diagnosis or research.

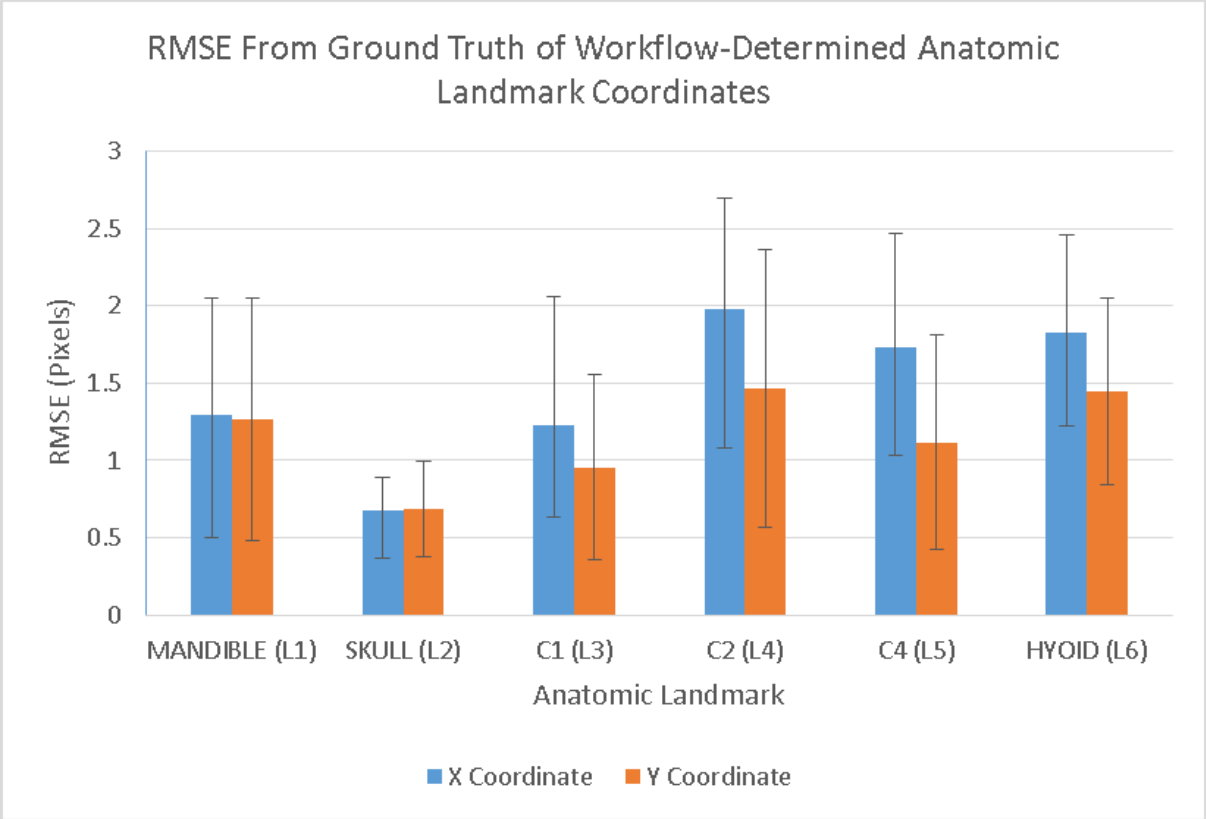

Figure 5.1: Per-landmark average root-mean-square-error (RMSE) between workflow-determined landmark coordinates (L1-L6) and ground truth manually-annotated coordinates (L1-L6) averaged across 8 VFSS. Error bars denote standard deviation from the mean

5.2.2 Accuracy of kinematic variables

Description

Once landmarks are annotated, the next step in research settings is to use the set of coordinates to determine some set of swallowing parameters that will help answer a particular research question. For the workflow to have research value, it must be able to determine swallowing parameters with acceptable accuracy. It is possible that individual landmarks that have been determined with acceptable accuracy may propagate their errors in calculations of swallowing parameters that use several landmark coordinates to calculate a single parameter. To evaluate whether the workflow determined swallowing parameters with acceptable accuracy, the kinematic variables K1-K8 were calculated using landmark coordinates from the previous evaluation and were compared to K1-K8 calculated from ground truth landmark annotations. These were calculated and reported for each VFSS video in the same manner as for landmark annotations, using RMSE as the error metric. Further, we investigated whether or not the workflow determined K1-K8 were correlated with ground truth determined K1-K8. Correlation was determined by calculating and averaging the Pearson correlation coefficient in the same manner as for RMSE of anatomical landmarks.

Results

Figure 5.2 presents RMSE of K1-K8 determined using the workflow compared to ground truth. Figure 5.3 presents the Pearson r for K1-K8.

5.2.3 Intra-rater sensitivity analysis

Description

The ability of the workflow to produce results that are correlated between repeated usages for analysis of the same VFSS is important. A user who repeats the workflow on the same VFSS is unlikely to choose the same pixels, whether to initialize the automatic tracker in the workflow or annotate manually. Instead, deviations of a few pixels are likely. How large the annotations are likely to be from the previous annotation is a question of how experienced the user is. More experienced users are more likely to perform repeat annotations with little variance in the coordinate locations than less experienced ones. An assumption of 2 to 4 pixels of variance between repeated runs for an experienced rater is used.

This evaluation attempts to determine whether repeated runs of the workflow on the same VFSS by the same user will yield highly correlated results. This is done by repeating the workflow for each VFSS 4 times. The same annotation strategy as described before is used, however, for each of the four runs, when the automatic tracker is used, the initialization pixel is chosen such that it is deviated from the ground truth by 2-4 pixels in each of the 4 compass directions. Thus, in the first repeat, for each initialization of the tracker, the initialization point is chosen to be 2-4 pixels North of the ground truth, and in the second repeat, 2-4

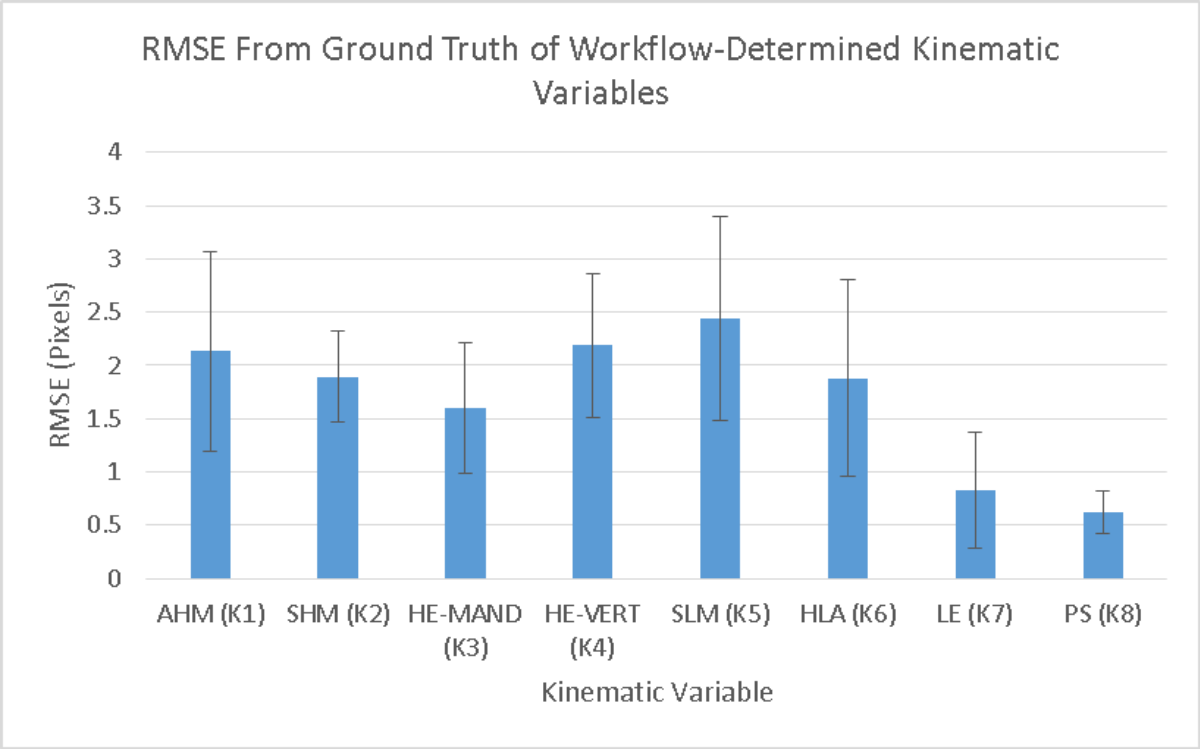

Figure 5.2: Per-kinematic variable RMSE between workflow-determined kinematic variables (K1-K8) and ground truth manually-annotated kinematic variables (K1-K8) averaged across 8 VFSS. Error bars denote standard deviation.

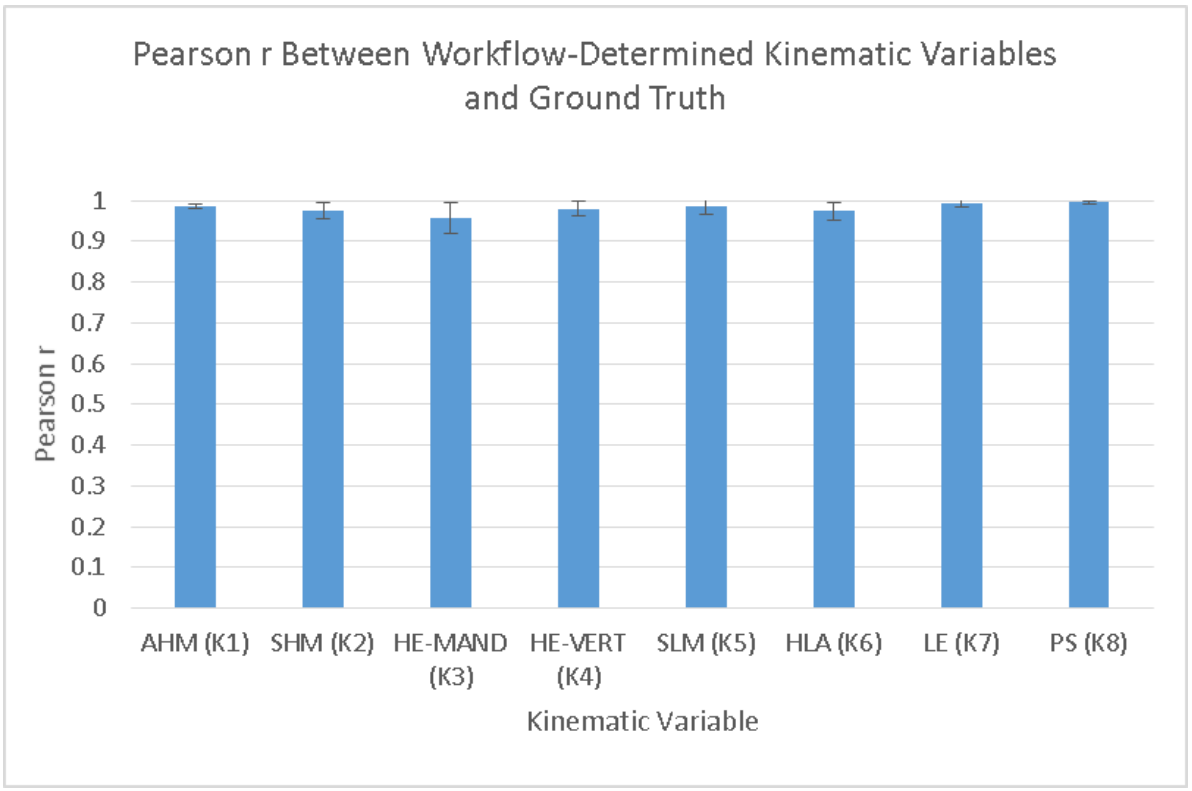

Figure 5.3: Per-kinematic variable Pearson r correlation between workflow-determined kinematic variables (K1-K8) and ground truth manually-determined kinematic variables (K1-K8) averaged across 8 VFSS. Error bars denote standard deviation.

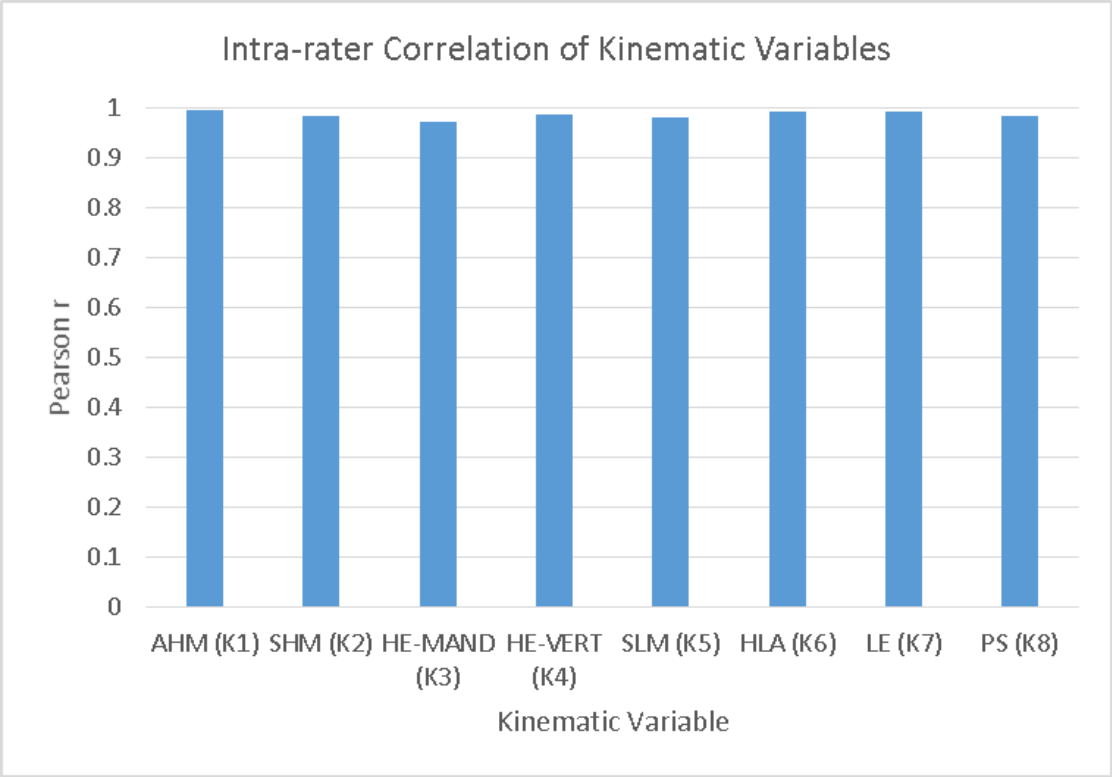

Figure 5.4: Per-kinematic variable average Pearson r correlation computed pairwise between repeated trials of the workflow-determined kinematic variables (K1-K8). The standard deviation of the correlation is less than 0.038 for all variables.

pixels West of the ground truth, and so on. The Pearson correlation was calculated for each VFSS in the same way for the previous evaluation. Then the average Pearson r for K1-K8 was determined by averaging across the 4 repeats. Then these were averaged across all 8 VFSSs. A proper analysis would call for a formal intra-rater reliability study in which several users use the workflow, however, due to a shortage of users trained in Coordinate Mapping and the difficulty of distributing the prototyped version of this MATLAB-centric workflow, this was not feasible for the thesis and is planned as future work.

Results

Figure 5.4 shows the Pearson r determined for each kinematic variable. $r > 0.90$ is generally considered an indication of strong correlation. The results argue that a trained rater who repeats the Coordinate Mapping -or workflow that requires frame-by-frame landmark annotation- will likely produce results that are strongly correlated.

Chapter 6

Discussion

In this section, the results of the evaluation of the workflow, its limitations, and external factors that can affect its efficiency are discussed.

6.1 Analysis of results

6.1.1 Overall accuracy of landmark coordinates

The results of section 5.2.1 show that the upper bound for the error is 1.9 pixels which is for the X coordinate of L4. In all cases, the error for the X coordinate is higher than the error for the Y coordinate. In the VFSSs used, the patient's head, and thus the landmarks L1-L5, all experience greater horizontal movement compared to vertical movement. When the tracker is invoked to annotate long sequences, the changes in intensity from the template image the tracker uses to initialize itself appear to cause it to drift away slightly from the actual location of the landmark. L2 was annotated with the greatest accuracy and this is likely due to the fact that the landmark is at the sharp intersection of two cranial skeletal features that are clearly visible in VFSS and the intensities of pixels representing those anatomical regions is stable as the VFSS progresses. L4 and L5 had higher errors because the pixels representing those vertebral bones' anterior-inferior edges of those vertebral bones often change intensity as a VFSS progresses resulting in images of the region that are different from the template image used by the KLT tracker. The hyoid bone (L6) was also one of the landmarks tracked with lower accuracy. Most VFSSs are taken in the lateral view (side view) of the patient, and in this view the outline of the hyoid bone appears as a small L shape. The small size of the hyoid bone and its lower radiodensity means that it does not appear as contrastingly as other larger skeletal features in VFSSs. In frames where the hyoid bone moves faster than the sampling rate of the VF instrument can faithfully capture its outline, it appears blurred. It is the least clear of all skeletal landmarks and moves the most and is generally harder to track. Overall, the workflow produces results that are accurate to within 2 pixels in the X and Y directions. Such pixel distances represent physical distances on the order of a few millimeters, which are considered to be clinically insignificant (William Pearson Jr., personal communication, August 22, 2014).

6.1.2 Overall accuracy of calculated kinematic variables

The results of section 5.2.2 can be understood in terms of the accuracy of the number of landmarks L1-L6 used in the calculation of the kinematic variable and the accuracy of landmark tracking as shown in section 5.2.2. The results show that generally, the more landmarks a kinematic variable's calculation depends on and the higher those landmarks' RMSEs, the lower the accuracy of the kinematic variable. Table 6.1 shows the number of landmarks used in each kinematic variable's calculation:

Table 6.1: Number of workflow determined landmarks that each kinematic variable's calculation depends on

Kinematic variable	Landmarks used in calculation
K1	L3, L5, L6
K2	L3, L5, L6
K3	L1, L3, L6
K4	L3, L5, L6
K5	L3, L4, L6
K6	L6
K7	L3
K8	L2

K8 was determined with the lowest average RMSE because its calculation depended only on L2 which was the most accurately annotated landmark. Similarly, K7, the second most accurately tracked landmark can be explained by noting that its calculation depended only on L3 which was the 2nd most accurately tracked landmark. Similarly for K3. K1, K2, K4, and K5 all use L3, L5, and L6 and have similar RMSEs.

For all landmarks however, the mean RMSE was bounded above by 2.5 pixels. As mentioned before, errors of such small magnitude are generally considered clinically insignificant(William Pearson Jr., personal communication, August 22, 2014).

6.1.3 Intra-rater sensitivity analysis

The results of section 5.2.3 show that perturbing the initial location of the landmark in the frames to be tracked by 2-4 pixels, simulating a level of intra-rater variation due to repeated uses of the workflow by the same user, resulted in annotations that produced kinematic variable calculations that were highly correlated ($r > 0.9$). This suggests that the workflow is not sensitive to small variations in the user's choice of the location of the landmark used in the initialization of the tracker algorithm, and could be used to produce annotation results that are similar between repeated uses of the workflow on the same VFSS by the same

user.

6.2 Limitations of the workflow

Although the workflow shows promising results for skeletal landmarks, it has several limitations.

Soft-tissue landmarks and landmarks with low radio-opacity cannot be reliably tracked because they are not suitable for tracking by the KLT tracker. One solution to this problem is to adjust the settings of VF instruments so that soft-tissue landmarks are visible with greater contrast in VFSS imagery. VF instruments settings are usually chosen by radiologists to minimize the radiation exposure from the VF instrument to the point where image quality allows qualitative judgements about swallowing health to be made. Physiologically speaking, landmarks L1-L5 do not move significantly during swallowing. In controlled settings like a VFSS recording, they move even less between frames, which is a key assumption for the KLT tracker. This is not always the case for L6 whose anterior-superior movement is significant during the oropharyngeal phase of swallowing. This rapid movement of L6 between frames and the low sampling rate of most VF instruments causes the hyoid to appear blurred in some frames and its appearance to change throughout the recording, presenting a challenge for the KLT tracking algorithm. Hence, the workflow currently does not deal well with anatomical landmarks that appear blurred in VFSS frames because the sampling rates of clinical VF instruments cannot faithfully capture the shapes of such objects. This restricts the kind of landmarks the workflow can be used to annotate to slow moving skeletal anatomy.

Currently, the workflow keeps the settings of the KLT tracker static across a tracking session. One particular setting, the size of the window around the feature of interest within which the spatial gradient matrix is computed (equation 4.7) is kept fixed. This can be a problem when the VFSS is of a patient swallowing a contrast-enhanced bolus and the bolus, and the image dimensions are small enough that the contrast-enhanced bolus encroaches into the window, causing a sudden distraction to the tracking algorithm. One potential solution is to ensure that VFSSs are taken with high resolutions and field-of-view's that are restricted to anatomies of interest only, ensuring that a reasonably sized window can be selected such that it is unlikely a distraction enters it.

6.3 External factors that affect workflow effectiveness

A number of factors beyond the control of the domain of computer vision affect the ability of VFSSs to be used with computer vision techniques like the developed workflow. VFSSs are usually taken by radiologists, who set VF instrument settings. Settings are chosen to minimize radiation exposure while at the same time maintaining a level of image quality that allows qualitative judgements about swallowing health to be made. Generally, radiologists do not take VFSSs with the interests of quantitative analysis in mind as qualitative diagnosis is still the norm. Digitized VFSSs that are to be used in computer vision based analysis should have the following qualities: uniform image brightness, high frame rates, high contrast of soft tissue and skeletal

landmarks, and high resolution. However, these require that VF instrument settings be adjusted in ways that would increase radiation dosage. If VFSSs are to be used in quantitative computer vision based analysis workflows, then radiologists must consider using VF instrument settings that allow for high quality imagery to be recorded. In clinical settings, VFSSs are usually digitized after being delivered from a VF instrument for end use and archival. Further care should be taken during this digitization step to ensure that VFSSs are digitized without introduction of digitization or transcoding artifacts like interlacing lines and compression blocks, which are easily avoidable distractions to the KLT Tracker.

Chapter 7

Conclusion

A novel semi-automatic workflow to semi-automate the process of annotating anatomical landmarks in VFSS imagery was developed and evaluated. The contributions of this thesis are:

1. The design and implementation of a computer-assisted workflow for the semi-automatic annotation of anatomical landmarks in VFSS images, to allow clinicians and researchers of swallowing function to efficiently annotate landmarks, in clinical and research processes that require frame-by-frame annotation of landmark coordinates to make swallowing measurements or answer research questions in oropharyngeal dysphagia. The workflow uses the KLT tracking algorithm to track landmarks between successive VFSS frames to enhance current manual annotation processes by reducing annotation time.

2. Evaluation of the accuracy of the workflow using Coordinate Mapping to determine swallowing parameters from annotated VFSS images, which showed that the workflow was able to determine the positions of anatomical landmarks' coordinates with a mean RMSE bounded above by 2 pixels in both the X and Y coordinates in image space. These annotations were used to determine several kinematic variables, the mean RMSEs of which were found to be bound above by 2.5 pixels. In VFSS imagery, such pixel distances correspond to clinically insignificant distances (William Pearson Jr., personal communication, August 22, 2014) in physical space. Thus, the workflow has been demonstrated to accurately annotate landmark positions and determine kinematic variables determined from them.

3. Evaluation of the intra-rater sensitivity of the workflow to determine the similarity of the annotation results produced between repeated uses of the workflow on the same VFSS by the same user, which suggests that it can be used to produce annotations of the same VFSS by the same user that are highly correlated.

4. Deployment of the workflow in a research setting to facilitate data collection for multivariate analysis of oropharyngeal dysphagia using Coordinate Mapping. Tran et al. (2016) used the workflow to annotate 18 VFSSs. Annotations produced by the workflow were then used in multi-variate analysis to determine the improvements in swallowing mechanics resulting from respiratory-swallow phase training (Tran et al., 2016).

7.1 Future work

The results of the workflow's evaluation demonstrate that it can enhance the process of annotating skeletal anatomical landmarks. This leads to some obvious research questions that might be pursued as follow up to this thesis. The workflow was only evaluated on skeletal landmarks. Soft-tissue and cartilaginous landmarks do not appear clearly and consistently in contemporary VFSS media and being in violation of the assumptions of image quality the KLT tracker was built upon, they are unsuitable for tracking with it. The soft tissue landmarks in Coordinate Mapping (L7-L9) are often not apparent in VFSS frames especially when the hyolaryngeal complex is in transit. The majority of landmarks in Coordinate Mapping are skeletal landmarks (L1-L6) and thus the overall annotation experience based on that technique was enhanced with the workflow. However, the number of soft-tissue landmarks needing to be tracked and their apparent quality in VFSS imagery will vary based on the experimental technique being used to answer a particular research question. The workflow is not suitable for use when the majority of landmarks needing to be annotated are soft-tissue types. There is a dearth of research on tracking soft-tissue landmarks. Only Aung et al. (2010a) has attempted the evaluation of a technique for semi-automatic tracking of soft-tissue landmarks: Active Shape Modeling (ASM). Future work should explore how ASM could be used in combination with a feature tracking technique like the KLT tracker to provide clinicians with the ability to semi-automatically annotate both skeletal and soft-tissue landmarks.

The workflow was evaluated with 8 VFSSs. Procuring VFSSs for evaluation of semi-automatic annotation techniques is a challenge because ethics approvals are a significant research hurdle and there is a dearth of VFSS imagery that has been taken with the intent of maintaining a level of image quality that is suitable for processing after digitization. Future work should evaluate the generalizability of this thesis's findings by repeating evaluation on a larger number of VFSSs.

Due to the heterogeneity in the diagnosis of dysphagia and the swallowing parameters used to make diagnoses, finding clinicians trained in the use of a particular swallowing protocol was difficult. As a result, a formal inter-rater study could not be conducted. There are several studies in the literature that assess the inter-rater reliability of various dysphagia assessment methods. However, Obeidin et al. (2014) conducted an inter-rater reliability study in which 6 raters determined the kinematic variables of Coordinate Mapping using the two key frames method described in section 4.1. 40 VFSSs were used in the study. It would be of research value to determine whether the workflow developed in this thesis would improve inter-rater reliability by conducting a formal study using a similar number of VFSSs. This thesis conducted a sensitivity study to determine the similarity of workflow produced annotations when the same clinician annotates the same VFSS repeatedly, in place of a formal intra-rater reliability study. It would also be of research value to determine whether the workflow can be used to enhance intra-rater reliability. A possible study might involve several researchers using the workflow to annotate several VFSSs multiple times and determine the similarity of repeated annotations.

This thesis demonstrated the research value of the workflow by demonstrating its ability to make annotation of skeletal landmarks efficient and producing accurate annotation and swallowing kinematic results using Coordinate Mapping. To further test the belief that the workflow enhances research processes where frame-by-frame annotation of skeletal landmarks is required, the workflow should be evaluated using other clinical techniques in dysphagia research.

References

M. Aung, S. Hamdy, M. Power, S. Stanschus, and J. Goulermas. Measuring bolus transit times from videofluoroscopy using image profiles and particle swarm optimisation. In *Developments in eSystems Engineering (DESE), 2009 Second International Conference on*, pages 117–122. IEEE, 2009.

M. Aung, J. Goulermas, S. Stanschus, S. Hamdy, and M. Power. Automated anatomical demarcation using an active shape model for videofluoroscopic analysis in swallowing. *Medical engineering & physics*, 32(10): 1170–1179, 2010a.

M. S. Aung, J. Y. Goulermas, S. Hamdy, and M. Power. Spatiotemporal visualizations for the measurement of oropharyngeal transit time from videofluoroscopy. *Biomedical Engineering, IEEE Transactions on*, 57 (2):432–441, 2010b.

L. Baijens, A. Barikroo, and W. Pilz. Intrarater and interrater reliability for measurements in videofluoroscopy of swallowing. *European journal of radiology*, 82(10):1683–1695, 2013.

S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International journal of computer vision*, 56(3):221–255, 2004.

A. Björck. *Numerical methods for least squares problems*. Siam, 1996.

T. A. Burnett, E. A. Mann, S. A. Cornell, and C. L. Ludlow. Laryngeal elevation achieved by neuromuscular stimulation at rest. *Journal of applied physiology (Bethesda, Md.: 1985)*, 94(1):128–134, Jan 2003. LR: 20131121; GR: Z01 NS-02980/NS/NINDS NIH HHS/United States; JID: 8502536; ppublish.

M. Ceccarelli, A. Colaprico, L. D. Vito, S. Marotta, and M. Piscitelli. A semi-automatic measurement system for the swallowing analysis in videofluoroscopy. In *Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on*, pages 125–130. IEEE, 2011.

Y. Chen, J. L. Barron, D. H. Taves, and R. E. Martin. Computer measurement of oral movement in swallowing. *Dysphagia*, 16(2):97–109, 2001.

L. R. Cherney, P. Gardner, J. A. Logemann, L. A. Newman, T. O'Neil-Pirozzi, C. R. Roth, N. P. Solomon, C. Sciences, and D. C. T. R. Group. The role of speech-language pathology and audiology in the optimal management of the service member returning from iraq or afghanistan with a blast-related head injury: position of the communication sciences and disorders clinical trials research group. *The Journal of head trauma rehabilitation*, 25(3):219–224, May-Jun 2010. LR: 20131121; JID: 8702552; RF: 49; ppublish.

M. A. Crary, M. K. Butler, and B. O. Baldwin. Objective distance measurements from videofluorographic swallow studies using computer interactive analysis: technical note. *Dysphagia*, 9(2):116–119, 1994.

A. M. Dietsch, W. G. Pearson, L. K. Heard, C. B. Rowley, K. E. Dietrich-Burns, and N. P. Solomon. Swallowing mechanics associated with penetration-aspiration in trauma patients, 2015.

S. Fels, F. Vogt, K. V. D. Doel, J. Lloyd, I. Stavness, and E. Vatikiotis-Bateson. Artisynth: A biomechanical simulation platform for the vocal tract and upper airway. In *International Seminar on Speech Production, Ubatuba, Brazil*, 2006.

M. H. Gerkema, M. H. M. Ketel, E. P. J. Smits, and M. D. van der Stoel. Eliminating global head movement during swallowing, 2015.

H. Gray. *Anatomy of the human body*. Lea & Febiger, 1918.

C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Citeseer, 1988.

I. Hossain, A. Roberts-South, M. Jog, and M. R. El-Sakka. Semi-automatic assessment of hyoid bone motion in digital videofluoroscopic images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2(1):25–37, 2014.

B. Humphreys, R. Mathog, R. Rosen, P. Miller, J. Muz, and R. Nelson. Videofluoroscopic and scintigraphic analysis of dysphagia in the head and neck cancer patient. *The Laryngoscope*, 97(1):25–32, 1987.

P. M. Kellen, D. L. Becker, J. M. Reinhardt, and D. J. V. Daele. Computer-assisted assessment of hyoid bone motion from videofluoroscopic swallow studies. *Dysphagia*, 25(4):298–306, 2010.

Y. Kim and G. H. McCullough. Maximum hyoid displacement in normal swallowing. *Dysphagia*, 23(3): 274–279, 2008.

C. P. Klingenberg. Morphoj: an integrated software package for geometric morphometrics. *Molecular ecology resources*, 11(2):353–357, 2011.

R. J. Leonard, K. A. Kendall, S. McKenzie, M. I. Gonalves, and A. Walker. Structural displacements in normal swallowing: a videofluoroscopic study. *Dysphagia*, 15(3):146–152, 2000.

J. A. Logemann, B. R. Pauloski, A. W. Rademaker, L. A. Colangelo, P. J. Kahrilas, and C. H. Smith. Temporal and biomechanical characteristics of oropharyngeal swallow in younger and older men. *Journal of Speech, Language, and Hearing Research*, 43(5):1264–1274, 2000.

B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.

C. L. Ludlow, I. Humbert, K. Saxon, C. Poletto, B. Sonies, and L. Crujido. Effects of surface electrical stimulation both at rest and during swallowing in chronic pharyngeal dysphagia. *Dysphagia*, 22(1):1–10, 2007.

B. Martin-Harris, M. B. Brodsky, Y. Michel, D. O. Castell, M. Schleicher, J. Sandidge, R. Maxwell, and J. Blair. Mbs measurement tool for swallow impairmentmbsimp: establishing a standard. *Dysphagia*, 23 (4):392–405, 2008.

R. Martino, G. Pron, and N. Diamant. Screening for oropharyngeal dysphagia in stroke: insufficient evidence for guidelines. *Dysphagia*, 15(1):19–30, 2000.

K. Matsuo and J. B. Palmer. Anatomy and physiology of feeding and swallowing: normal and abnormal. *Physical Medicine and Rehabilitation Clinics of North America*, 19(4):691–707, 2008.

S. Noorwali. Semi-automatic tracking of the hyoid bone and the epiglottis movements in digital videofluoroscopic images. 2013.

T. Obeidin, B. Pearson, and A. Person. Coordinate mapping of hyolaryngeal mechanics in swallowing. *Journal of visualized experiments : JoVE*, 1(2):3, 2014.

W. G. Pearson, J. Blair, and B. Martin-Harris. Swallowing mechanics associated with swallowing impairment, 2015.

J. M. Pisegna, W. P. G., S. E. Langmore, and S. Kumar. Morphometric analysis of the swallowing mechanism defines differences in post-stroke aspirators, 2015 2015.

O. Remesz. Larynx - antero-lateral view, with external muscles of larynx visible, 2008. URL `https://commons.wikimedia.org/wiki/File:Larynx_external_en.svg`.

J. Robbins, J. Coyle, J. Rosenbek, E. Roecker, and J. Wood. Differentiation of normal and abnormal airway protection during swallowing using the penetrationaspiration scale. *Dysphagia*, 14(4):228–232, 1999.

D. F. Roden and K. W. Altman. Causes of dysphagia among different age groups: a systematic review of the literature. *Otolaryngologic clinics of North America*, 46(6):965–987, 2013.

J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994.

C. M. Steele, C. Greenwood, I. Ens, C. Robertson, and R. Seidman-Carlson. Mealtime difficulties in a home for the aged: not just dysphagia. *Dysphagia*, 12(1):43–50, 1997.

L. Sura, A. Madhavan, G. Carnaby, and M. A. Crary. Dysphagia in the elderly: management and nutritional considerations. *Clin Interv Aging*, 7(287):98, 2012.

C. Tomasi and T. Kanade. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.

T. T. A. Tran, B. M. Harris, and W. G. P. Jr. Improvements resulting from respiratory-swallow phase training visualized in patient-specific computational analysis of swallowing mechanics. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 0(0):1–7, 2016. doi: 10.1080/21681163.2016.1152567. URL `http://dx.doi.org/10.1080/21681163.2016.1152567`.

Appendix A

Formulas for calculating kinematic variables K1-K8

The formulas for calculating the kinematic variables used in the Coordinate Mapping technique are not fully explained in Obeidin et al. (2014). The formulas are themselves scattered across several publications. They are presented here for convenience.

These formulas calculate kinematics between two frames of a VFSS based on distances between sets of landmarks in these frames. The frames do not necessarily have to be adjacent, however, because in this thesis, the kinematics were calculated for each frame assuming the first frame of the VFSS depicted the hyoid at rest, the first of the two frames is the first frame of the VFSS. The notation $|LX^{(i)}LY^{(i)}|$ refers to the Euclidean distance between landmarks LX and LY in the i-th frame. Similarly, $|LX^{(j)}LY^{(j)}|$ refers to the Euclidean distance between landmarks LX and LY in the j-th frame. The notation $\theta^{(i)}_{LX,LY,LZ}$ refers to the angle between landmarks LX, LY, and LZ in the i-th frame.

As mentioned in section 4.1, kinematic measurements can be scaled to either physical coordinates, if a calibration reference is visible in the VFSS, or to the patient's C2-C4 distance. In the subsequent formulas, s refers to a scaling factor that is either the length of the calibration reference in image coordinates or the patient's C2-C4 distance in image coordinates. In both cases, the measurement is usually taken from the frame at which the hyoid is at rest.

A.1 K1: Anterior hyoid movement

$$K1 = \frac{\sin\left(\theta^{(j)}_{L6,L3,L5}\right) \times |L6^{(j)}L3^{(j)}| - \sin\left(\theta^{(i)}_{L6,L3,L5}\right) \times |L6^{(i)}L3^{(i)}|}{s}$$

A.2 K2: Superior hyoid movement

$$K2 = \frac{\sin\left(\theta^{(j)}_{L1,L6,L3}\right)|L6^{(j)}L1^{(j)}| - \sin\left(\theta^{(i)}_{L1,L6,L3}\right)|L6^{(i)}L1^{(i)}|}{s}$$

A.3 K3: Hyoid excursion with respect to mandible

$$K3 = \frac{\sin\left(\theta^{(j)}_{L3,L1,L6}\right)|L6^{(j)}L1^{(j)}| - \sin\left(\theta^{(i)}_{L3,L1,L6}\right)|L6^{(i)}L1^{(i)}|}{s}$$

A.4 K4: Hyoid excursion with respect to vertebra

$$K4 = \frac{\sqrt{K1^2 + K2^2}}{s}$$

A.5 K5: Superior laryngeal movement with respect to vertebrae

$$K5 = \frac{\cos\left(\theta^{(j)}_{L6,L3,L5}|L9^{(j)}L3^{(j)}|\right) - \cos\left(\theta^{(i)}_{L6,L3,L5}|L9^{(i)}L3^{(i)}|\right)}{s}$$

A.6 K6: Hyolaryngeal approximation

$$K6 = \frac{|L6^{(j)}L9^{(j)}| - |L6^{(i)}L9^{(i)}|}{s}$$

A.7 K7: Laryngeal elevation

$$K7 = \frac{|L3^{(j)}L8^{(j)}| - |L3^{(i)}L8^{(i)}|}{s}$$

A.8 K8: Pharyngeal shortening

$$K8 = \frac{|L2^{(j)}L7^{(j)}| - |L2^{(i)}L7^{(i)}|}{s}$$