AUTOMATIC SEGMENTATION AND CLASSIFICATION OF CERVICAL ACCELEROMETRY SIGNALS: TOWARDS THE INSTRUMENTAL DETECTION OF UNSAFE SWALLOWING

by

Helia Mohammadi

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

© Copyright 2017 by Helia Mohammadi
Abstract

Automatic Segmentation and Classification of Cervical Accelerometry Signals: Towards the Instrumental Detection of Unsafe Swallowing

Helia Mohammadi
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2017

Dual-axis cervical accelerometry has been proposed as a method for non-invasive clinical evaluation of swallowing function in people with risk for dysphagia (swallowing impairment). Acceleration signals in anterior-posterior and superior-inferior anatomical directions are processed and automatically segmented for use in classification experiments. However, automatic segmentation is often too liberal, admitting both pre- and post-swallowing activity and non-swallow segments. These segmentation shortcomings adversely affect feature extraction and ultimately classification of swallowing function. We propose a kernel density estimation-based algorithm to adaptively trim the swallow segments, and energy and noise-floor algorithms to reduce the number of false positive swallow segments. Dramatic reductions (-85.4%) in false positives can be achieved with a moderate loss of true positives (-15.1%).

Furthermore, we propose a detection system for instructed and reflexive coughs, discriminating accelerometry signals associated with coughs from those representing swallows, tongue movements, and speech. Using binary genetic feature selection and a support vector machine, the proposed system achieved a cough detection accuracy of 99.26±0.12\% when discriminating between instructed cough and rest state accelerometry signals. An accuracy of 90±13.9\% was achieved using elastic net feature selection and a support vector machine when classifying reflexive coughs and rest signals. When discriminating
instructed and reflexive coughs from all other non-cough artefacts, the proposed system achieved accuracies of 90.2±3.6% and 80.3±10.5%, respectively.

Additionally, all previous cervical accelerometry studies have relied on videofluoroscopy-demarcated segments rather than the automatic localization of segments of interest, which precludes direct translation of these algorithms into a stand-alone tool. Moreover, previous studies have adopted an ipsative classification paradigm, wherein, for every unseen case, the classifier must make a forced choice (e.g. between safe and unsafe swallowing). To address these shortcomings, bolus length estimation and instance selection were introduced as enhancements to swallowing accelerometry classification, on one hand liberating classification algorithms from manual segmentation of swallows and secondly affording the classifier the freedom to abstain from a decision in the face of uncertainty. We demonstrated that together these enhancements lead to an improvement in area under the curve (83.6±5.7%) in the discrimination between safe and unsafe swallows in a sizable clinical data set.
This is dedicated to my dearests:

my mom, Fariba Vandehvar,

my dad, Ali Mohammadi, and

my twin sister, Romena Mohammadi
Acknowledgements

I would like to express my deepest and strongest gratitude to my supervisor Professor Tom Chau, whose limitless support, knowledge, encouragement, and help made this laborious journey possible. I am forever grateful that you generously welcomed me to your research center, after the unfortunate passing of my late supervisor. I learned countless lessons from you, technical and research-related, but more importantly those relating to life and leadership. You have created an exceptional research center with an unwritten rule that impels everyone to assist their peers in order to build devices or technologies that help individuals in need. I feel extremely privileged to be a part of PRISM lab and to have worked under your extraordinary supervision.

I would like to pay tribute to my late supervisor, Professor Anastasios (Tas) Venetsanopoulos. I will forever remember his unreserved support and wisdom. May he rest in peace. I would also like to thank Professor Catriona Steele for providing domain knowledge and her helpful suggestions.

Special thanks to my mentor, Professor Sadeghian, whose supporting and encouraging words has always motivated me to continue pursuing my goals. Thank you for inspiring me to begin this journey and for always standing beside me through tough times.

I would like to thank the members of my thesis committee: Professor Deepa Kundur (an inspiring role model), Professor Frank Rudzicz, and Professor Azadeh Kushki for their valuable insight, support, and advice. I would also like to thank Professor Adrian Chan for serving as my external examiner.

Thank you to the PRISM family, specially to Dr. Ali-Akbar Samadani for his generous guidance and the long brainstorming sessions, to Ka Lun Tam and Pierre Duez for their unconditional technical help and consultations, to Fanny Hotze, Marcela Correa Villada, and Leslie Mumford for their comforting conversations during the stressful times, to Siva Rajaratnam, Claire Tasker, and Sophie Wang for helping with the meeting schedules and
financial arrangements between the Department of Electrical and Computer Engineering and the Institute of Biomaterials and Biomedical Engineering, to Katherine Plewa and Amanda Fleury for their emotional and moral support during the frustrating times, and to the rest of the PRISM family for their collegiality and the constructive synergy.

I would like to extend my appreciation to my summer students, Maryam Mokhberi and Alana Esty, and to my research assistant, Kamran Masteri Farahani, for their help with data demarcations.

I would like to recognize the financial support of the NSERC (Natural Sciences and Engineering Research Council of Canada), QEII-GSST (Queen Elizabeth II Graduate Scholarships in Science and Technology), OGS (Ontario Graduate Scholarship), Ewing Rae Graduate Scholarship, Walter C. Sumner Memorial Fellowship, J.L. (Allen) Yen Scholarship, Department of Electrical and Computer Engineering at the University of Toronto, Holland Bloorview Kids Rehabilitation Hospital, and the industry partner.

Special appreciations to Mr. Hezarkhani and his family, for their encouraging words and invaluable advice. Thanks to my friends, you know who you are, for the occasional distractions and much needed moral support. You may have been annoying at points, but your humour, vigilance, and supporting words gave me comfort during the challenging times.

Above all, I would like to acknowledge the tremendous sacrifices that my parents and twin sister, Fariba Vandehvar, Ali Mohammadi, and Romena Mohammadi, have made to make this journey less stressful and challenging. I cannot thank you enough for your unconditional support, love, kindness, care, and patience. For this and much more, I am forever in your debt. It is to you that I dedicate this dissertation.
Contents

Dedication ......................................................... iv

Acknowledgements ............................................... v

List of Tables .................................................... xiii

List of Figures .................................................... xv

List of Acronyms ................................................ xvi

1 Introduction ................................................... 1

   1.1 Motivation ................................................. 1

   1.2 Research Questions and Objectives ...................... 4

   1.3 Thesis Organization ...................................... 6

2 Background .................................................... 8

   2.1 Physiology and Phases of Normal Deglutition .......... 8

       2.1.1 Oral Preparatory Phase .......................... 9

       2.1.2 Oral Transport Phase ............................. 9

       2.1.3 Pharyngeal Phase ................................. 10

       2.1.4 Esophageal Phase ................................. 10

   2.2 Dysphagia ................................................ 11

   2.3 Existing Dysphagia Detection Methodologies .......... 12

       2.3.1 Videofluoroscopic Swallowing Study .............. 12
2.3.2 Fiberoptic Endoscopic Swallowing Study . . . . . . . . . . . . . . 13
2.3.3 Cervical Auscultation . . . . . . . . . . . . . . . . . . . . . . . . . 13
2.3.4 Dual-axis Accelerometry . . . . . . . . . . . . . . . . . . . . . . . 13
2.4 Cough as an Airway Protection Mechanism . . . . . . . . . . . . . . . . 17
  2.4.1 Causes of Cough . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
  2.4.2 Airway Protection Mechanisms . . . . . . . . . . . . . . . . . . . 19
    Pharyngeal Swallow . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
    Laryngeal Cough Reflex . . . . . . . . . . . . . . . . . . . . . . . . . . 19
    Tracheobronchial Cough Reflex . . . . . . . . . . . . . . . . . . . . . . 20
2.5 Dimensionality Reduction Algorithms . . . . . . . . . . . . . . . . . . . 21
  2.5.1 Binary Genetic Algorithm (BGA) Feature Selection . . . . . . . 21
  2.5.2 Filter-based Feature Selection . . . . . . . . . . . . . . . . . . . . 21
  2.5.3 Elastic Net Feature Selection . . . . . . . . . . . . . . . . . . . . 22
  2.5.4 Principle Component Analysis (PCA) Feature Extraction . . . . 23
2.6 Classification Techniques . . . . . . . . . . . . . . . . . . . . . . . . . . 23
  2.6.1 Support Vector Machines (SVM) . . . . . . . . . . . . . . . . . . . 23
  2.6.2 Linear Discriminant Analysis (LDA) . . . . . . . . . . . . . . . . . 24
  2.6.3 Artificial Neural Networks (ANN) . . . . . . . . . . . . . . . . . . 24

3 Post-Segmentation Swallowing Accelerometry Signal Trimming and
False Positive Reduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
  3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
  3.2 Swallowing accelerometry signals . . . . . . . . . . . . . . . . . . . . . 27
  3.3 Preprocessing and Swallow Segmentation . . . . . . . . . . . . . . . . . 28
  3.4 Adaptive Swallow Trimming . . . . . . . . . . . . . . . . . . . . . . . . 28
  3.5 False Positive Reduction . . . . . . . . . . . . . . . . . . . . . . . . . . 32
    3.5.1 Performance Metrics . . . . . . . . . . . . . . . . . . . . . . . . . 32
    3.5.2 Energy-based False Positive Reduction . . . . . . . . . . . . . 34
4 Automatic Discrimination Between Cough and Non-cough Accelerometry Signal Artefacts

4.1 Introduction

4.1.1 Existing Cough Monitors

4.1.2 Cervical Accelerometry

4.2 Experimental Setup

4.2.1 Data Sets

Instructed Cough Data Set

Reflexive Cough Data Set

4.3 Methodology

4.3.1 Pre-Processing

4.3.2 Meta-Feature-Based Representation of Signals

4.3.3 Salient Feature Identification

Binary Genetic Algorithm (BGA)

Filter-Based Feature Selection

Elastic Net

Principal Component Analysis

4.3.4 Classification

Artificial Neural Network

Support Vector Machines

4.3.5 Validation

4.4 Results
4.4.1 Cough versus Rest ........................................... 51
4.4.2 Cough versus Non-Cough Artefacts ......................... 51
    Instructed Coughs ............................................. 51
    Reflexive Coughs ............................................. 52
4.5 Discussion ..................................................... 53
    4.5.1 Cough versus Rest Classification ......................... 53
    4.5.2 Optimal Feature Dimensionality ......................... 56
    4.5.3 Salient Features ......................................... 57
    4.5.4 Cough Acceleration Trajectories ......................... 58
    4.5.5 SVM versus ANN Classifiers ............................. 58
4.6 Conclusion .................................................... 59
4.7 Acknowledgement .............................................. 59

5 Classification of Dual-Axis Swallowing Accelerometry Signals with Bolus Length Estimation and Instance Selection ............... 60
    5.1 Introduction ............................................... 60
    5.2 Automatic Classification of Swallowing Accelerometry Signals .......... 61
    5.3 Instance Selection ......................................... 64
    5.4 Data Set .................................................. 66
    5.5 Methodology ............................................... 67
        5.5.1 Pre-processing and Swallow Segmentation ............... 67
        5.5.2 Feature Selection and Extraction ..................... 68
        5.5.3 Noise-floor Bolus Length Estimation .................. 69
        5.5.4 Instance Selection .................................... 71
            Multidimensional Feature-based Interquartile-range .......... 71
            Classification Probability Threshold Band ................. 72
        5.5.5 Classification and Evaluation ........................... 74
    5.6 Results ................................................... 74
5.7 Discussion .................................................. 76
  5.7.1 Reduced Dependence on VFSS-based determination of bolus of interest ......................... 76
  5.7.2 Value of Instance Selection .................................. 78
  5.7.3 Exploration of Removed Cases ................................ 78
  5.7.4 Classification Performance .................................. 81

5.8 Conclusion .................................................. 81

5.9 Acknowledgement ............................................ 81

6 Conclusion .................................................... 82
  6.1 Overall Framework ........................................... 82
  6.2 Summary of Contributions ..................................... 82
  6.3 Future Work ................................................ 84
  6.4 Publications ................................................ 85
    6.4.1 Journal Articles .......................................... 85
    6.4.2 Conference Presentations ................................ 86
    6.4.3 Patents ................................................ 87

A Accelerometry Signal Feature Calculations ......................... 88
  A.1 Time Domain Features ........................................ 89
    A.1.1 Mean ................................................... 89
    A.1.2 Mean Absolute Deviation ................................ 89
    A.1.3 Median .................................................. 89
    A.1.4 Variance ................................................ 89
    A.1.5 Skewness ............................................... 90
    A.1.6 Kurtosis ............................................... 90
    A.1.7 Memory ............................................... 90
    A.1.8 Maximum Energy ....................................... 91
List of Tables

2.1 8-point Penetration-Aspiration scale [1] ................................. 12
3.1 Estimation of $\Lambda^{AP}$ and $\Lambda^{SI}$ scalars ............................ 34
4.1 Accelerometry signal labelling ............................................. 46
5.1 Automatic classification of swallowing accelerometry signals .......... 63
5.2 Number of participants and boluses (thin consistency) ............... 67
5.3 Comparison of the classification performance using the proposed instance
    selection approaches ................................................... 76
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Thesis roadmap</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Anatomy of the deglutition apparatus</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Four deglutition phases</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>Location and orientation of the sensor</td>
<td>14</td>
</tr>
<tr>
<td>3.1</td>
<td>Swallow trimming based on dual-directional energy differences.</td>
<td>30</td>
</tr>
<tr>
<td>3.2</td>
<td>Definitions of TP, FP, and FN segments</td>
<td>33</td>
</tr>
<tr>
<td>3.3</td>
<td>MaxED FP and TP changes</td>
<td>36</td>
</tr>
<tr>
<td>3.4</td>
<td>Noise-floor FP Reduction</td>
<td>37</td>
</tr>
<tr>
<td>4.1</td>
<td>Dual-axis accelerometer</td>
<td>43</td>
</tr>
<tr>
<td>4.2</td>
<td>AP and SI signals of swallows an reflexive cough</td>
<td>44</td>
</tr>
<tr>
<td>4.3</td>
<td>Instructed cough vs. non-cough artefact accuracy comparisons</td>
<td>52</td>
</tr>
<tr>
<td>4.4</td>
<td>Instructed cough vs. non-cough artefact - Wilcoxon ranksum</td>
<td>53</td>
</tr>
<tr>
<td>4.5</td>
<td>Acceleration trajectories for instructed and reflexive coughs</td>
<td>54</td>
</tr>
<tr>
<td>4.6</td>
<td>Reflexive cough vs. non-cough artefact accuracy comparisons</td>
<td>55</td>
</tr>
<tr>
<td>4.7</td>
<td>Reflexive cough vs. non-cough artefact - Wilcoxon ranksum</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Location and orientation of the sensor</td>
<td>62</td>
</tr>
<tr>
<td>5.2</td>
<td>Noise-floor annotated AP and SI signals of a bolus</td>
<td>70</td>
</tr>
<tr>
<td>5.3</td>
<td>Scalar analysis over different object function penalty values ($\beta$)</td>
<td>72</td>
</tr>
<tr>
<td>5.4</td>
<td>Classification probability threshold band</td>
<td>73</td>
</tr>
<tr>
<td>5.5</td>
<td>Histogram of VFSS-determined and algorithmically estimated bolus lengths for different scalars ($\alpha$)</td>
<td>75</td>
</tr>
</tbody>
</table>
5.6 Box plot of AUC values for classification with instance selection by posterior probability bands .................................. 77
5.7 PCA components of selected (red) and non-selected (black) instances .................................................. 79
5.8 Parallel features of selected and non-selected instances ................................................................. 80
6.1 Overall framework .......................................................................................................................... 83
List of Acronyms

**ACC** Accuracy  
**ANN** Artificial Neural Network  
**AP** anterior-posterior  
**FEES** Fiberoptic Endoscopic Swallowing Study  
**FFT** Fast Fourier Transform  
**FN** False Negative  
**FNR** False Negative Ratio  
**FP** False Positive  
**FPR** False Positive Ratio  
**LDA** Linear discriminant analysis  
**MMG** mechanomyography  
**PCA** principal component analysis  
**PPV** Positive Predictive Value  
**RBF** radial basis function  
**SA** swallowing apnea  
**SI** superior-inferior  
**SLP** speech-language pathologist  
**SVM** Support Vector Machine  
**TN** True Negative  
**TNR** True Negative Ratio
**TP**  True Positive

**TPR**  True Positive Ratio

**USPTO**  United States Patent and Trademark Office

**VFSS**  videofluoroscopic swallowing study
Chapter 1

Introduction

1.1 Motivation

Dysphagia (swallowing impairment) refers to any swallowing disorder including any difficulties in the process of transferring food or liquid from the mouth to the stomach [2]. Through an increased risk for aspiration (the entry of material into the airway below the true vocal folds), dysphagia negatively affects both the health and the quality of life of patients.

In a 2012 national health survey, an estimated $9.44 \pm 0.33$ million adults (prevalence estimate ranges from 11% to 68% [3, 4]) in the US reported a swallowing problem [5]. The health care cost associated with dysphagia is estimated at $547$ million annually in the United States [3, 4].

Dysphagia is common in individuals who have suffered stroke [6, 7, 8]. Due to various reported medical conditions associated with dysphagia and different diagnostic procedures for the identification of swallowing impairment, the exact epidemiological numbers associated with the various neurological diseases are poorly defined. [5]. Using cursory screening procedures, the incidence of dysphagia among the stroke population was reported to be as low as 37% [7]. When dysphagia was identified using instrumental clinical
testing, this number rose to 78% [7]. Dysphagia is also common in neurological conditions such as Alzheimer’s, Parkinson’s, cerebral palsy, and acquired brain injury [5]. Stroke and neurological conditions together are the two primary etiologies of dysphagia with 700,000 cases (18.4%) of the reported swallowing problems in the US [5]. Changes in the physiological apparatus due to cancer, trauma, surgery or mucosal inflammation can also lead to dysphagia [5, 9]. Additionally, congenital abnormalities in the oral cavity, nasopharynx, hypopharynx, larynx, trachea, or esophagus can cause dysphagia [10].

The videofluoroscopic swallowing study (VFSS) is widely considered the gold standard instrumental method for the detection and diagnosis of dysphagia [11]. Nevertheless, due to the exposure to ionizing radiation, the limited availability of radiology suites and long waiting lists, it is not suitable for ongoing day-to-day monitoring [12]. As the existing clinical protocols for dysphagia detection have poor specificity and accuracy [13] there is a tendency to miss silent aspirations (due to a lack of clinical signs of aspiration) and this may lead to failure to implement necessary diet texture restrictions or treatment. There is also a risk of over-identifying clinical signs, leading to unnecessary diet and texture restrictions among other interventions caused due to misdiagnosis of dysphagia. In addition, due to the subjective rating of VFSS, the inter- and intra-rater reliability of the assessment is reported as inconsistent, inaccurate, and highly variable [3, 14, 15, 16, 17, 18].

Currently, there are no non-invasive devices available for point-of-care aspiration detection. As a non-invasive adjunct to clinical (bedside) swallowing assessment, swallowing accelerometry has been proposed where a dual-axis accelerometer is placed on the patient’s neck to measure epidermal vibrations in two directions (superior-inferior (SI) and anterior-posterior (AP)). However, previous monitoring and classification algorithms, proposed in literature, are highly dependent on the VFSS annotations during the data collection, pre-processing, and feature extraction phases. Consequently, the existing systems rely on the VFSS specifically for the identification of swallowing onset and offset.
In addition, subsequent to signal pre-processing, signal segmentation attempts to identify the swallowing activities within both channels of the acceleration signal. However, the previous non-adaptive segmentation algorithms have developed algorithms that are prone to isolate non-swallowing activities and often lead to the introduction of false positive segments [19, 20, 21].

Swallowing is a complex mechanism that involves transferring food or liquid from the mouth to the esophagus through the throat or pharynx. The pharynx is a pathway to both food, when swallowing, and air, when breathing. Therefore, an inherent risk of potential aspiration is expected when swallowing. A cough is a protective mechanical response [22] consisting of rapid contractions of the thoracic cavity, which generates a forceful and rapid expulsion of air that clears the airway of foreign material, fluid or mucus [23]. Hence, knowledge of cough severity, including intensity and frequency, may inform clinical decision-making in terms of appropriate treatment of the underlying issue. However, clinical assessments of cough often involve subjective judgement of symptoms and symptom severity, leading to inconsistent symptom reports between patients and caregivers. Cough scores, diaries, symptom questionnaires and visual analogue scales generally lack validation as tools for evaluating cough severity [24].

The overall objective of this thesis was to improve the reliability and performance of the screening and detection of dysphagia using dual-axis cervical epidermal accelerometry signal and to evaluate the proposed framework on a larger data set. The proposed framework comprises an algorithm to reduce the dependency of the acquired signals on the corresponding VFSS annotations (marked by speech-language pathologists). Another objective of this research was to reduce the number of false positive segments, which are erroneously detected using the existing automatic swallowing segmentation algorithms and to design and evaluate a swallow trimming algorithm to remove pre- and -post swallowing activities from the automatically segmented swallows. Additionally, identifying
the uncertain data points for the purpose of instance selection was another primary goal of this research. Instance selection aims to increase the overall performance of the dysphagia classification system through identifying the uncertain-to-classify signals. Such machine capability would allow a system to ask caregivers and clinicians to repeat the protocol as required. Moreover, the performance of a proposed cough detection module was evaluated in the discrimination between cough segments and non-cough accelerometry signals. Taken as a whole, this thesis contributes to improving the use of dual-axis cervical accelerometry signals in the identification and classification of safe and unsafe swallowing events. This research proposes a framework as a means for early screening of dysphagia. Individuals identified at being at heightened risk for swallowing disorders would then be referred to a swallowing clinic for follow-up assessment, which may include VFSS.

1.2 Research Questions and Objectives

This thesis aims to answer the following research questions with focus on making contributions to the corresponding objectives:

RQ1 Can the automatic segmentation of dual-axis swallowing accelerometry signals improve by identifying and removing false positive swallowing segments and trimming the pre- and post-swallowing activity signals?

O1 Develop an algorithm that reduces the number of false positive automatically detected swallows within dual-axis accelerometry recordings spanning bolus (a mass of food particles processed mechanically and chemically in the oral cavity ready to be swallowed) entry into and bolus clearance from the pharynx, with a high degree of agreement with clinical segmentation via VFSS.

O2 Develop an algorithm that trims the automatically segmented swallowing activities to remove pre- and post-swallowing signals.
RQ2 With what accuracies can dual-axis instructed and reflexive cough accelerometry signals be discriminated from rest state and non-cough artefact?

O3 Develop a robust algorithm for the discrimination of cough-like artefacts based on both instructed and reflexive cough reflex mechanisms.

RQ3 Can the dependency of the bolus onset and offset timestamps on VFSS annotation be reduced to obtain a more independent feature vector to increase the reliability of the swallowing safety discrimination system?

O4 Reduce the dependencies of the bolus onset and offset timestamps on VFSS annotations, by developing an algorithm to estimate bolus lengths after including a cushion at both ends of the bolus signals, to introduce more liberal onset and offset timestamps before estimating the bolus lengths with a high degree of agreement with clinical segmentation via VFSS.

RQ4 What level of discrimination between safe and unsafe swallowing of boluses of thin liquid consistency is achievable when considering dual-axes accelerometric signals recorded from the time of bolus entry into the pharynx through to bolus clearance?

O5 Identify and remove uncertain boluses to increase the overall performance of the discriminative system, providing the means to the clinicians and pathologists to repeat the screening test after encountering such uncertain boluses.

O6 Design and evaluate a discriminative system that leverages both quantitative bolus and automatic swallow segmented features to provide high sensitivity and specificity discrimination between safe and unsafe swallowing in patients living with dysphagia.

1.3 Thesis Organization

Following this introductory chapter, Chapter 2 presents a brief review of the background literature, mainly on the physiology and phases of deglutition, dysphagia, the existing
dysphagia monitoring and detection methodologies, dual-axis accelerometry, and cough and protection mechanisms. This chapter also includes a brief review of the dimensionality reduction and classification algorithms used in the proposed framework.

To address the aforementioned research questions and objectives, four research studies were completed during the course of this thesis. These studies are presented in Chapters 3, 4, and 5.

Chapter 3 presents the results of the first research study. This chapter addresses objectives one and two of this thesis which are to reduce the number of false positive swallowing segments, derived from an automatic segmentation algorithm and to remove pre- and post-swallowing activities through a trimming algorithm.

The second research study focuses on the third objective. This objective is addressed in Chapter 4. This chapter focuses on discriminating between cough and non-cough swallowing accelerometry artefacts, both on instructed and reflexive coughing reflex.

The third and fourth research questions are detailed in Chapter 5 which addresses research objectives four to six. In this chapter the dependency reduction of the bolus onset/offset timestamps to the VFSS annotations is explored and evaluated using an algorithm that estimates bolus lengths. Moreover, an evaluation of classification performance of the dysphagia monitoring and detection system is presented in this chapter.

Finally, Chapter 6 presents the overall framework of the proposed system, summarizes the contributions of this research, and suggests areas for future work.

Each of Chapters 3 to 5 are verbatim excerpts of a journal article that is either currently in print (Chapter 3), under review (Chapter 4), or submitted to the industry partner for approval (Chapter 5).

The technical content of Chapters 3, and 4 constitute separate patent filings.

Necessary permissions for reprinting the copyrighted published work have been obtained. The reader is advised that the introduction and parts of the methodology sections of each chapter may contain redundant information. The reader may want to skip these
redundant sections.

Figure 1.1: Thesis roadmap.
Chapter 2

Background

2.1 Physiology and Phases of Normal Deglutition

Deglutition or the act of swallowing refers to the transfer of food, water, saliva, and in general, foreign materials to the stomach from the oral cavity. Swallowing is an essential function that provides one’s body with the nutrients required to remain healthy and alive. However, nutrients, and thus energy, can only be absorbed and utilized if they are transferred successfully to the stomach. The process of deglutition contains four phases (oral preparatory phase, oral transport phase, pharyngeal phase, and esophageal phase) [2], all of which need to function properly and in a coordinated fashion to perform a healthy transport of nutrients. For these phases to be executed successfully, four anatomical regions are involved: (1) oral cavity, (2) pharynx, (3) larynx, and (4) esophagus. Figure 2.1 illustrates the anatomy of the swallowing apparatus.

It has been shown that the swallowing mechanism is affected by the bolus volume and the individual’s age [25]. Swallowing is an automatic function. It occurs approximately once every minute depending on whether the individual is awake or asleep.

The typical phases of deglutition are briefly explained in the following sections:
2.1.1 Oral Preparatory Phase

This phase includes the placement of the liquid or food in the mouth and the preparation of the bolus. The oral preparatory phase involves two stages [26, 27]: (1) Transport; and (2) Processing. In the transport stage, food is transported from the incisal to the molar area of the oral cavity for chewing. During the processing stage, the food is mechanically broken down to smaller particles using the teeth, saliva, and tongue to form a swallowable bolus. At the end of this phase, the prepared food is carried backwards into the pharynx on the surface of the tongue and liquids are squeezed backwards by the tongue [25, 26, 28].

2.1.2 Oral Transport Phase

During the oral transport phase, the prepared bolus is directed to an area where the next phase of deglutition is triggered. The force used by the tongue to propel the bolus depends on the bolus’ viscosity [29]. In this phase, the oral cavity is sealed in order to
prevent leakage of the bolus either from the lips or into the pharynx.

2.1.3 Pharyngeal Phase

This phase is considered reflexive and involuntary. The presence of liquid, food, or saliva is necessary to trigger this phase [2]. The sensory receptors in the oropharynx are triggered in this phase as the bolus is propelled by the tongue. The stimulation of these sensory receptors leads to information being sent along afferent pathways to the central pattern generator in the medulla of the brainstem where the motor command for the pharyngeal phase of the swallow is generated [30, 31]. The pharyngeal phase typically initiates when the bolus reaches the area between the anterior faucial arches and the epiglottis [2]. The velum lifts to seal the entrance to the nasopharynx, the hyoid bone and larynx move in an upward and forward direction in order to position the entrance to the airway out of the pathway of the bolus. The epiglottis retroverts and closes over the airway entrance. Biomechanical traction associated with hyolaryngeal excursion assists with opening the upper esophageal sphincter [30, 32]. The pharyngeal constrictor muscles contract behind the tail of the bolus, and make contact with the tongue base to sweep any residual material down through the pharynx towards the esophagus [33, 34].

2.1.4 Esophageal Phase

The esophageal phase is completely involuntary and involves transfer of the bolus from the pharynx to the stomach by peristalsis (the muscular movement of the inner wall of the esophagus) and gravity. The bolus travels through the esophagus at a velocity of 2–4 cm/sec [35]. Figure 2.2 illustrates the four phases of deglutition.
2.2 Dysphagia

Any difficulty in the processes of transferring food or liquid from the mouth to the stomach is referred to as dysphagia [2]. There are two primary functional concerns involved in dysphagia: a) impaired swallowing safety (in which material enters the airway, known as penetration-aspiration); and b) impaired swallowing efficiency (in which residual material remains in the pharynx after the swallow).

The severity of airway invasion events is typically graded using the 8-point Penetration-Aspiration scale [1] (shown in Table 2.1). A swallow is considered to be normal when the material either does not enter the airway or enters the supraglottic area at a point above
Table 2.1: 8-point Penetration-Aspiration scale [1]

<table>
<thead>
<tr>
<th>Scale</th>
<th>Description</th>
<th>Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Material does not enter airway.</td>
<td>Normal</td>
</tr>
<tr>
<td>2</td>
<td>Material enters the airway, remains above the vocal folds, and is ejected from the airway.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Material enters the airway, remains above the vocal folds, and is not ejected from the airway.</td>
<td>Abnormal</td>
</tr>
<tr>
<td>4</td>
<td>Material enters the airway, comes into contact with the vocal folds, and is ejected from the airway.</td>
<td>(Penetration)</td>
</tr>
<tr>
<td>5</td>
<td>Material enters the airway, contacts the vocal folds, and is not ejected from the airway.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Material enters the airway, passes below the vocal folds, and is ejected into the larynx or out of the airway.</td>
<td>Abnormal</td>
</tr>
<tr>
<td>7</td>
<td>Material enters the airway, passes below the vocal folds, and is not ejected from the trachea despite effort.</td>
<td>(Aspiration)</td>
</tr>
<tr>
<td>8</td>
<td>Material enters the airway, passes below the vocal folds, and no effort is made to eject.</td>
<td></td>
</tr>
</tbody>
</table>

the vocal folds, and then is successfully ejected. Penetration, on the other hand, occurs when the foreign material (bolus) either enters the supraglottic space and is not ejected, or makes contact with the true vocal folds. Aspiration is the entry of foreign material into the airway below the true vocal folds.

When an individual experiences aspiration, a reflexive cough is a normal reaction. However, studies show that among the acute stroke population, 2% to 25% of aspirations occur “silently” with no noticeable physiological signs [36].

### 2.3 Existing Dysphagia Detection Methodologies

#### 2.3.1 Videofluoroscopic Swallowing Study

Currently, the videofluoroscopic swallowing study (VFSS), commonly known as a modified barium swallow (MBS), is considered the state-of-the-art standard and reference instrumental test for detection of dysphagia [3, 37, 38, 39]. During this procedure patients are asked to swallow different consistencies of barium-infused or barium-coated
liquid and food [11, 40]. The severity of penetration-aspiration can be directly viewed. The patient must ingest barium and is exposed to ionizing radiation [11]. In addition, the expensive equipment and experienced clinicians to acquire VFSS data are not available in every medical establishment [12, 41], thus precluding the use of VFSS on a day-to-day basis [42].

2.3.2 Fiberoptic Endoscopic Swallowing Study

Another procedure used for the detection of dysphagia is Fiberoptic Endoscopic Swallowing Study (FEES). During this procedure, a flexible tube is inserted into the patient's throat and the individual is asked to ingest dye-colored food or liquid with different consistencies. This procedure is invasive and uncomfortable [43, 44]. However, the advantage over VFSS is that the FEES equipment is that there is no radiation exposure.

2.3.3 Cervical Auscultation

Cervical auscultation is another technique that is sometimes included to collect additional information during clinical bedside swallowing assessments. This procedure is non-invasive and it produces an assessment based on swallow sounds, recorded by a stethoscope, accelerometer, or laryngeal microphone, and it can, in some cases, record pre- and post-swallow sounds [45, 46]. One of the limitations of this method is in its inability to precisely pinpoint the source of the sounds recorded [45].

2.3.4 Dual-axis Accelerometry

Currently there are no non-invasive and reliable devices available for point-of-care aspiration detection. As a non-invasive potential adjunct to clinical (bedside) swallowing assessment, swallowing accelerometry has been investigated. An accelerometer is placed on the surface of the patient’s neck on and slightly below the laryngeal prominence (com-
monly known as Adam’s apple) to record epidermal vibrations (i.e. signals generated from physiological vibrations of the swallowing apparatus) [47, 48] along two anatomical axes: anterior-posterior (AP) and superior-inferior (SI). Figure 2.3 demonstrates the location and orientation of the sensor on a patient’s neck.

![Figure 2.3: The location and orientation of the sensor on a participant’s neck.](image)

Quantitative evaluations of the recorded signals obtained from the accelerometer and discrimination between safe and unsafe swallows may be possible through digital signal processing, pattern recognition, and machine learning techniques. Various studies have been conducted on swallowing tools during the last two decades [13, 40, 49, 50, 51, 52]. The majority of these studies have focused on swallowing acoustics, acquired using microphones. However, this method is sensitive to ambient noise. As an alternative approach, cervical epidermal or laryngeal acceleration studies have been proposed [53, 54, 55, 56].

Several research studies identified a similar swallowing pattern for healthy swallowing accelerometry signals and discuss an absence of such pattern among the unhealthy instances [40, 48, 57, 58, 59, 60, 61, 62]. It has been suggested that the hyoid, larynx, and hyolaryngeal complex movements contribute to the accelerometry signal vibrations.
These studies investigated the movements of these regions using the VFSS recordings and their correlation to the timestamps of the swallowing accelerometry signal spikes and patterns. Specifically, correlation between the magnitude and exact timing of the bolus movement along the swallowing accelerometry signals and the extent of the laryngeal rise was reported [63].

Several past studies have been performed using a single-axis accelerometer in the AP direction [40, 60, 61, 62, 64, 65, 66], ignoring the potential information from the SI direction. Research has shown that the maximum hyoid displacement in swallowing activities occur in both AP and SI axes [67, 68]. The study of the accelerations in the SI axis provides additional information when detecting swallowing difficulties [19, 42]. In addition to the studies reporting a correlation between the AP axis of the swallowing accelerometry signal and the extent of the laryngeal elevation [60], the maximum excursion of the hyolaryngeal complex is reported to exhibit similar movements to the anterior and superior directions during swallowing [63, 67, 68, 69]. This suggests that both axes of the acceleration signal carry non-redundant information regarding hyolaryngeal displacement [69]. In addition, safe and unsafe swallow discrimination accuracies are increased using both AP and SI axis [53].

Swallowing accelerometry segmentation has been investigated in the literature [19, 20, 70]. A multi-sensor based study used Artificial Neural Network (ANN) to segment swallowing activities within dual-axis accelerometry, submental mechanomyography (MMG), and nasal airflow signals [20]. In this study, the AP direction is reported as the leading contributor in the accuracy of swallowing signal segmentation followed by the SI channel [20]. Another study uses a two-class piecewise sequential fuzzy c-mean partitioning to segment dual-axis swallowing accelerometry signals [19]. It is shown that the removal of the vocalization segments of accelerometry signals increases the segmentation accuracy [70]. The aforementioned last two studies reported over 90% segmentation accuracy. However, the algorithm evaluation is conducted using estimated swallow timestamps
rather than a more reliable validation reference such as the VFSS.

Effects of the head movement on dual-axis cervical accelerometry signals were reported on healthy individuals [71]. This study was conducted using accelerometry signals without swallowing activities in the presence of head movement. The peak frequency is not affected by the head motion tasks, however different bandwidths and spectral centroids between the two AP and SI directions were reported, suggesting the presence of an additional non-dominant low frequency component associated with head movement [71, 72].

Classification of healthy and unhealthy swallowing is investigated in adults based on the combination of dual-axis accelerometry and nasal airflow signals [42]. In another study, a radial basis classifier is presented to discriminate between healthy and unhealthy swallowing accelerometry signals in children [40]. A research study reports Support Vector Machine (SVM) classification with the radial basis function (RBF) kernel and principal component analysis (PCA) dimensionality reduction algorithm on pediatric swallowing accelerometry signals [55]. In addition, a reputation-based classification algorithm is proposed to distinguish between safe and unsafe swallowing accelerometry signals [53]. Another study discriminates healthy and unhealthy swallowing accelerometry signals using a Bayes classifier based on wavelet packet analysis and Linear discriminant analysis (LDA) as the feature reduction algorithm [56].

The majority of the conducted research reports on a small sample size, usually acquired at a single site. Additionally, the data points are collected from patients of the same target population such as stroke survivors [73], children [40], or those with a specific disability [13]; therefore, limiting the screening/detection tool to a specific population. According to recent systematic reviews [13, 74, 75], the swallowing screening systems are recommended to be further evaluated and verified with larger data sets consisting of patients from different populations and medical conditions.

The studies on discriminative analysis of dysphagia vary in sample size, the uti-
lized screening tool, discriminatory problem at hand (aspiration, penetration, or pharyngeal residue), characteristic of the study sample, and validation techniques (k-fold cross-validation, leave-one-out, leave-one-subject-out). Hence a fair comparison between these research studies is challenging. A comprehensive review of detection of and screening for dysphagia can be found in [13].

One of the shortcomings of the existing swallowing accelerometry discrimination algorithms is that the onset and offsets of the acquired acceleration signals are annotated based on the demarcations of the VFSS, which are recorded concurrent to the acceleration signal. The inevitable VFSS annotation-dependency during the accelerometry signal demarcation phase (bolus onset and offset timestamps annotations) affects the consequent segmentation, feature calculation, and classification phases, thus precludes the stand-alone use of the dysphagia screening system.

2.4 Cough as an Airway Protection Mechanism

Cough can be symptomatic of various respiratory conditions such as asthma, rhinitis and gastrooesophageal reflux disease in adults and protracted bronchitis in children [76]. Cough is also a normal reflexive response to aspiration, the entry of foreign material into the airway, seen in people with swallowing difficulties [77].

2.4.1 Causes of Cough

A cough may be due to different causes including the normal physiological reflex. The main causes of cough include:

- **Foreign body aspiration**: Entry of a foreign body into the airway can be life-threatening and individuals with swallowing reflex dysfunction are at greater risk for aspiration [78]. When the foreign body is subtle and small enough, it can remain undetected for long periods of time [78, 79, 80]. As an unusual example,
a 46-year-old white male was reported with a case of foreign body aspiration that had likely occurred 3 years before his diagnosis [78]. Foreign body aspiration may also be misdiagnosed as bronchitis, chronic pneumonia, asthma, or tumour [78]. Cough due to foreign body aspiration may be suspected when coughing commences immediately after the patient begins to eat [79].

- **Infections:** A cough can be due to respiratory tract infection such as cold, acute bronchitis, tuberculosis, pertussis, or pneumonia [81].

- **Reactive airway disease:** Individuals who meet the criteria for asthma suffer from two conditions including atopic cough and eosinophilic bronchitis. Asthma is a common cause of chronic cough in both children and adults. Chronic bronchitis, on the other hand, is referred to persistent cough with mucus and sputum.

- **Gastroesophageal reflux:** This type of cough occurs when acidic contents of the stomach come rise into the esophagus. In patients with unexplained source of cough, gastroesophageal reflux is considered [82].

- **Psychogenic cough:** When a physical disease is absent, a psychogenic cough, habit cough, or tic cough may be the cause. Psychogenic cough is more common in children compared to the adults and may be developed in a child with a chronically ill sibling [83].

- **Neurogenic cough:** Sensory neuropathic disorder may cause chronic cough [84]. In tic disorders (such as Tourette syndrome), coughing may occur. This type of cough should be distinguished from throat-clearing in this type of disorder.

- **Air pollution:** Poor air quality and air pollution including tobacco smoke, irritant gases, dampness in a closed environment, and particulate matter may cause coughing [82].

- **Other causes:** Cough can also be cause by lung diseases such as cystic fibrosis, bronchiectasis, or lung tumors. Additionally, cardiovascular diseases such as heart failure and aortic aneurysm may also cause coughing [85].
2.4.2 Airway Protection Mechanisms

Due to the risk of potential aspiration during deglutition, three airway protection mechanisms are available: (1) the pharyngeal swallow, (2) the laryngeal cough reflex, and (3) the tracheobronchial cough reflex.

Pharyngeal Swallow

The pharyngeal swallow is a complex reflex mechanism [86] that achieves both bolus transfer and airway protection [87]. Although introducing potential risks of aspiration by transferring food or liquid to the vicinity of the airway, the pharyngeal phase of swallowing is considered as the best defence mechanism against aspiration. Pharyngeal swallow contains two mechanisms: (1) laryngeal closure, and (2) hyolaryngeal complex movement [2].

Laryngeal Cough Reflex

Laryngeal coughing, also known as the expiration reflex [88, 89, 90], provides secondary airway protection when the pharyngeal swallowing mechanism fails to protect against aspiration. Through the forceful expulsion of air, the laryngeal cough reflex ejects foreign material from the airway. This reflex is triggered when the laryngeal receptors, specifically receptors of the recurrent laryngeal nerve, are stimulated as a result of foreign material entry or aspiration.

The laryngeal cough reflex contains three phases [91]: (1) compressive, (2) expulsive, and (3) constrictive. The first phase includes glottis closure for a brief moment while the expiratory muscles contract. This phase creates the necessary pressure. In the second phase, the laryngeal adductor muscles relax abruptly to open the glottis to exhale the pressurized air. Finally, the last phase includes relaxation of the expiratory muscles followed by contraction of the vocal cord adductors to create laryngeal resistance.

The laryngeal coughing reflex includes minimum inhalation during the first phase [91]
to prevent foreign material from travelling further down the airway [92].

**Tracheobronchial Cough Reflex**

Another type of cough reflex is the tracheobronchial cough. The laryngeal and tracheobronchial cough reflex mechanisms are often indistinctive in the literature; however they seem to have different neuronal pathways and incorporate distinct reflex mechanisms [93].

The tracheobronchial cough reflex includes three phases [88]: (1) inspiratory, (2) compressive, and (3) expiratory. The compressive and expiratory phases of the tracheobronchial cough are similar to the compressive and expulsive phases of the laryngeal cough reflex, respectively. One of the important distinctions between the laryngeal and tracheobronchial cough reflex mechanism is that the tracheobronchial cough includes an inspiratory phase prior to the compressive and expiratory phases.

The first phase of the tracheobronchial includes inhalation of a variable volume of air with expiratory muscles stretched. The next phase is similar to the first phase of the laryngeal cough reflex as the glottis is closed briefly while the expiratory muscles contract to build the necessary pressure. During the last phase, the glottis is opened to exhale the pressured air.

Deeper receptors are involved in the tracheobronchial reflex compared to the laryngeal cough. Additionally, this type of coughing reflex is triggered volitionally either voluntarily induced or generated using the cortical control [94].

Therefore, since the laryngeal cough reflex cannot be initiated volitionally, unlike the tracheobronchial cough reflex, and considering the physiological differences between the two cough reflexes, it may be concluded that instructed coughing differs from the reflexive coughing. Research has suggested that the involuntary or reflex cough mechanism differs from instructed coughs [95, 96]. This is further explored in chapter 4.
2.5 Dimensionality Reduction Algorithms

The following dimensionality reduction algorithms are used to achieve parsimonious and salient features.

2.5.1 Binary Genetic Algorithm (BGA) Feature Selection

A genetic algorithm (GA) [97] is a population-based, stochastic optimization algorithm originally derived from the concept of human evolution [97]. To invoke GA-based feature selection, candidate feature vectors are coded as a chromosome of boolean values, each gene indicating whether the corresponding feature is selected. In other words, a candidate solution, \( x \), to the optimization problem is a binary string of the same length as the number of features:

\[
x = \{f_1, f_2, \ldots, f_n\}, \quad f_i \in \{0, 1\}, i \in [1, n]
\]

where \( n \) is the number of features and \( f_i = 1 \) indicates whether or not the \( i^{th} \) feature is included in the reduced feature set. As a simplified example, \( x = \{0, 0, 1, 1, 0\} \) would indicate that only the 3\(^{rd}\) and 4\(^{th}\) features are included in the reduced feature set.

The following fitness function, \( G \), is then minimized:

\[
G(x_j) = \epsilon_j \frac{1}{N_j} + \exp\left(-\frac{1}{N_j}\right)
\]

where \( \epsilon_j \) is the classification error based on a k-nearest neighbour classifier, and \( N_j \) is the cardinality of selected features of the \( j^{th} \) solution candidate.

2.5.2 Filter-based Feature Selection

In this approach, features are ranked based on their uni-dimensional class separability score [98]. While inter-feature correlations were not considered, this approach offered
computational simplicity. Individual features were ranked in descending order on the basis of a feature-specific criterion value, \( C(k) \), given by

\[
C(k) = \frac{|m_1(k) - m_2(k)|}{s_p(k)\sqrt{1/n_1 + 1/n_2}},
\]

(2.2)

\[
s_p(k)^2 = \frac{(n_1 - 1)s_1(k)^2 + (n_2 - 1)s_2(k)^2}{n_1 + n_2 - 2},
\]

(2.3)

where the subscripts 1 and 2 denote respectively samples from classes 1 and 2, \( m_1(k) \) and \( m_2(k) \), and, \( s_1(k) \) and \( s_2(k) \) are the sample estimates of the population means and variances, respectively, and \( n_1 \) and \( n_2 \) are the sample sizes, while \( k = 1, 2, ..., n \) indexes the features. The criterion \( C(k) \) is the absolute value of the two-sample t-test with pooled variance estimate, \( s_p(k) \) [98]. The top ranking features were then selected as the reduced feature vector.

### 2.5.3 Elastic Net Feature Selection

The elastic net is a regularized binomial logistic regression which is used to select a subset of features [99]. It linearly combines the penalties of the LASSO (Least Absolute Shrinkage and Selection Operator) and ridge regularization methods [99]. For observations \( x_i \in \mathbb{R}^p \) and responses \( y_i \in \mathbb{R}, i = 1, 2, ..., N \), the objective function for the binomial family is given as

\[
\min_{(\beta_0, \beta)} - \left[ \frac{1}{N} \sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right] + \lambda \left[ (1 - \alpha)\|\beta\|_2^2/2 + \alpha\|\beta\|_1 \right],
\]

(2.4)

where \( N \) is the number of observations, \( x_i \) are the features and \( y_i \) the corresponding class label, at observation \( i \), \( \beta_0 \) and \( \beta \) are the intercept and predictor coefficients respectively,
$0 \leq \alpha \leq 1$ is a compromise between ridge ($\alpha = 0$) and LASSO ($\alpha = 1$) and interpolates between the $L^1$ norm and the $L^2$ norm of $\beta$ ($||\beta||_1$ and $||\beta||_2^2$). $\lambda \geq 0$ is a complexity parameter which introduces sparsity over the feature space. All features are selected when $\lambda = 0$.

### 2.5.4 Principle Component Analysis (PCA) Feature Extraction

In addition to the above feature selection approaches, PCA was used to generate a reduced set of transformed features [100]. PCA is a non-parametric, generic statistical method, used for dimensionality reduction by rejecting low-variance features. It is useful for extracting new and uncorrelated features that are more informative. PCA is one of the most commonly used feature extraction algorithms that uses an orthogonal transformation in order to convert a set of features that may be correlated into a set of components that are linearly uncorrelated [101]. The new set of components is generated such that the first component has the largest possible variance and provides the most variability of the data set. The succeeding components are calculated such that each are orthogonal to the previous component and has the highest possible variance. These components are orthogonal since they are the eigenvectors of the symmetric covariance matrix.

### 2.6 Classification Techniques

The following classification algorithms are used in the proposed framework.

#### 2.6.1 Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression challenges. SVM was initially introduced for binary classifications with primary focus on class separation using an optimal separating hyperplane by maximizing the margins between the boundary data points of the two classes. The data
points on the boundaries are commonly referred to as support vectors and the middle of the margin between the support vectors is the desired separating hyperplane \[102\]. When a linear hyperplane cannot be found, the data points are transferred to an often higher-dimension space using different kernel models, such that the new projections are linearly separable. An extensive tutorial on SVM and the different kernel techniques is given in \[103\].

### 2.6.2 Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is a classification algorithm that searches for a linear combination of variables to form a separating boundary that best discriminates between the target classes. LDA was first introduced in 1936 \[104\] and provides a simple mathematically robust model with comparable accuracy to that of the more complex methods. Using the LDA algorithm, the belonging of a new set of inputs to a particular class is predicted by estimating the probabilities with which the new set belongs to each class. These probabilities are estimated using Bayes theorem \[105\].

### 2.6.3 Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) is a classification method commonly used for solving machine learning problems. ANN is comprised of connected units (or neurons) with weighted connections. This method takes the features as its input, passes the features through different layers of neurons (input, hidden, and output layer), and generates a classification output. ANN is trained by example rather than explicitly programmed. This attribute is desired when the solution or feature detection is difficult. However, the lack of transparency of these systems is a disadvantage compared to rule-based techniques \[106\].
Chapter 3

Post-Segmentation Swallowing
Accelerometry Signal Trimming and False Positive Reduction


In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University of Toronto’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.
3.1 Introduction

Any difficulty in the process of transferring food or liquid from the mouth to the stomach is referred to as dysphagia [2] or swallowing impairment. Dysphagia negatively affects the quality of life of patients through an increased risk of aspiration (the entry of material into the airway below the true vocal folds). Causes of dysphagia include: changes in the anatomical structures necessary for swallowing as a consequence of surgery, cancer, trauma, or inflammation; genetic malformations of the swallowing apparatus; and neurological impairments due to stroke, Parkinson’s disease, cerebral palsy, and acquired brain injury [107]. According to the Agency for Healthcare Research and Quality, 41.7% of stroke patients in the US (320,476 patients) are affected by dysphagia. When considering all etiologies, the number of new cases in the US rises to 624,757 per year, while approximately 6,288,116 patients are currently living with dysphagia.

The videofluoroscopic swallowing study (VFSS) is the gold standard for the diagnosis of dysphagia [38]. During this x-ray procedure, patients are asked to swallow different consistencies of liquid and food mixed with barium. In addition to radiation exposure, VFSS requires access to expensive equipment and clinician expertise that is not available in every medical establishment [41].

Significant correlations between laryngeal movement and the anterior-posterior (AP) acceleration signal have been reported [41] and hyolaryngeal excursion has been implicated as the primary physiological source of the two-dimensional signal via semi-partial correlations [63]. Therefore, as a non-invasive adjunct to clinical swallowing assessment, swallowing accelerometry has been introduced [48], deploying a dual-axis accelerometer on the surface of the patient’s neck slightly below the laryngeal prominence (commonly known as Adam’s apple) to measure epidermal vibrations accompanying swallowing [48], in two anatomical directions: anterior-posterior (AP) and superior-inferior (SI). Dual-axis accelerometric measurements provide more information relating to swallowing than measurements in either the AP or SI axis alone [69]. Subsequent to signal pre-processing,
signal segmentation attempts to isolate the information-bearing portions of the signal. However, existing non-adaptive swallow segmentation algorithms are prone to identifying non-swallowing activities. This chapter introduces methods to mitigate this undesirable effect of segmentation.

3.2 Swallowing accelerometry signals

Dual-axes swallowing acceleration signals were acquired using a biaxial accelerometer (ADXL327) with sensitivity of $\pm 2g$ from 264 consenting adult participants. The protocol was approved by the research ethics boards of the 8 participating North American hospitals. The accelerometry signals were collected during videofluoroscopy via a two-channel custom USB audio interface, consisting of a high-pass filter with 0.1 Hz cut-off to remove the DC or the gravity component from the signals and a low-pass filter with 3 kHz cutoff for each channel. The signals from each axis were sampled at 10 kHz with 12-bit resolution. Data were stored by a custom LabVIEW program, which captured the time-synchronized videofluoroscopy recordings and accelerometry signals and stored these on a laptop computer for subsequent offline analysis. Participants were first asked to perform calibration tasks including rest, cough and counting at the beginning of the session. They were then instructed to take 6 sips of thin liquid barium (Bracco Varibar Thin Liquid Barium diluted to a 20% w/v concentration). A bolus refers to a sip of thin liquid barium which can be ingested in one or more swallows. The videofluoroscopy recordings were reviewed in duplicate by trained speech-language pathologists to identify the time-codes corresponding to the onset and offset of swallowing events within each bolus. Inter-rater agreement for event detection was excellent, with intra-class correlations for specific events ranging from 0.81 to 0.99 (95% confidence intervals from 0.79 to 0.99). A total of 1,243 usable swallowing events were identified.
3.3 Preprocessing and Swallow Segmentation

Signals were preprocessed by de-noising [108], head movement removal [71, 72], and speech removal [70]. High frequency noise was further suppressed via wavelet packet decomposition with a 4-level discrete Meyer wavelet and Shannon entropy. Shannon entropy was used to estimate the level of decomposition [108].

AP and SI variance signals were computed by estimating the sample variance within windows of size 200 data points, shifted along each of the AP and SI signals with 50% overlap. The swallows were then segmented by subjecting the variance signals to a sequential fuzzy c-means algorithm [19]. Automatic segmentation by this method as well as by neural network [20] or quadratic variation [21] tend however, to yield segment boundaries that are too lenient, admitting non-swallow activity pre- and post-swallow (Fig. 3.1-e). Likewise, segmentation is prone to identify non-swallow artefacts, resulting in false positive segments (Fig. 3.4-b). The following algorithms were designed to address both of these issues.

3.4 Adaptive Swallow Trimming

This algorithm aims to trim the swallow, segmented using the sequential fuzzy c-means algorithm [19], so that it includes only the portion of the signal corresponding to the physiological vibrations associated with swallowing while excluding the pre- and post-swallow signal fluctuations. We first calculate the base energy, $E_{\text{base}}$, within a window of size, $w = 500$ samples, centered at the location of the peak amplitude value of the
segmented swallow, $S$.

$$E_{base}^{AP} = \sum_{i=p-\frac{w}{2}}^{p+\frac{w}{2}} (z_{i}^{AP})^2 \tag{3.1}$$

where $z_{i}^{AP}$ is the pre-processed AP signal, $w$ is the window size, and $p$ is the peak index, that is the timestamp at which the value of the accelerometry signal reaches its maximum ($\arg\max_i z_{i}^{AP}$).

Let $L_S$ be the length of the initially segmented swallow. We define a corresponding non-swallow segment, $NS$, as the segment of length $L_S$ with the minimum signal energy within the first 10 seconds of the calibration signal (likely during quiet breathing), given that typical swallows are approximately 1 second in duration.

We then slide the window by an increment of $s = 50$ samples along the swallow and non-swallow segments (i.e., with 90% overlap) and calculate the energy differences between the base energy and the energy within the moving windows. For the AP signal, these differences are written as:

$$\Delta E_S^{AP}(j) = |E_{base}^{AP} - E_S^{AP}(j)|, \tag{3.2}$$

$$\Delta E_{NS}^{AP}(j) = |E_{base}^{AP} - E_{NS}^{AP}(j)|, \quad 1 \leq j \leq \left\lceil \frac{L_s - w + s}{s} \right\rceil,$$

where $\Delta E_S^{AP}$ and $\Delta E_{NS}^{AP}$ are the energy differences of the swallow and non-swallow segments, respectively, and

$$E_S^{AP}(j) = \sum_{i=1+(j-1)s}^{w+(j-1)s} (z_{i}^{AP})^2, \quad 1 \leq j \leq \left\lceil \frac{L_s - w + s}{s} \right\rceil \tag{3.3}$$
$E_{NS}^P(j)$ is similarly defined using the non-swallow segment, $NS$. Fig. 3.1-a depicts an example of these energy differences for one swallow. The same formulation applies to the SI signal.

Figure 3.1: Swallow trimming based on dual-directional energy differences.

The probability density of energy differences for both swallow and non-swallow segments are then estimated from their respective histograms (Fig. 3.1-b) using kernel density estimation [109]. Let $x_i$ denote the histogram bin counts of energy difference values $i = 1, \ldots, N$. The estimated kernel density of energy differences, $\hat{d}(x)$ is computed as in
where $K$ is the kernel function, $N$ is the number of energy difference distribution bins, and $h$ is the kernel smoothing bandwidth. Given the versatile estimation capabilities of a Gaussian mixture, we adopted a Gaussian kernel:

$$K\left(\frac{x - x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}. \quad (3.5)$$

The bandwidth of the kernel is estimated by [111]

$$h = 1.06\sigma N^{-\frac{1}{4}} \quad (3.6)$$

where $\sigma$ is the standard deviation of the energy differences.

Let $C(x) \in \{\text{swallow, non-swallow}\}$ represent the predicted label for an energy difference $x$. The probability of an energy difference, $x$, belonging to the swallow class is computed as [109]

$$P(C(x) = \text{swallow}|X = x) = \frac{p_S d_S(x)}{p_S d_S(x) + p_N d_N(x)} \quad (3.7)$$

where $d_S(x)$ and $d_N(x)$ are the estimated densities for swallow and non-swallow segments, while $p_S = p_N = 0.5$ are the swallow and non-swallow priors, respectively [109].

Fig. 3.1-c shows the probability density estimations of both swallow and non-swallow segments. Again, note that the above formulation applies to both AP and SI signals.

Integrating the densities, we obtain probability distributions for swallow and non-swallow segments. By setting a probability cutoff, we obtain energy thresholds $T^{AP}$ and $T^{SI}$ for each channel. The higher the probability cutoff, the more aggressive the
trimming. A probability cutoff 0.9 was determined to be suitable for the problem at hand, as exemplified in Fig. 3.1-d, where $T^{SI} = 1.26 \times 10^6$. The vertical line marks the energy difference where the swallowing class probability exceeds 0.9. This energy threshold is also plotted on Figs. 3.1-b and 3.1-c.

In summary, to trim the swallow segments, the location of the peak amplitude is found, overlapping windows of size $w$ are shifted to the left and to the right of the peak by increments of size $s$ and the energy difference is calculated within each window. Bilaterally, windowed segments with energy difference below the threshold are removed from the candidate swallow segment.

Fig. 3.1-e illustrates the SI signal of a swallow segment. The black dashed rectangle marks the trimming boundary achieved considering only the SI channel. In order to select the same portion of the AP and SI signals, we adopted two approaches. The first approach, called AP/SI min-max, selects the left- and right-most boundaries of the AP and SI segments (marked by the green thin rectangle). The second approach, called AP/SI average, calculates the midpoint of the two boundaries of AP and SI segments (marked by a red thick rectangle). Using the SI boundaries, the distance between the VFSS onset and the corresponding swallow segment onset were reduced by 71.7%. The AP, AP/SI min-max and AP/SI average approaches resulted in reduction of 48.1%, 20.6%, and 56.7%, respectively.

### 3.5 False Positive Reduction

#### 3.5.1 Performance Metrics

Before describing the proposed false positive reduction methods, we first outline the relevant performance metrics.

- **True Positive (TP)** refers to an automatically segmented swallow candidate where a videofluoroscopy-demarcated swallow onset (marked by a vertical lines
in Fig 3.2) precedes or falls within the candidate segment boundaries (shown as rectangles in Fig 3.2).

- **False Positive (FP)** refers to an automatically segmented swallow candidate that does not have a videofluoroscopy-demarcated swallow onset neither within nor preceding the candidate segment boundaries (Fig 3.2).

- **False Negative (FN)** occurs when no swallows are segmented for a particular videofluoroscopic swallow onset (Fig 3.2).

- **TP change** refers to the percent change in the number of TP cases

\[
TP_{\text{change}} = \frac{(TP_{\text{new}} - TP_{\text{existing}})}{TP_{\text{existing}}}.
\]

- **FP change** refers to the percent change in the number of FP segments.

- **Recall (R)** also known as sensitivity, measures the proportion of swallow segments that are correctly identified, i.e., \( R = \frac{TP}{TP + FN} \).

- **Precision (P)** or the positive predictive value is a measure of fidelity and equals one minus the FP rate, i.e., \( P = \frac{TP}{TP + FP} \).

- **Harmonic average (F)** is a combined measure of recall and precision, i.e., \( F = \frac{(2 \cdot R \cdot P)}{(P + R)} \).

- **False Discovery Rate (FDR)** measures the proportion of the detected swallow segments that are falsely identified as swallow candidates, i.e., \( FDR = \frac{FP}{TP + FP} \).

To reduce the instances of falsely identified swallows, two algorithms are proposed: (1) energy-based and (2) noise floor-based false positive reduction. Each is explained in turn.
Table 3.1: Estimation of $\Lambda^{AP}$ and $\Lambda^{SI}$ scalars

<table>
<thead>
<tr>
<th>Approach</th>
<th>$\Lambda^{AP}$</th>
<th>$\Lambda^{SI}$</th>
<th>TP change (%)</th>
<th>FP change (%)</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>$F$ (%)</th>
<th>FDR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing</td>
<td>-</td>
<td>-</td>
<td>1118</td>
<td>226</td>
<td>125</td>
<td>89.94</td>
<td>83.18</td>
<td>86.43</td>
</tr>
<tr>
<td>mean</td>
<td>0.860</td>
<td>0.838</td>
<td>704</td>
<td>9</td>
<td>539</td>
<td>-37.03</td>
<td>-96.02</td>
<td>56.64</td>
</tr>
<tr>
<td>mean-std</td>
<td>0.599</td>
<td>0.564</td>
<td>834</td>
<td>21</td>
<td>409</td>
<td>-25.4</td>
<td>-90.71</td>
<td>67.1</td>
</tr>
<tr>
<td>mean-2xstd</td>
<td>0.339</td>
<td>0.290</td>
<td>981</td>
<td>74</td>
<td>262</td>
<td>-12.25</td>
<td>-67.26</td>
<td>78.92</td>
</tr>
<tr>
<td>MaxED ($\beta$: .95)</td>
<td>0.022</td>
<td>0.078</td>
<td>1113</td>
<td>216</td>
<td>130</td>
<td>-0.45</td>
<td>-4.42</td>
<td>89.54</td>
</tr>
<tr>
<td>MaxED ($\beta$: .9-.8)</td>
<td>0.064</td>
<td>0.078</td>
<td>1113</td>
<td>206</td>
<td>130</td>
<td>-0.45</td>
<td>-8.85</td>
<td>89.54</td>
</tr>
<tr>
<td>MaxED ($\beta$: .75-.7)</td>
<td>0.079</td>
<td>0.078</td>
<td>1113</td>
<td>200</td>
<td>130</td>
<td>-0.45</td>
<td>-11.5</td>
<td>89.54</td>
</tr>
<tr>
<td>MaxED ($\beta$: .65-.45)</td>
<td>0.185</td>
<td>0.078</td>
<td>1079</td>
<td>163</td>
<td>164</td>
<td>-3.49</td>
<td>-27.88</td>
<td>86.81</td>
</tr>
<tr>
<td>MaxED ($\beta$: .4-.2)</td>
<td>0.316</td>
<td>0.078</td>
<td>1040</td>
<td>203</td>
<td>115</td>
<td>-6.98</td>
<td>-49.12</td>
<td>83.67</td>
</tr>
<tr>
<td>MaxED ($\beta$: .15)</td>
<td>0.317</td>
<td>0.235</td>
<td>1009</td>
<td>92</td>
<td>234</td>
<td>-9.75</td>
<td>-59.29</td>
<td>81.17</td>
</tr>
<tr>
<td>MaxED ($\beta$: .1)</td>
<td>0.458</td>
<td>0.326</td>
<td>942</td>
<td>54</td>
<td>301</td>
<td>-15.74</td>
<td>-76.11</td>
<td>75.78</td>
</tr>
<tr>
<td>MaxED ($\beta$: .05)</td>
<td>0.672</td>
<td>0.384</td>
<td>865</td>
<td>23</td>
<td>378</td>
<td>-22.63</td>
<td>-89.82</td>
<td>69.59</td>
</tr>
</tbody>
</table>

3.5.2 Energy-based False Positive Reduction

The first method is based on adaptive energy-based thresholding. We derive for each axis, a bolus-specific threshold, $T_b$ based on the axial energy of the bolus, $T_b^{AP} = \Lambda^{AP} \times \hat{E}_b^{AP}$, where $\hat{E}_b^{AP}$ is the maximum energy calculated within a moving window of size 500 samples on the AP channel for a specific bolus $b$, and $\Lambda^{AP}$ is a data-dependent scalar. $T_b^{SI}$ is determined using the same procedure, but for the SI channel. For each candidate swallow, the energy was estimated within 50% overlapping windows of 500 samples each. If the maximum energy value across these windows was less than the corresponding bolus-adaptive threshold (either $T_b^{AP}$ and $T_b^{SI}$), then the candidate swallow was discarded.

The scalars, $\Lambda^{AP}$ and $\Lambda^{SI}$, were estimated by the following approaches.

Energy Ratio

Segmental scalars for each channel ($\lambda_{b,i}^{AP}$ and $\lambda_{b,i}^{SI}$) were calculated for each swallow segment as

$$\lambda_{b,i}^{AP} = \frac{\hat{E}_{b,i}^{AP}}{\hat{E}_b^{AP}}, \quad \text{and} \quad \lambda_{b,i}^{SI} = \frac{\hat{E}_{b,i}^{SI}}{\hat{E}_b^{SI}}$$

(3.8)

where $i \geq 1$ indexes the number of the detected swallow segment within bolus $b$, and
\( \hat{E}_{b,i}^{AP} \) and \( \hat{E}_{b,i}^{SI} \) are the maximum axial energies of the \( i^{th} \) swallow segment of bolus \( b \). The denominators are the maximum energies over the entire bolus signal (defined as the signal between bolus onset and offset). As above, all energies were estimated within a 500-sample moving window with 50% overlap.

The scalars for the AP and SI channels (\( \Lambda^{AP} \) and \( \Lambda^{SI} \)) were then estimated as a linear combination of the average (\textit{mean}) and standard deviation (\textit{std}) of the candidate scalars. For instance, in this investigation,

\[
\Lambda^{AP} = mean(\lambda^{AP}_{b,i}) - j \times std(\lambda^{AP}_{b,i}) \tag{3.9}
\]

was considered as an estimate for the AP scalar with \( j = 0, 1, 2 \). Rows 2-4 of Table 3.1 document the effect of these scalar estimates on false positive reduction metrics.

**Maximum Energy Difference (MaxED)**

The energy of FP segments often falls below that of the lowest energy TP segment within the same bolus. Fig. 3.3(a) portrays the maximum energy of 30 randomly selected boluses (crosses), and their TP (open circles) and FP (dots) swallow segments. We can see generally, that FP segments have lower maximum windowed energy than do TP segments. An energy difference approach was thus devised; the maximum energy difference for bolus \( b \) was defined as,

\[
\delta_{b,*}^{AP} = \hat{E}_{b}^{AP} - \hat{E}_{b,*}^{AP} \tag{3.10}
\]

where \( \hat{E}_{b}^{AP} \) is the maximum windowed energy of bolus \( b \) and \( \hat{E}_{b,*}^{AP} = \min_{i} \hat{E}_{b,i}^{AP} \). Consequently we defined \( \Delta^{AP} \) as the set of \( \delta_{b,*}^{AP} \) that satisfy the following:

\[
\Delta^{AP} = \{ \delta_{b,*}^{AP} \mid \forall b : \delta_{b,*}^{AP} \geq max_{b}(\delta_{b,*}^{AP}) \times \beta \} \tag{3.11}
\]
where $\beta \in (0, 1]$ is an empirically tuned scalar to suit the characteristics of the signals of interest. In this set, there exists swallow segment $i'$ within bolus $b'$, for which the energy ratio, $\lambda_{b,i}^P$, as defined in Eq. 3.8 is maximized, i.e.,

$$\{ \exists! b', i' \mid \lambda_{b',i'}^P \geq \lambda_{b,i}^P ; \forall b : \delta_{b,i} \in \Delta^P, i \in I_b, i' \in I_{b'} \}$$  \hspace{1cm} (3.12)

where $I_b$ and $I'_b$ are the sets of candidate swallow indices for boluses $b$ and $b'$, respectively.

Finally, the scalar $\Lambda^P$ is set as the energy ratio of swallow segment $i'$ within bolus $b'$:

$$\Lambda^P = \lambda_{b',i'}^P$$

We estimated $\Lambda^S$ following the same procedure for the SI axis. Table 3.1 summarizes different estimations of these scalars.

Fig. 3.3(b) portrays FP and TP changes for different values of scalar $\Lambda^P$. The vertical line delineates the scalars ($\Lambda^P = 0.079, \Lambda^S = 0.078$) that yield the highest harmonic average, decreasing false positives by 11.5% with minimal change to the number of true positives.

Fig. 3.3(b) portrays FP and TP changes for different values of scalar $\Lambda^P$. The vertical line delineates the scalars ($\Lambda^P = 0.079, \Lambda^S = 0.078$) that yield the highest harmonic average, decreasing false positives by 11.5% with minimal change to the number of true positives.
3.5.3 Noise-floor False Positive Reduction

To further reduce false positives, we only accept candidates whose range exceeds that of the noise floor. This algorithm first computes the amplitude histogram of the bolus signal. The idea is that most of the noise will be low energy. Therefore, we estimate the range of the noise signal as $\alpha \times \sigma$, where $\sigma$ is initially the bolus signal variance ($\sigma = 1/N \sum (x - \mu)^2$) and $\alpha$ is a scalar multiplier. This is an estimate of the range of the noise (i.e., assuming that the noise resided within $\mu + \alpha\sigma$ and $\mu - \alpha\sigma$). The axial thresholds are then determined as:

$$T^{AP} = \alpha^{AP} \times \sigma^{AP}, \text{ and } T^{SI} = \alpha^{SI} \times \sigma^{SI}$$ (3.13)

Figure 3.4: Example of (a) a raw AP bolus signal, (b) segmentation with a FP case (first rectangle), and (c) segmentation after FP removal. The vertical lines indicate the VFSS-identified swallow onsets and the rectangles identify the segmented swallows.

To estimate the optimum values for $\alpha^{AP}$ and $\alpha^{SI}$, the following criterion function is considered:

$$J(\alpha^{AP}, \alpha^{SI}) = n_{TP}(\alpha^{AP}, \alpha^{SI}) - n_{FP}(\alpha^{AP}, \alpha^{SI})$$ (3.14)
where \( n_{TP} \) and \( n_{FP} \) are the number of TP and FP cases, expressed as a function of AP and SI scalars (\( \alpha^{AP}, \alpha^{SI} \)).

The optimal AP and SI scalars are given by:

\[
\alpha^{AP*}, \alpha^{SI*} = \arg\max_{\alpha^{AP}, \alpha^{SI}} J
\]

(3.15)

where \( \alpha^{AP*} = 7 \) and \( \alpha^{SI*} = 4 \) leads to a 74% FP reduction with only a 12% decrease in TP cases. Fig. 3.4 exemplifies a case where a FP swallow segment was removed after the application of this noise-floor FP reduction algorithm.

The energy and noise-floor false positive reduction methods were applied in parallel on segmented, pre-processed data. We only admitted candidate segments that were identified as valid by at least one of the two FP reduction methods. If the loss of TPs is capped at 20%, then the proposed methods lead to a dramatic reduction in FPs (-85.4%) and a low FDR (3.36%), while sacrificing only 15.1% of TPs, leading to 96.64% precision and 76.35% recall. (\( \Lambda^{AP} = 0.458, \Lambda^{SI} = 0.326, \alpha^{AP*} = 7, \) and \( \alpha^{SI*} = 4 \)). Although aggressive reduction of FP segments may lead to reduction of TP swallows, it ensures that the automatically segmented portion of the bolus contains swallowing activity. This is particularly important when extracting swallow level features for the purpose of discriminating between safe and unsafe swallowing.

### 3.6 Conclusion

The combined effect of the proposed energy and noise-floor methods was a definitive decrease in the number of false positives post-segmentation. The balance between FP reduction and loss of TPs can be fine-tuned to suit the specific accelerometric application by adjusting the axial thresholds. On the other hand, the proposed swallow trimming
algorithm successfully excluded the pre- and post-swallow signal fluctuations.

3.7 Acknowledgement

The authors would like to thank the Natural Sciences & Engineering Research Council of Canada (NSERC), J.L. (Allen) Yen Scholarship, Walter C. Sumner Memorial Fellowship, Edward S. Rogers Sr. Scholarship, and Ewing Rae Graduate Scholarship for providing the generous funding of this project. The authors would like to thank Dr. Ali-Akbar Samadani and Ka Lun Tam for their valuable insights.
Chapter 4

Automatic Discrimination Between Cough and Non-cough Accelerometry Signal Artefacts

This chapter was excerpted in entirety from the following submitted journal article: Mohammadi, H., Samadani, A. A., Steele, C., & Chau, T., Automatic Discrimination Between Cough and Non-cough Accelerometry Signal Artefacts. Journal of Biomedical Signal Processing and Control, submitted in April 2017. This article is currently under review with Biomedical Signal Processing and Control. The journal’s homepage is located at https://www.journals.elsevier.com/biomedical-signal-processing-and-control.

4.1 Introduction

A cough is a protective mechanical response [22] consisting of rapid contractions of the thoracic cavity, which generates a forceful and rapid expulsion of air that clears the airway of foreign material, fluid or mucus [23]. Cough can be symptomatic of various respiratory conditions such as asthma, rhinitis and gastro-oesophageal reflux disease in adults and protracted bronchitis in children [76]. Cough is also a normal reflexive response to
aspiration, the entry of foreign material into the airway, seen in people with swallowing
difficulties [77]. Hence, knowledge of cough severity, including intensity and frequency,
may inform clinical decision-making in terms of appropriate treatment of the underly-
ing issue. However, clinical assessments of cough often involve subjective judgement of
symptoms and symptom severity, leading to inconsistent symptom reports between pa-
tients and caregivers. Cough scores, diaries, symptom questionnaires, and visual analogue
scales generally lack validation as tools for evaluating cough severity [24].

4.1.1 Existing Cough Monitors

Given the above, there has been a growing interest in the development of accurate cough
monitoring devices [112, 113] in both human and veterinary health [114], [115]. Auto-
matic cough detectors may be more suitable for real-life day-to-day monitoring, where
continuous manual counting of coughs is not practical. Cough monitoring devices can be
used as a clinical informatic for a myriad of conditions, including tuberculosis [116] and
gastroesophageal reflux [113]. For example, Tracey et al. [117] emphasize the need for
a laboratory-free test for the diagnosis of pulmonary tuberculosis.

A number of considerations need to be taken into account when designing a cough
monitoring device. Firstly, cough monitors must reject ambient noise [118]. Secondly,
cough monitors must accommodate variation in the characteristics of coughs across indi-
viduals and conditions [116]. Thirdly, the device should be able to monitor the client over
long periods of time, especially during the night when self-reporting is not feasible [113].

Currently, there are a number of commercially available cough monitoring devices,
some of which are reviewed in [119]. Generally, these microphone-based systems are un-
able to distinguish true coughs from ambient noise and non-cough patient sounds [115].
In a recent validation against manually identified coughs, a commercial cough detector
yielded low sensitivity [120]. Likewise, in a separate comparative analysis, the perform-
ance of another commercial cough monitor was inconsistent across subjects [115]. The
To circumvent some of the above limitations, recent research on automatic cough detection has invoked multiple sensors. Drugman et al. [115] compared 6 different sensors against a commercial cough monitor, the Karmelsonix system (currently the PulmoTrack), finding that an omnidirectional lapel microphone was the most sensitive to coughs. On a separate front, Turner et al. [120] compared the counts of coughs detected by human experts against those identified by a sensor combination consisting of thoracic respiratory belt and tracheal and chest microphones. They found that the fully automatic system achieved very low sensitivity (32%). Recently, Hirai et al. [122] used a microphone (over the second intercostal muscle) and an accelerometer (positioned over the abdomen) to count the number of overnight coughs. By thresholding the envelope of the cough sounds, they achieved 95% correlation with the current gold standard of cough detection, which is expert review of video-audio recordings. However, their study only considered spontaneous, nocturnal coughs in hospitalized children with respiratory illness. It would be desirable to develop a cough detection system that works in a broader population.

Although multi-transducer approaches have produced promising results, they do require careful sensor positioning and attachment. Further, most still retain a microphone, precluding their use in noisy environments. An alternative approach may be to exclusively deploy a sensor, such as an accelerometer, that is insensitive to ambient acoustic noise.

This chapter presents a framework for detection of cough and non-cough events, using dual-axis accelerometry signals from a single accelerometer on the patient’s neck.
4.1.2 Cervical Accelerometry

Cervical accelerometry is a novel non-invasive and non-radiographic assessment technique where the patient wears a dual-axis accelerometer midline, below the laryngeal prominence (commonly known as the Adam’s apple). The accelerometer captures epidermal vibrations in the anterior-posterior (AP) and superior-inferior (SI) directions, thus facilitating day-to-day monitoring of pharyngeal vibrations. This chapter proposes an algorithmic approach to accurately differentiate coughs from a resting state, swallowing, tongue movements and speech on the basis of dual-axis accelerometry signals.

Figure 4.1: Location and orientation of the Dual-axis accelerometer sensor (shaded oval) on a participant’s neck.

4.2 Experimental Setup

4.2.1 Data Sets

To develop and validate a novel cough detection algorithm, we used two different data sets of dual-axis accelerometry signals (herein referred to as the ‘instructed’ and ‘reflexive’ cough data sets). For both data sets, hyo-laryngeal accelerations were recorded at 10 kHz with 12-bit resolution using a dual-axis accelerometer (ADXL327, with measure-
Figure 4.2: AP and SI signals containing three swallows (black dotted rectangles) and one reflexive cough (red rectangles).

ment range of ±2.5g and a sensitivity of 420mV/g), placed on and slightly below the laryngeal prominence (commonly known as the Adam’s apple) to measure epidermal vibrations accompanying swallowing, speech, tongue movements, and coughs [48] (shown in Figure 4.1). The sensor axes were aligned with the superior-inferior (SI) and anterior-posterior (AP) anatomical directions. Signals were filtered in hardware with a passband between 0.1 Hz and 3 kHz to remove the DC and high frequency noise. The digitized samples were stored on a laptop computer for subsequent offline analysis.

**Instructed Cough Data Set**

Fifteen subjects participated in the instructed cough data collection. Each participant attended two data collection sessions, each lasting approximately 45 minutes. The protocol was approved by the research ethics board of the participating hospital. Each participant provided written, informed consent.

The first session consisted of only tongue motions (tongue protruding out of the mouth with lips pursed, tongue contacting the inside of the left and right cheeks separately,
and tongue at rest). The second session comprised coughing, swallowing water, and saying “on” or “off” out loud. Prior to data collection in each session, the experimenter demonstrated the required tasks and provided participants with 5 minutes to practice the tasks. Within each session, participants were cued to perform the tasks in a pseudo-random order through a LabVIEW interface. Participants were instructed to perform the task within the 4 seconds immediately following the presentation of each cue. Each task was repeated 20 times by each participant. In total, 300 examples of each task were obtained (15 participants × 20 examples of each task/participant). The experimenter noted when the participant performed the incorrect task. The data set thus included accelerometry signals pertaining to tongue movements, coughs, swallows, rest state, and speech. All signals were trimmed automatically by identifying the 1 second segment with maximum energy within the 4 second recording. In particular, the trimmed signal was derived by centering a one second window around the location of the signal peak in the maximum energy segment.

**Reflexive Cough Data Set**

Involuntary reflexive coughs were derived from a previously reported dataset [123]. These coughs are associated with swallowing activity, reflecting aspiration events, as opposed to coughs elicited in a cough reflex test (where an irritant like citric acid is infused through a nebulizer to observe the expected cough reflex response).

Dual-axis accelerometry signals were collected from 196 consenting adults living with the effects of stroke or brain injury, or with otherwise unrelated suspicion of dysphagia. Each participant performed a series of 6 discrete sips of thin liquid barium (Bracco Varibar Thin Liquid Barium, diluted to a 20% w/v concentration).

Segments of the accelerometry signals were manually annotated with the labels listed in Table 4.1, using a graphical user interface (GUI) designed in MATLAB that enabled simultaneous visual and aural review of the signals. The GUI enabled marking the start
Table 4.1: Accelerometry signal labelling

<table>
<thead>
<tr>
<th>Label</th>
<th>Type</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cough</td>
<td>Clearly audible cough</td>
</tr>
<tr>
<td>2</td>
<td>Swallow</td>
<td>Clearly audible swallow</td>
</tr>
<tr>
<td>3</td>
<td>Speech</td>
<td>Vocalizations</td>
</tr>
<tr>
<td>4</td>
<td>Clearing swallow</td>
<td>Attenuated swallow sound subsequent to a clearly audible swallow</td>
</tr>
<tr>
<td>5</td>
<td>Throat clear</td>
<td>Not a cough but an audible throat clearing attempt</td>
</tr>
<tr>
<td>6</td>
<td>Breathing</td>
<td>Audible inspiration</td>
</tr>
<tr>
<td>7</td>
<td>Ambiguous</td>
<td>Indistinguishable sound</td>
</tr>
</tbody>
</table>

and end times of different events (Table 4.1). Through this procedure, a total of 51 coughs (average duration 862.61±536.1 ms) were identified. To facilitate the development of a cough detector, we further extracted 45 swallow segments (average duration 1198.17±493.6 ms) from the signals containing the identified coughs. Additionally, 51 rest segments were extracted from the first 10 seconds of recorded data prior to swallowing task commencement. In particular, for a given cough, the pre-task signal segment of the same duration and minimum energy was chosen as the corresponding rest segment. Rest segments were only selected from recordings containing at least one cough segment.

Figure 4.2 exemplifies manually annotated coughs and swallows for a participant in the reflexive cough data set. This recording contained three swallows, outlined by the dotted black rectangles, and one cough event, indicated by the red rectangles.

4.3 Methodology

Signal pre-processing included the removal of noise and head movements from the acceleration signals. Meta-feature-based representation of the pre-processed signals were then computed followed by feature selection/extraction to identify the most salient features. The salient features were then classified over 10 runs of 5-fold cross-validation.
The following sections elaborate upon the analytical framework in detail.

4.3.1 Pre-Processing

Preprocessing included de-noising [108] and head movement suppression [71, 72]. Additionally, high frequency noise was filtered by wavelet packet decomposition using a 4-level discrete Meyer wavelet and Shannon entropy [123].

4.3.2 Meta-Feature-Based Representation of Signals

A total of 35 temporal [69, 124], time-frequency [40, 56, 69, 124], frequency [124], information-theoretic [124] features for each segment (i.e., cough, speech, swallow, rest) were computed from the AP and SI axes separately. Additionally, we computed the mutual information, cross-entropy rate, and cross-correlation between the corresponding AP and SI signals [69, 124, 125]. Details of the calculated features are provided in Appendix A.

4.3.3 Salient Feature Identification

To determine parsimonious and discriminatory feature vectors, three selection algorithms were considered: binary genetic algorithm (BGA) [97], elastic net [99], and filter-based feature selection [98]. Additionally, a reduced feature set was also derived via principal component analysis (PCA).

Binary Genetic Algorithm (BGA)

To invoke GA-based feature selection, we coded candidate feature vectors as a chromosome of boolean values, each gene indicating whether the corresponding feature is selected. In other words, a candidate solution, \( x \), to the optimization problem was a binary string of the same length as the number of features:
\[ x = \{ f_1, f_2, ..., f_n \}, \quad f_i \in \{0, 1\}, \quad i \in [1, n] \]

where \( n \) is the number of features and \( f_i = 1 \) indicates that the \( i^{th} \) feature is included in the reduced feature set. As a simplified example, \( x = \{0, 0, 1, 1, 0\} \) would indicate that only the 3\(^{rd}\) and 4\(^{th}\) features are included in the reduced feature set.

A population size of 50 was selected along with a tournament size of 2. Optimization proceeded for a maximum of 100 generations. Crossover and mutation rates of 0.8 and 0.1 were selected respectively. Additionally, in order to keep the best solutions in the population pool, elitism of size 2 was selected. The entire optimization was iterated 30 times. The aforementioned algorithmic parameters were derived empirically based on training data.

**Filter-Based Feature Selection**

In this approach, features were ranked based on their uni-dimensional class separability score [98]. While inter-feature correlations were not considered, this approach offered computational simplicity. The top 5 to 30 features were considered for the subsequent classification experiments.

**Elastic Net**

The elastic net is a regularized binomial logistic regression which is used to select a subset of features. With the elastic-net penalty of [126], a set of 10 equally spaced ridge-LASSO penalty \( \alpha \) values in the range of \([0.1, 1]\) and 100 values of the penalty parameter \( \lambda \) were tested. A pair of \( \alpha \) and \( \lambda \) values yielding the minimum 5-fold cross-validated squared-error on the training data was selected using the generalized binomial logistic regression models toolbox from [127].
Principal Component Analysis

In addition to the above feature selection approaches, we also generated a reduced set of transformed features using principal component analysis (PCA) [100]. The components were then sorted in descending order based on their corresponding eigenvalues. Classification was then evaluated using different subsets selected from the top of the sorted components in the inner cross-validation.

4.3.4 Classification

In order to classify cough segments versus rest states, and all artefacts, we deployed artificial neural networks (ANN) and support vector machines (SVM) as our classification algorithms.

Artificial Neural Network

Neural networks with a single hidden layer of 20 units and two output units were implemented. This configuration was selected empirically based on the training performance. The inputs were feature values from the reduced feature subsets described above. Networks were trained using Bayesian regularized back-propagation with a mean-squared error criterion function and evaluated via 5-fold cross-validation with a 80-20 split into training and validation on the training folds.

Support Vector Machines

A support vector machine (SVM) with a radial basis function (RBF) kernel with scaling factor of size 2 was deployed [128]. The sequential minimal optimization algorithm [129] was used to train a soft-margin SVM classifier.
4.3.5 Validation

To validate the proposed cough detection system, comparisons between pairs of feature selection and classification approaches (e.g., elastic net + SVM) were conducted based on classification performance and model complexity such as number of features. The comparison was conducted for a feature set size ranging from 1 to 35 (i.e., entire feature set).

Feature selection and classifier pairings were evaluated using 10 runs of 5-fold cross-validation. The model performance was evaluated based on the mean±standard deviation of the pair’s accuracy, as well as true positive and true negative rates over 10 runs of 5-fold cross-validation of the test cases. In each run, the data set was divided into 5 folds. Each fold was considered as the test set while the feature selection and classifier pair was trained using the remaining four folds and blind to the test cases. This process was repeated 10 times.

The elastic net and SVM hyper-parameter (RBF variance) and SVM slack parameter were tuned using the training data set based on inner cross-validation. The inner cross-validation accuracy values of different pairs were evaluated using the Wilcoxon ranksum test.

The accuracy of each feature reduction-classification combination was compared against the 7 other possible combinations using a one-sided Wilcoxon rank sum test. These comparisons were iterated with the differently sized feature subsets (from 1 to 35 features), yielding the feature reduction-classification combination with the highest accuracy for each feature dimensionality.
4.4 Results

4.4.1 Cough versus Rest

When discriminating between instructed coughs and the rest state, the BGA-SVM combination triumphed over other feature reduction-classification combinations, yielding an accuracy of 99.26$\pm$0.12%, TPR of 99.96$\pm$0.16% and TNR of 98.6$\pm$0.15%.

In the discrimination between reflexive coughs and the rest state, the elastic net-SVM combination exhibited superiority over other feature reduction-classification combinations, achieving an accuracy of 90$\pm$13.9%, TPR of 100$\pm$0% and TNR of 95$\pm$6.9%.

4.4.2 Cough versus Non-Cough Artefacts

The more challenging classification problem was to discriminate between cough segments and other non-cough artefacts (combination of swallow, speech, and tongue movement segments).

Instructed Coughs

Figure 4.3 illustrates test accuracies (top row) and training and testing error rates (bottom row) for ANN and SVM classification of instructed cough segments and non-cough artefacts, for feature sets of increasing dimensionality. Figure 4.4 is a heat-map of the p-values from a right-tailed Wilcoxon rank sum test comparing the best accuracies of each feature reduction-classifier combination (i.e., accuracies corresponding to its best performing feature set) against the best accuracies of those specified across the columns. A small p value means that the feature reduction-classifier combination on a given row has a significantly higher median accuracy than that of the feature reduction-classifier combination in the intersecting column. From Figure 4.4, it is evident that the leading feature reduction-classifier combination was the elastic net-SVM combination ($p < 0.001$), with accuracy, TPR and TNR of 90.2$\pm$3.6%, 91.2$\pm$4.8%, and 89$\pm$5.5%, respectively. Among
Figure 4.3: Instructed cough versus non-cough artefact accuracy (top 2 graphs) and error rate (bottom 2 graphs) comparisons among feature reduction-classification combinations. Only testing accuracies are shown in the top graphs whereas the lower graphs depict both training and testing error rates. Note that the elastic net did not converge for feature dimensionalities less than four and hence the incomplete trend (for lower feature dimensionalities) for some feature reduction-classification combinations.

The salient features selected by elastic net over the 5-fold cross-validation runs, features that estimate the complexity of the signals were frequently used. The plots on the left side of Figure 4.5 display the acceleration trajectories (i.e., SI versus AP) of instructed cough segments.

Reflexive Coughs

Figure 4.6 presents the accuracies and error rates of classifying reflexive coughs versus non-cough artefacts, over different feature subsets using different feature reduction-classification combinations. Figure 4.7 compares the best accuracies of each feature reduction-classifier combination (i.e., accuracies corresponding to its best performing
Figure 4.4: Comparisons of accuracies of distinguishing between instructed cough and non-cough artefacts. Each row highlights the comparison between a specific feature reduction-classification combination and all other combinations (specified in the column labels). Entries depict magnitude of \( p \)-values of the Wilcoxon rank sum test.

feature set) against the best accuracies of those specified across the columns. From this heat-map, the leading feature reduction-classification combination for discerning between reflexive coughs and non-cough artefacts was the BGA-SVM combination, yielding accuracy, TPR and TNR of 80.3\( \pm \)10.5\%, 80.9\( \pm \)15.8\%, and 79.8\( \pm \)18.6\%, respectively. The plots on the right side of Figure 4.5 show the acceleration trajectories of reflexive coughs.

4.5 Discussion

4.5.1 Cough versus Rest Classification

For the discrimination between coughs and the resting state, the proposed system exceeded the sensitivity and specificity reported in previous cough studies. Validated methods of cough assessment currently include [130]: (1) methods based solely on audio-recordings [131, 132, 133], (2) those that combine audio-recordings and electromyography.
Figure 4.5: Acceleration trajectories for instructed (left) and reflexive coughs (right) coughs. Signals shown are from one participant each from the instructed and reflexive cough data sets. The time elapsed is implicit in the 2-dimensional trajectories (bottom) and indicated by the color coding.

(EMG) [134, 135], (3) multiparametric devices based on plethysmography, electrocardiography, and an accelerometer [136], and (4) methods that use vibrations of the suprasternal notch, recorded using an accelerometer [137]. Sensitivity and specificity have varied: 78.1% and 99.6% respectively in [136], 80% and 96% in [133], and 91% and 99% in [132]. While our sensitivity and specificity for the cough versus rest problem exceed all of these numbers, some important methodological and analytical differences are noteworthy. Past systems focused on manually counting the number of coughs rather than automatically discriminated between cough signals and non-cough artefacts. The sensors used in past devices also differed. None of the previously reported cough monitoring systems conducted automated signal analyses [130]. For these reasons, a direct comparison with previous literature is difficult.
Figure 4.6: Reflexive cough versus non-cough artefact accuracy (top 2 graphs) and error rate (bottom 2 graphs) comparisons among feature reduction-classification combinations. Only testing accuracies are shown in the top graphs whereas the lower graphs depict both training and testing error rates. Note that the elastic net did not converge for feature dimensionalities less than four and hence the incomplete trend (for lower feature dimensionalities) for some feature reduction-classification combinations.
### 4.5.2 Optimal Feature Dimensionality

Although the SVM emerged as the preferred classifier for both instructed and reflexive cough classification problems, classifier accuracies and error rates suggested different maximum feature dimensionalities. The bottom right plot of Figure 4.3 shows that for the instructed cough versus non-cough artefact problem, the SVM error rates of the training and test data diverged after 15 features. This divergence of error rates is attributed to over-fitting, suggesting a cap of 15 features for the instructed cough classifier. Correspondingly, the testing accuracies generally saturated or decreased after 15 features. In contrast, in the discrimination between reflexive coughs and non-cough artefacts, as seen in the bottom right plot of Figure 4.6, error rates diverged after 11 features, suggesting a cap of 11 features for the reflexive cough classifier.

![Comparison of accuracies of distinguishing between reflexive cough and non-cough artefacts.](image)

Figure 4.7: Comparisons of accuracies of distinguishing between reflexive cough and non-cough artefacts. Each row highlights the comparison between a specific feature reduction-classification combination and all other combinations (specified in the column labels). Entries depict magnitude of p-values of the Wilcoxon rank sum test.
4.5.3 Salient Features

Five features were selected frequently for the instructed and reflexive cough classification problems (i.e., discrimination between cough versus non-cough artefacts): mean SI, Lempel-Ziv SI, maximum energy AP, variance AP, and skewness AP. Features from both AP and SI axis were selected, emphasizing the complementary information afforded by each axis, which is in accordance with findings from previous studies [19, 69].

In addition, a unique set of salient features were selected for the reflexive classification problem (cough versus non-cough artefacts) from the information theoretic domain (e.g., entropy and entropy rate) and the combination of two axis (e.g. mutual information and cross-correlation). The entropy rate characterizes a stochastic process and measures the regularity of the accelerometry signal [125]. Entropy reflects the information content within a channel while mutual information measures the redundancy of information between channels. The selection of these features indicates that the amplitude distributions of reflexive coughs have morphological characteristics that are distinct from those of non-cough artefacts. Additionally, the appearance of cross-correlation among the salient features suggests that the correlation between the AP and SI axes is particularly discriminatory for reflexive rather than instructed coughs.

In contrast, the majority of salient features for the instructed classification were from the time domain (e.g. memory and kurtosis). The memory of a signal measures the temporal extent of the correlation of the neighbouring data samples. The selection of this feature suggests that the autocorrelation of the signal differs between instructed coughs and non-cough artefacts. The kurtosis of a signal [69] measures the peakedness of the amplitude distribution, suggesting that low amplitude signal values have a differential contribution to instructed coughs and non-cough artefacts.

The finding that the most salient feature subsets were problem-dependent suggests that the signals corresponding to instructed coughs and those that spontaneously occur (i.e., reflexive) are very different in nature. Thus, studies based strictly on instructed
coughs should be interpreted with caution when generalizing to spontaneous coughs.

### 4.5.4 Cough Acceleration Trajectories

Upon closer examination of Figure 4.5, there seem to be clearly defined trajectories in both the AP and SI axes for coughs from the instructed cough data set (individuals without dysphagia). In contrast, there is no obvious, unified pattern for coughs from the reflexive cough data set. This lack of conspicuous pattern was observed in all participants from the reflexive cough data set, highlighting the non-trivial inter- and intra-subject signal variability among individuals with dysphagia.

By examining the 2-dimensional cough trajectories (plots in the lower half of Figure 4.5), we note that the magnitude of acceleration for reflexive coughs exceeds that of the instructed coughs by an order of magnitude. This observation suggests that the forceful and rapid expulsion of air is markedly more intense in spontaneous coughing.

### 4.5.5 SVM versus ANN Classifiers

SVM performed better than ANN in the majority of comparisons (Wilcoxon ranksum $p < 0.05$). This may be due to the tendency of SVM classifiers to hone in on the global minimum, while ANN classifiers are more susceptible to local minima traps [138]. Moreover, SVM training was generally faster than that of ANN, making SVM a more suitable candidate for online training and classification.

One of the advantages of the proposed system is its simplicity, deploying only a single accelerometer. Additionally, the proposed system is not affected by ambient noise, therefore suitable for day to day monitoring in noisy environments. Consequently, potential applications such as cough frequency monitoring during sleep studies and veterinary medicine applications may benefit from this algorithm.
4.6 Conclusion

An automatic cough detection and monitoring system discriminated cough accelerometry signals from other artefacts such as rest state, swallowing, tongue movements, and speech. Both instructed and reflexive coughs were considered. The proposed system discriminated between coughs and rest state with accuracies of 99.64% and 90% for instructed and reflexive coughs, respectively. Additionally, the cough segments were discriminated from the non-cough artefacts with accuracy values of 90.2% and 80.3% for instructed and reflexive data sets.

4.7 Acknowledgement

The authors would like to thank the Natural Sciences & Engineering Research Council of Canada (NSERC), J.L. (Allen) Yen Scholarship, Walter C. Sumner Memorial Fellowship, and Ewing Rae Graduate Scholarship for providing the generous funding of this project. The authors would also like to thank Dr. Khondaker Mamun, Kamran Mastery, and Laura Cuevas Moreno for their valuable help in data acquisition and labelling phase.
Chapter 5

Classification of Dual-Axis Swallowing Accelerometry Signals with Bolus Length Estimation and Instance Selection

This chapter will eventually be submitted to Medical Engineering & Physics following approval by the industry partner.

5.1 Introduction

Dysphagia refers to any difficulties in the processes of transferring food or liquid from the mouth to the stomach [2]. The safety of a swallow can be scored on an 8-point Penetration-Aspiration scale [1]. A swallow is considered to be safe when the material either does not enter the airway or enters the supraglottic area at a point above the vocal folds, and then is successfully ejected. Penetration, on the other hand, occurs when the foreign material (bolus) either enters the supraglottic space and is not ejected, or
makes contact with the true vocal folds. Aspiration is the entry of foreign material into the airway below the true vocal folds. Individuals with dysphagia are at higher risk of aspiration [2]. When an individual experiences aspiration, a reflexive cough is a normal reaction. However, studies show that among the acute stroke population, 2% to 25% of aspirations occur “silently” with no noticeable physiological signs [36].

The videofluoroscopy swallowing study (VFSS) is the gold standard for the detection of dysphagia [38]. During this procedure patients are asked to swallow different consistencies of barium-coated boluses [40] and are exposed to ionizing radiation. In addition, VFSS requires expensive equipment and experienced clinicians [41]. Single [48, 57, 62] and subsequently dual-axis [125] accelerometry has been proposed as potential non-invasive adjunct to non-instrumental clinical swallowing assessment, where an accelerometer is positioned below the laryngeal prominence (commonly known as Adam’s apple) to measure vibrations associated with swallowing. In the dual-axis case, the sensitive axes of the accelerometer are aligned to the anterior-posterior (AP) and superior-inferior (SI) anatomical axes. Figure 5.1 demonstrates the location and orientation of the sensor on a patient’s neck. Our previous research has identified a significant correlation between the magnitude of hyolaryngeal excursion and integrated dual-axis acceleration signals [63] and inter-axis signal differences in both time and time-frequency domains [69]. These findings justify the consideration of dual-axis accelerometry over its single-axis counterpart.

5.2 Automatic Classification of Swallowing Accelerometry Signals

Past attempts at automatically classifying swallowing accelerometry signals are summarized in chronological order in Table 5.1. The first 4 studies [40, 64, 65, 66] deployed single-axis accelerometers, and had either low specificity or low sensitivity for detecting
penetration-aspiration events. The last three classified at the participant level whereas the fourth classified manually segmented signals. The first 2 dual-axis studies [42, 53] reported high sensitivity but much lower specificity in discriminating between safe and unsafe swallows (i.e., those with airway entry). Both studies had modest sample sizes. Merey et al. [55] and Sejdic et al. [56] achieved impressive detection of aspiration among pediatric and older adult samples, respectively, but required high-dimensional feature sets. Generally, previous dual-axes swallowing accelerometry studies in Table 5.1 share the following two limitations.

1. All the dual-axes studies have relied on VFSS-demarcated segments rather than automatic localization of segments of interest, which precludes direct translation of these algorithms into a stand-alone tool.

2. Previous studies have adopted an ipsative classification paradigm, wherein, for every unseen case, the classifier must make a forced choice, for example, between safe and unsafe swallowing.

Swallowing accelerometry is susceptible to numerous sources of noise, including, but not limited to, head movements [71], vocalizations [70], vasomotion [139] and other physiological and motion artifacts [140]. Moreover, cervical accelerometry signal characteristics are known to exhibit intra- and inter-subject variability [139, 141]. Hence, it is sensible to afford the classifier an option to abstain from a decision when the classifier is uncertain,
Table 5.1: Automatic classification of swallowing accelerometry signals

<table>
<thead>
<tr>
<th>Study</th>
<th>Classification problem</th>
<th>Bolus consistency</th>
<th># Axes</th>
<th>Classification/validation</th>
<th># Participants</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lazareck &amp; Moussavi [65]</td>
<td>“normal” vs. “dysphagic” patient</td>
<td>Thin</td>
<td>1</td>
<td>LDA/LOO</td>
<td>27</td>
<td>85±8</td>
<td>63±9</td>
</tr>
<tr>
<td>Aboofazeli &amp; Moussavi [66]</td>
<td>“normal” vs. “dysphagic” patient</td>
<td>Various</td>
<td>1</td>
<td>3-NN/LOSO</td>
<td>26</td>
<td>Accuracy:83</td>
<td></td>
</tr>
<tr>
<td>Lee et al. [40]</td>
<td>swallows vs. aspirations</td>
<td>Various</td>
<td>1</td>
<td>RBF/10-fold CV</td>
<td>100</td>
<td>88±12</td>
<td>75±16</td>
</tr>
<tr>
<td>Yadollahi &amp; Moussavi [64]</td>
<td>“normal” vs. “dysphagic” patient</td>
<td>Thick</td>
<td>1</td>
<td>LDA/LOO</td>
<td>8</td>
<td>66</td>
<td>90</td>
</tr>
<tr>
<td>Lee et al. [42]</td>
<td>depth of airway invasion (none vs. aspiration)</td>
<td>Various</td>
<td>2</td>
<td>LDA/10-fold CV</td>
<td>24</td>
<td>50±6</td>
<td>100±0</td>
</tr>
<tr>
<td></td>
<td>bolus clearance from valleculae (no residue vs. substantial residue)</td>
<td></td>
<td></td>
<td>LDA/10-fold CV</td>
<td></td>
<td>75.5±15.4</td>
<td>91.9±8.6</td>
</tr>
<tr>
<td></td>
<td>bolus clearance from pyriform sinuses (no residue vs. substantial residue)</td>
<td></td>
<td></td>
<td>LDA/10-fold CV</td>
<td></td>
<td>81.7±25.4</td>
<td>86.8±7.3</td>
</tr>
<tr>
<td>Nikjoo et al. [53]</td>
<td>safe vs. unsafe (airway entry) swallows</td>
<td>Various</td>
<td>2</td>
<td>SVM-RBF/10-fold CV</td>
<td>30</td>
<td>64±9</td>
<td>97±2</td>
</tr>
<tr>
<td>Merey et al. [55]</td>
<td>swallows without vs. with airway invasion</td>
<td>Various</td>
<td>2</td>
<td>SVM-RBF/8-fold CV</td>
<td>29</td>
<td>92.2±2</td>
<td>89.6±1</td>
</tr>
<tr>
<td>Sejdic et al. [56]</td>
<td>safe swallows vs. aspirations</td>
<td>Thin</td>
<td>2</td>
<td>LDA/LOO</td>
<td>40</td>
<td>95.6</td>
<td>92.5</td>
</tr>
<tr>
<td>Steele et al. [54]</td>
<td>safe (PAS&lt;3) vs. unsafe (PAS≥3)</td>
<td>Thin</td>
<td>2</td>
<td>LDA/LOO</td>
<td>37</td>
<td>77</td>
<td>90</td>
</tr>
</tbody>
</table>

Abbreviations: LDA=linear discriminant analysis; RBF=radial basis function; SVM = support vector machine; 3-NN = 3-nearest neighbor; CV = cross-validation; LOO=leave-one-out; LOSO = leave-one-subject-out; PAS=penetration-aspiration scale

due for example, to poor signal quality or signal characteristics that do not resemble any known category.

To address the aforementioned shortcomings of previous research, we propose a classification framework using swallowing accelerometry signals recorded in both AP and SI directions, with reduced dependency on VFSS-demarcations and the possibility of rejecting a test case on the basis of classifier uncertainty. The ability of the proposed system to discriminate between safe and unsafe (airway invasion at or below the true vocal folds) swallowing activities is evaluated using a total of 1,649 signal segments, collected from 305 consenting patients from 8 different hospitals. Since this data set represents a much larger sample than in previous studies and contains a more diverse population with dif-
ferent medical conditions, it is more challenging compared to the previous classification studies with fewer data points [40, 42, 64, 65, 66].

5.3 Instance Selection

Instance selection refers to a family of methods in machine learning that aims to reduce the volume of a given data set to accelerate the training and testing processes while maintaining or surpassing the classification accuracies obtained with the full data set [142]. In general, instance selection algorithms extract a subset of instances from data sets that are suspected of containing ambiguous, superfluous, or noisy data points [143]. The intent is that the extracted subset optimizes classification performance. Ambiguous data points are the instances with classification posteriors close to the classification threshold, while superfluous data points bring no additional value to classification and noisy data points lead to false classification predictions [143]. The choice of instance selection algorithms is problem-specific and no one algorithm is superior over others in all contexts [143]. A complete review of instance selection algorithms may be found in [144, 145].

Instance selection algorithms can be categorized according to the process of deriving the data subset (i.e. incremental, decremental, batch, mixed, and fixed) [144], the type of discarded instances (i.e. boundary, central, or both), and the selection criterion (i.e., classification performance or feature values) [146]. Based on the process of deriving the data subset, instance selection algorithms can be organized into five categories [144]:

- **Incremental:** Instance selection begins with an empty subset and incrementally adds data points by analyzing the instances in the training set [147].

- **Decremental:** The decremental algorithm begins with the entire training set and removes data points that are suspected of being unnecessary or superfluous; these data points meet the predefined selection criterion [148].

- **Batch:** The batch instance selection algorithms does not remove instances until
all data points have been analyzed. Instances that meet the selection criterion are marked but not removed until all data points have been considered, at which time, all the marked instances are discarded [149, 150].

- **Mixed**: Mixed instance selection starts with a preselected subset of data points and either adds instances to or removes instances from this subset.

- **Fixed**: Fixed instance selection algorithms constitute a subfamily of the mixed algorithm, where a predetermined subset size is maintained while adding instances to or removing instances from the subset.

Instance selection algorithms can also be classified according to the type of discarded data, namely, points from the decision boundary, “central” points within the boundaries, or combinations thereof [144]:

- **Condensation**: These methods retain data points at the border among classes while selecting central (internal) instances for removal. They argue that the instances closer to the decision boundary play a key role in the classification process while the central data points have relatively little effect on classification performance [151]. Although training accuracy may be preserved with this scheme, the overall test accuracies are often negatively affected [144]. Since the number of central data points are often larger than the border instances, the condensation algorithms generally achieve high rates of data reduction [144].

- **Edition**: These instance selection algorithms retain the central data points. These methods aim to identify instances that are ambiguous and not well-classified [144], specifically by their nearest neighbours. However, superfluous central data points that do not necessarily contribute to classification are not removed in these algorithms. The general test accuracies are positively affected while data reduction is modest compared to the condensation instance selection algorithms [144].

- **Hybrid**: Hybrid instance selection algorithms combine condensation and edition approaches to select both boundary and central instances to maintain or improve
classification accuracies [144].

Lastly, instance selection algorithms can be understood in terms of their selection criterion [152]. Wrapper algorithms embed instance selection in the process of classifier evaluation. Generally, instances with negligible contribution to model prediction are discarded from the training set. The majority of the wrapper algorithms are based on some measure of misclassification of the instances [146]. In contrast, filter-based instance selection rejects instances based on a selection criterion which is independent of the training algorithm but usually relating to the feature values of the instances [144]. The filter approaches either find representative instances from different subspaces of the data set, or base selection on the similarities between pairs of instances. An extensive review of wrapper and filter instance selection algorithms is presented in [146].

5.4 Data Set

The data set analyzed here was an expanded version of that reported in [123]. Briefly, acceleration signals were collected from both axes (anterior-posterior (AP) and superior-inferior (SI)) of a dual-axis accelerometer (ADXL327, with measurement range of $\pm 2.5 \, g$ and a sensitivity of $420 \, mV/g$) situated on and slightly below the laryngeal prominence (commonly known as the Adam’s apple) of participants with suspicion of swallowing difficulties. Acceleration signals were recorded at 10 kHz with 12-bit resolution and filtered in hardware using a passband between 0.1 Hz and 3 kHz. The digitized samples were then stored on a computer with concurrent videofluoroscopy for offline analysis. Signals were recorded while patients took 6 sips of thin liquid barium (Bracco Varibar Thin Liquid Barium diluted to a 20% w/v concentration). A sip of barium-coated liquid is referred to as a bolus, which can be ingested in one or multiple swallows. Bolus onset and offset were marked in the accelerometry signals according to expert annotations of the corresponding videofluoroscopy recordings. A total of 1,649 usable boluses were
Table 5.2: Number of participants and boluses (thin consistency)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Bolus</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>243</td>
<td>Safe 1532</td>
</tr>
<tr>
<td>Unhealthy</td>
<td>52</td>
<td>Unsafe 117</td>
</tr>
<tr>
<td>No label</td>
<td>10</td>
<td>No label 0</td>
</tr>
<tr>
<td>Total</td>
<td>305</td>
<td>Total 1649</td>
</tr>
</tbody>
</table>

identified. A bolus was labeled as unsafe if it contained at least one swallow with a Penetration-Aspiration Scale (PAS) [1] score of 3 or higher while a safe label was given otherwise. For the purpose of this research, only swallows pertaining to thin liquid barium consistency were considered. Table 5.2 summarizes the characteristics of the data set.

5.5 Methodology

5.5.1 Pre-processing and Swallow Segmentation

AP and SI signals were de-noised using 10-level wavelet decomposition with Daubechies-8 mother wavelets [53]. Signal artefacts relating to head movement were removed by subtracting a B-spline approximation of low frequency (<5 Hz) signal components [72] while vocalizations were suppressed by eliminating signal segments with periodic behaviors, as detected by pitch tracking [70]. Channel-specific normalization was applied to the bivariate bolus signals, to scale the signals to [0, 1].

AP and SI variance signals were computed by estimating the sample variance within windows of size 200 data points, shifted along each of the AP and SI signals with 50% overlap. The swallows were then segmented by subjecting the variance signals to a sequential fuzzy c-means algorithm [19]. The aforementioned segmentation algorithm was too liberal, admitting pre- and post-swallowing activity while also giving rise to non-swallow segments or false positives. A kernel density estimation-based algorithm [123] was used to adaptively trim the swallow segments, while energy and noise floor
algorithms [123] reduced the number of false positive swallow segments.

5.5.2 Feature Selection and Extraction

Time, frequency, time-frequency, information theoretic domain features for both AP and SI axis and channel combination features at both bolus- and swallow-level were calculated. Details of the calculated features are provided in Appendix A.

The elastic net is a regularized binary logistic regression which is used to select a subset of features. It linearly combines the penalties of the LASSO (Least Absolute Shrinkage and Selection Operator) and ridge regularization methods [99]. For observations \(x_i \in \mathbb{R}^p\) and responses \(y_i \in \mathbb{R}, i = 1, 2, \ldots, N\), the objective function for elastic net logistic regression is given as

\[
- \left[ \frac{1}{N} \sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right] + \lambda \left[ (1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1 \right],
\]  

(5.1)

where \(N\) is the number of observations, \(x_i\) are the features and \(y_i\) the corresponding class label, at observation \(i\), \(\beta_0\) and \(\beta\) are the intercept and predictor coefficients respectively, \(0 \leq \alpha \leq 1\) is a compromise between ridge (\(\alpha = 0\)) and LASSO (\(\alpha = 1\)) and interpolates between the \(L^1\) norm and the \(L^2\) norm of \(\beta\) (\(\|\beta\|_1\) and \(\|\beta\|_2^2\)). \(\lambda \geq 0\) is a complexity parameter which introduces sparsity over the feature space. The goal is to find the \(\beta_0\) and \(\beta\) that minimize Eq. 5.1.

With the elastic-net penalty given in Eq. 5.1 [126], a set of 10 equally spaced ridge-LASSO penalty \(\alpha\) values in the range of \([0.1, 1]\) and 100 values of the penalty parameter \(\lambda\) were tested. An optimal pair of \(\alpha\) and \(\lambda\) values were selected by minimizing the 5-fold cross-validated squared-error on the training data [127].
5.5.3 Noise-floor Bolus Length Estimation

The majority of the existing studies are dependent on VFSS to demarcate the bolus onset and offset of the acquired acceleration signals. As a result, the existing systems are not completely automated and rely on an external point of reference to segment the signal portions of interest. The proposed noise-floor bolus length estimation reduces the level of VFSS-dependency of the acquired acceleration signals by adding a cushion of 5000 samples before and after the VFSS annotated boluses and subsequently re-estimating the bolus boundaries. This is possible since the recordings of the accelerometer were continuous. By shifting the VFSS annotated onset to the left and the offset to the right, a more liberal bolus length is selected. The noise-floor algorithm then automatically estimates the bolus length to be as close as possible to the VFSS annotated onset and offsets.

To calculate the noise-floor of the bolus signals, the amplitude histogram of both AP and SI channels of the expanded signal were first computed (Figure 5.2). After removal of head motions and vocalizations, the remaining noise will generally be of low energy. We thus estimate the range of the noise signal as \( \alpha \times 2\sigma \), where \( \sigma \) is initially the bolus signal standard deviation \( \sigma = \sqrt{1/N \sum (x - \mu)^2} \) and \( \alpha \) is a scalar multiplier. This expression provides an estimate of the range of the noise (i.e., assuming that the noise resided within \( \mu + 2\alpha\sigma \) and \( \mu - 2\alpha\sigma \)). The axial thresholds are then determined as:

\[
T^{AP} = \alpha \times 2\sigma^{AP}, \quad \text{and} \quad T^{SI} = \alpha \times 2\sigma^{SI}
\]

(5.2)

To estimate the optimum values for \( \alpha^{AP} \) and \( \alpha^{SI} \), the following criterion function was
Figure 5.2: Noise-floor annotated AP and SI signals of a bolus. The signal portion that is above the noise-floor threshold is marked in light green.

considered:

\[ J(\alpha) = \sum_{i=1}^{2} \left( [\delta_i' - \delta_i(\alpha)]^2 \times (1 - \beta)^{(-1)^{i-1}\text{sign}(\delta_i' - \delta_i(\alpha))} \right) \]  

(5.3)

where \( \delta_1' \) and \( \delta_2' \) are the new estimated bolus onset and offset, respectively, and \( \delta_1 \) and \( \delta_2 \) are the VFSS onset and offset respectively, expressed as a function of the threshold scalar \( \alpha \). The parameter \( 0 \leq \beta < 1 \) is used to tune the objective function. Larger values of \( \beta \)
yield more liberal estimates of onsets and offsets, i.e., further away from VFSS values, whereas smaller values of $\beta$ provides more conservative estimates. The optimal scalar is given by:

$$\alpha^* = \underset{\alpha}{\arg \min} J(\alpha)$$

(5.4)

The optimal value of $\alpha$ for the data set under consideration was determined via leave-one-out cross-validation with different values of $\beta$. The differences between predicted values of bolus onsets and offsets and those determined via VFSS were minimized with $\alpha = 0.81$. For this optimal $\alpha$, Figure 5.3 depicts the objective function values at different values of $\beta$. As seen in Figure 5.3, a $\beta$ of 0.35 provided an objective function that yielded the lowest error (i.e., boluses closest in length to those annotated by VFSS) in the neighborhood of the optimal $\alpha$ value. Once $\alpha$ and $\beta$ were optimized, those values were used in the bolus length estimation algorithm described above in classifier evaluation, i.e., to predict bolus lengths for each training and testing case.

### 5.5.4 Instance Selection

To reduce the effect of noisy instances on classification, we first attempted a filter approach to instance selection and subsequently proposed a posterior probability-guided wrapper approach.

**Multidimensional Feature-based Interquartile-range**

In this section, a simple multidimensional feature-based interquartile-range filter is proposed for instance selection. The 10 most salient features were considered. Let $J$ represent the dimensionality of the feature space and $N$ the total number of instances. Let $b_i = [f_{i,1}, f_{i,2}, ..., f_{i,J}]$ denote a single $J$-dimensional feature vector corresponding to the $i^{th}$ bolus. Let $Q_1 = [Q_{11}, Q_{12}, ..., Q_{1J}]$ and $Q_3 = [Q_{31}, Q_{32}, ..., Q_{3J}]$ be the lower and upper interquartile values, respectively, for the $J$ features. Let $\textbf{IQR} = [IQR_1, IQR_2, ..., IQR_J]$
Figure 5.3: Scalar analysis over different object function penalty values ($\beta$). The vertical line denotes the optimal value of $\alpha$.

denote the interquartile ranges of the $J$ features.

The set of $J$-dimensional excluded instances, $\Theta$, is then defined by

$$\Theta = \{\forall i, \ b_i \mid b_i < Q_1 - \delta \times \text{IQR}$$
$$\vee \ b_i > Q_3 + \delta \times \text{IQR},$$
$$1 \leq i \leq N\}$$

(5.5)

where $\delta = 1.5$ in the classical definition of outlying cases.

**Classification Probability Threshold Band**

An alternative, wrapper-based approach to instance selection is to deploy the classification posterior threshold in a selection criterion. A receiver operating characteristic (ROC) curve was calculated using the posteriors of the training data set where each point on
this curve results defines a sensitivity and specificity pairing. To account for class imbalance (in this case, minority positive class) the classification posterior threshold was tuned, using only the training set in each cross-validation run, to maximize sensitivity while maintaining 60% classification specificity [153].

![Figure 5.4: Instance selection on the basis of proximity to the posterior classification probability threshold.](image)

In this approach, an instance was selected for removal, if the corresponding classification posterior fell within the vicinity of the tuned threshold. The reasoning is that the uncertainty in the classifier’s decision is maximal at the decision threshold and decreases as posterior values depart from the threshold, either increasing in value towards unity or decreasing in value towards zero. In order to limit the number of selected instances, a marginal window was set (Figure 5.4). After tuning a classification posterior threshold in each cross-validation run, a probability window of size 0.02 centred around the threshold was considered. The size of this window was then incremented by 0.01 in each direction (above and below the threshold), admitting more instances while not exceeding a selection cap of 5%. This margin along with the tuned threshold was then applied to the test
data set. In other words, instances that met the following condition were selected for removal.

\[ \hat{T} - \delta \hat{T} < P(C(x)|X = x) < \hat{T} + \delta \hat{T} \]

where \( \hat{T} \) is the tuned threshold, \( \delta \) is the margin based on the instance removal cap, \( P(C(x)|X = x) \) is the posterior probability of instance \( x \) and \( C(x) = \{ \text{`safe', `unsafe'} \} \) is the bolus target class label.

### 5.5.5 Classification and Evaluation

A Linear Discriminant Analysis (LDA) classifier was evaluated over 1000 runs of a random hold-out cross-validation test. The entire data set was randomly divided into training and test participants (80% and 20%, respectively) in each run and the cross-validation runs were completely independent. In each run, a classifier was trained, using only the boluses of the training participants and then tested using the remaining 20% of the participants that were held out. It is important to note that the training and test data sets were selected at participant level, such that the test data set did not contain any boluses from the participants whose data were selected as part of the training data set. Moreover, the classifiers in each run were oblivious to the test and training sets of other runs. Classification performance was assessed in terms of sensitivity, specificity, and area under curve (AUC) across the cross-validation runs. Incidentally, artificial neural network (ANN) and support vector machine (SVM) classifiers were also trained but did not demonstrate any added value in terms of the above classification metrics.

### 5.6 Results

Using the noise-floor bolus length estimation algorithm with the scalar \((\alpha)\) value of 0.81, the performance of the classification system remained unchanged when compared
to classification based on VFSS-demarcated boluses. Figure 5.5 shows that there is no systematic bias in the length of the boluses before and after application of the noise floor bolus length estimation algorithm using the scalar ($\alpha$) value of 0.81 ($p = 0.36$, Kolmogorov-Smirnoff test). A kernel density estimate of the VFSS bolus lengths provided the null hypothesis cumulative distribution function against which each distribution of bolus lengths for a given $\alpha$ were tested using the Kolmogorov-Smirnoff goodness-of-fit test.

Table 5.3 compares classification performance with and without the different instance selection algorithms after 1000 runs of hold-out cross-validation. As shown, a maximum AUC of 83.6% was achieved for the discrimination of safe and unsafe boluses of thin consistency. There is an improvement in AUC ($p < 0.001$, Wilcoxon rank sum test) over the no instance selection case when applying the threshold band algorithm with either a 5 or 10% removal cap. This is further elucidated in Figure 5.6 where the notches of the box plots for 5 and 10% instance removal do not intersect the notches of the boxplot of the default case.
Table 5.3: Comparison of the classification performance using the proposed instance selection approaches.

<table>
<thead>
<tr>
<th>Instance selection (max % of instances removed)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>87 ± 8.7</td>
<td>60.2 ± 4.8</td>
<td>82.4 ± 5.7</td>
</tr>
<tr>
<td>Multidimensional IQR (5%)</td>
<td>82.8 ± 9</td>
<td>60.5 ± 5.3</td>
<td>78.9 ± 6.1</td>
</tr>
<tr>
<td>threshold band (3%)</td>
<td>88.2 ± 8.4</td>
<td>60.6 ± 4.9</td>
<td>82.8 ± 5.8</td>
</tr>
<tr>
<td>threshold band (5%)</td>
<td>88.9 ± 8.3</td>
<td>60.9 ± 5.1</td>
<td>83.2 ± 5.8</td>
</tr>
<tr>
<td>threshold band (10%)</td>
<td>90.9 ± 7.7</td>
<td>61.1 ± 5.4</td>
<td>83.6 ± 5.7</td>
</tr>
</tbody>
</table>

5.7 Discussion

This chapter has introduced bolus length estimation and instance selection as new elements to swallowing accelerometry classification. The former estimates the onset and offset of the bolus signals, based on the noise-floor distribution of both the AP and SI channels, and hence reduces classifier dependency on VFSS-based annotation. This reduced reliance on manual segmentation sets the stage for the development of a standalone, practical device for assessing swallowing safety. Instance selection, on the other hand, objectively identifies instances that diminish classification performance. The aforementioned classification framework achieves improved bolus-level AUC.

5.7.1 Reduced Dependence on VFSS-based determination of bolus of interest

As shown on Figure 5.5, larger values of the noise-floor bolus length estimation algorithm scalar (\(\alpha\)) forces the algorithm to estimate shorter bolus lengths. Smaller values of \(\alpha\) on the other hand yields longer boluses. An optimal value of \(\alpha\), which is achieved by minimizing the objective function given in Equation 5.2, produced bolus lengths closest to those obtained via VFSS, while maintaining classification performance.

By reducing the dependency on VFSS annotations, a standalone system can eventually be achieved. The addition of the cushion to the beginning and end of the bolus
mimics the demarcations one might obtain from operator button presses to bookmark the swallowing activity pertaining to each bolus. The proposed noise floor algorithm then provides an estimate of the bolus boundaries that one might obtain from VFSS review. To our knowledge, all previous swallowing accelerometry studies performed feature calculation, analysis, and classification on the basis of VFSS-demarcated signals, which precludes those algorithms from direct implementation into an independent swallow monitoring system.
5.7.2 Value of Instance Selection

The multidimensional feature-based interquartile-range approach to instance selection discarded boluses with extreme feature values. Since the extreme data points had a defining role in classification training and performance, this approach, although commonly used in the literature, failed to increase the performance of the classifier.

Instance selection using the classification probability threshold band, on the other hand, demonstrated very promising results. This approach leveraged classifier uncertainty as expressed through posterior probabilities. In the cases where the classification probability of the data points were close to the tuned threshold, there was uncertainty in the discrimination between the two classes. By removing instances within the uncertain band enveloping the tuned threshold, the overall performance of the classification algorithm increased significantly, even when only a modest fraction of instances were discarded (5-10%).

5.7.3 Exploration of Removed Cases

In this section, we investigate the selected instances for the case of the 5% removal cap.

Although the classification posterior of the selected instances were marginal (i.e., close to the decision boundary), the feature values of these instances where interior to the feature clusters. Figure 5.7 shows the first two components (derived using PCA) of these instances. As shown, the majority of the selected instances reside inside the class clusters. Figure 5.8 illustrates the parallel coordinate plot of the 10 salient features for the selected instances, again corroborating the observation that the selected instances are interior to the feature clusters rather than outlying observations.

The origin of the selected instances was as follows: 28.6% were drawn from unhealthy participants while 71.4% came from healthy participants. Note that the original data set was imbalanced with 17.6% and 82.4% of unhealthy and healthy participants, respectively. Despite this class imbalance, the instance selection algorithm disproportionately
oversampled the unhealthy participants, suggesting a tendency for indeterminate cases to stem from unhealthy participants. In the original data set, 7% of boluses were unsafe while 92.9% were safe. Of the instances identified by the probability threshold band instance selection algorithm, 4.9% were unsafe boluses and 95.1% were safe boluses. Considering the algorithm’s oversampling of unhealthy participants, this latter finding indicates that many safe boluses of unhealthy participants were selected as uncertain. Additionally, 3.4% of the total unsafe boluses and 5.1% of the total safe boluses were selected as uncertain instances. This further emphasizes that most of the selected instances were safe and potentially from unhealthy participants. These safe but uncertain boluses may possess characteristics that are very different from the safe boluses of the healthy participants.
To further investigate the selected instances, we performed a 5-fold cross-validation classification between the selected instances and the remaining (unselected) cases. We found that the selected instances could be discriminated from the rest of the data set with a high accuracy of 98%. This finding confirms that the selected instances exhibit very different signal characteristics from the rest of the data set.

Additionally, the majority of the selected instances were collected from 3 sites (31.52%, 22.42%, and 22.42% of instances from sites 1, 4, and 7, respectively). Further investigation of site-specific protocol compliance, as well as inter- and intra-participant variation may provide additional insight into the tendency of uncertain cases to originate from these 3 data collection sites.
5.7.4 Classification Performance

The safe and unsafe bolus-level classification performance achieved in this study is competitive when considering clinical detection rates reported in the literature. According to a recent study, sensitivity and specificity of clinical evaluations are reported to be 39% and 80% respectively, for penetration and 55.6% and 80.5% for aspiration [154]. In other studies, detection sensitivity and specificity have been cited as 88 ± 8% and 50 ± 13, respectively [155] and 93 ± 21% and 56 ± 20% [156].

5.8 Conclusion

Bolus length estimation and instance selection were introduced as enhancements to swallowing accelerometry classification, on one-hand liberating classification algorithms from manual segmentation of swallows and secondly affording the classifier the freedom to abstain from a decision in the face of uncertainty. We demonstrated that together these enhancements lead to an improvement in AUC in the discrimination between safe and unsafe swallows in a sizable clinical data set.

5.9 Acknowledgement

The authors would like to thank the Walter C. Summer Memorial Fellowship, and Ewing Rae Graduate Scholarship for providing the generous funding of this project.
Chapter 6

Conclusion

6.1 Overall Framework

Figure 6.1 depicts the overall framework of the proposed system. This framework contains three main modules that address the proposed research questions: (1) segmentation false positive reduction and adaptive trimming of the automatically segmented swallow signals (Chapter 3), (2) discrimination of cough versus non-cough accelerometry artefacts (Chapter 4), (3) bolus length estimation and instance selection to enhance the automatic classification of safe and unsafe swallowing accelerometry signals (Chapter 5).

6.2 Summary of Contributions

This thesis makes several original contributions to swallowing accelerometry research. The major contributions of the thesis are listed below. For reference, the corresponding research question (RQ) and chapter number for each contribution are indicated parenthetically.

1. Developed an algorithm that reduces the number of falsely detected swallows within dual-axes accelerometry recordings with a high degree of agreement with clinical segmentation via VFSS (RQ1; Chapter 3).
2. Developed an algorithm that trims the automatically segmented swallowing activities to remove pre- and post-swallowing signals (RQ1; Chapter 3).

3. Developed a robust algorithm for the discrimination between cough and non-cough artefacts for both instructed and reflexive coughs (RQ2; Chapter 4).

4. Reduced classifier dependency on the bolus onset and offset timestamps determined via VFSS by developing an algorithm to estimate bolus length with a high degree of agreement with clinical segmentation via VFSS (RQ3; Chapter 5).

5. Identified and removed uncertain boluses through posterior probability instance selection to increase the overall performance of automatic classification system (RQ4; Chapter 5).

Figure 6.1: Overall framework.
6. Integrated the above developments into a classification system that leverages both bolus and swallow level features to provide high sensitivity and specificity discrimination between safe and unsafe swallowing in patients living with dysphagia (RQ4).

A unifying theme underlying many of the aforementioned contributions (trimming, bolus length estimation, False Positive (FP) reduction, and classification) is the adaptive estimation of algorithmic thresholds. In other words, thresholds are estimated on a bolus by bolus basis, rather than on a global (entire data set) basis. The latter would yield static universal thresholds that are specific to the present problem domain. In contrast, the data-centric approach facilitates broader application of the proposed algorithms to different signals. Additionally, the present contributions collectively contribute to a higher level of automation than that reported with previous swallowing accelerometry classifiers.

6.3 Future Work

Several suggestions for future work can be considered to further improve the performance of swallowing accelerometry signal classification. The following areas may be of interest for future studies.

The manual segmentation of the proposed cough discrimination algorithm isolates this module from the rest of the framework. An accurate automatic cough segmentation algorithm that can achieve high correlation with the existing manual demarcations may facilitate the seamless integration of the cough module with other modules within the framework.

Moreover, additional cough features may be introduced to the feature set. These features may include: number of identified coughs within each bolus, the proximity of coughs to their preceding swallow segment, duration of detected coughs, and peak amplitude of the cough segments, to name a few.
Considering the high accuracy achieved in discriminating between selected and non-selected instances in chapter 5, future research might entertain the notion of an instance selection classifier. By integrating such a classifier into the overall framework, boluses can then be automatically classified as rejected or acceptable. Only acceptable boluses would proceed to safety classification. In a practical dysphagia detection system, the rejection of a bolus (by the instance selection classifier) would conceivably trigger a warning (i.e., that the previously recorded sample was unusable) and thereby alert clinicians or care-givers to administer another bolus.

Finally, sensor placement is another suggested study for future work. Although there is currently one published study on the effect of sensor placement on signal characteristics [124], research on the specific effect of sensor positioning on classification is necessary. Eventually, a feedback system may be beneficial to guide the optimal placement of the sensor.

6.4 Publications

6.4.1 Journal Articles

The following manuscripts comprise the majority of this thesis.


swallowing accelerometry signals with instance selection and bolus length estimation. to be submitted to Medical Engineering and Physics after industry partner approval.

The following secondary publication is not directly related to the thesis topic but was completed during the course of this PhD.


### 6.4.2 Conference Presentations

I delivered the following podium presentations during the course of my doctoral studies.


Additionally, I was a co-author on the following conference paper.

6.4.3 Patents

The following patents were filed during my doctoral studies, based on intellectual property developed as part of this thesis.


Appendix A

Accelerometry Signal Feature Calculations

This appendix contains feature calculations of dual-axis cervical accelerometry signals.

The following temporal [69, 124], time-frequency [40, 56, 69, 124], frequency [124], information-theoretic [124] features were calculated from the AP and SI axes separately. Additionally, we computed the mutual information, cross-entropy rate, and cross-correlation between the corresponding AP and SI signals [69, 124, 125].

- **Time Domain**: duration, mean, mean absolute deviation, median, variance, skewness, kurtosis, memory, maximum energy
- **Time-Frequency Domain**: wave energy
- **Frequency Domain**: peak frequency, bandwidth, Lempel-Ziv complexity, centroid frequency
- **Information Theoretic Domain**: entropy, entropy rate
- **AP and SI Combination**: mutual information, cross-entropy rate, and cross-correlation

Let $X$ be either the AP or SI axis of the cervical accelerometry signal, such that $X = \{x_1, x_2, ..., x_n\}$, where $n$ is the noise-floor estimated duration or the number of samples of $X$ (discussed in Chapter 5). Feature calculations are described in the following sections in detail.
A.1 Time Domain Features

A.1.1 Mean

The sample mean of the time series \( X \) is the average amplitude of the signal samples and can be calculated as

\[
\mu_X = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (A.1)
\]

A.1.2 Mean Absolute Deviation

Mean absolute deviation of a signal is the average distance between each data sample and the mean of the time series and is calculated as

\[
mad_X = \frac{1}{n} \sum_{i=1}^{n} |x_i - \mu_X| \quad (A.2)
\]

A.1.3 Median

The median is a robust location estimate of the amplitude distribution. For the sorted signal \( \hat{X} \), the median can be calculated as

\[
\text{med}_X = \begin{cases} 
  x_{m+1}, & \text{if } n = 2m + 1 \\
  \frac{x_m + x_{m+1}}{2}, & \text{if } n = 2m
\end{cases} \quad (A.3)
\]

A.1.4 Variance

The variance of a time series signal is a reflection of the signal power. It measures the signal spread around the mean and can be measured as

\[
\sigma^2_X = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \mu_X)^2 \quad (A.4)
\]
A.1.5 Skewness

The skewness of a distribution measures the symmetry of a distribution. Skewness is computed as

\[ \gamma_X = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_X)^3}{\left( \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_X)^2 \right)^{1.5}} \]  
(A.5)

A.1.6 Kurtosis

Kurtosis measures the peakedness of a distribution and can be calculated as

\[ \zeta_X = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_X)^4}{\left( \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_X)^2 \right)^2} \]  
(A.6)

A.1.7 Memory

The temporal extent of the correlation between the neighbouring data samples is quantified by the memory of the signal. The autocorrelation of the signal from zero to the maximum time lag (length of the signal) was calculated and normalized, such that the autocorrelation at zero lag was unity. The memory was then measured by the time duration from zero lag to the sample where autocorrelation decays to \( 1/e \approx 0.3679 \) or less [125]. This threshold has been used in previous studies for estimation of dual-axis accelerometry signal memory [69, 125].
A.1.8 Maximum Energy

Maximum energy is the maximum windowed energy calculated within a moving window of size 500 samples along the signal.

\[ E_X = \max \left( \sum_{i=1+(j-1)s}^{w+(j-1)s} (x_i)^2 \right), \quad 1 \leq j \leq \left\lfloor \frac{L_X - w + s}{s} \right\rfloor \]  

(A.7)

where \( w \) is the window size (500 samples), \( L_X \) is the length of the signal \( X \), and \( s = 50 \) is the moving step (window is moving with 90% overlap).

A.2 Time-Frequency Domain Features

A.2.1 Wavelet Energy

The energy in the approximation signal is calculated after wavelet decomposition based on 10-level discrete wavelet decomposition of the signal using the discrete Mayer wavelet [125]. The decomposition is defined as \( W_X = [a_{10}, d_{10}, d_9, ..., d_1] \), where \( a_{10} \) is the 10th-level approximation signal and \( d_{10}, d_9, ..., d_1 \) are the detail signals. The wavelet energy in \( a_{10} \) is measured as

\[ E_{a_{10}} = ||a_{10}||^2 \]  

(A.8)

where || • || is the Euclidean norm.

Similarly, the \( k^{th} \)-level detail signal energy is calculated as

\[ E_{d_k} = ||d_k||^2, \quad k = 1, 2, ..., 10 \]  

(A.9)
The wavelet energy is then calculated as

\[ \hat{E} = E_{a10} + \sum_{K=1}^{10} E_{dk} \] (A.10)

### A.3 Frequency Domain Features

#### A.3.1 Peak Frequency

The peak frequency is the peak magnitude of the Fast Fourier Transform (FFT) of the signal which is associated with the maximum spectral power and is determined by

\[ f_p = \arg\max_{f \in [0, f_{max}]} |F_X(f)|^2 \] (A.11)

where \( F_X(f) \) is the Fourier transform of the signal and \( f_{max} \) is the Nyquist frequency (effectively 500 Hz after down-sampling).

#### A.3.2 Centroid Frequency

The spectral centroid of the signal is calculated by

\[ \hat{f} = \frac{\int_0^{f_{max}} f |F_X(f)|^2 df}{\int_0^{f_{max}} |F_X(f)|^2 df} \] (A.12)

#### A.3.3 Bandwidth

The bandwidth of the spectrum is calculated by

\[ BW = \sqrt{\frac{\int_0^{f_{max}} (f - \hat{f})^2 |F_X(f)|^2 df}{\int_0^{f_{max}} |F_X(f)|^2 df}} \] (A.13)
A.3.4 Lempel-Ziv Complexity

The predictability of the signal is measured by the Lempel-Ziv complexity [157]. The signal $X$ is first quantized into 100 equally spaced levels between the signal minimum and maximum amplitudes, where $\hat{X} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_n\}$ is the quantized signal of length $n$. Then $\hat{X}$ is decomposed into $k$ blocks, such that $\hat{X}_1^n = \{B_1, B_2, ..., B_k\}$. The notion $\hat{X}_i^j = \{\hat{x}_i, \hat{x}_{i+1}, ..., \hat{x}_j\}, 1 \leq i, j \leq n$ denote consecutive samples of $\hat{X}$ from $i$ to $j$.

The first block is initialized to the first symbol (i.e. $B_1 = \hat{X}_1^1 = \hat{x}_1$). The subsequent blocks are defined as

$$B_{m+1} = \hat{X}_h^{h_m+1}, m \geq 1$$ (A.14)

where $h_m$ is the ending index for block $B_m$. The block $B_{m+1}$ is a previously unrepeated sequence of minimal length in the sequence $\hat{X}_1^{h_m+1}$.

To measure the Lempel-Ziv complexity, the normalized complexity is computed as

$$C = \frac{k \log_{100} n}{n}$$ (A.15)

Logarithmic base 100 is selected since the signal was initially quantized into 100 quantization levels.

A.4 Information Theoretic Domain Features

A.4.1 Entropy

The signal entropy reflects the amount of uncertainty or information within the signal [158]. After signal standardization to zero mean and unit variance, the entropy is measure
by

\[ En = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \] (A.16)

where \( p(x_i) \) is the probability that \( X = x_i \).

### A.4.2 Entropy Rate

The extent of regularity in a signal is measured by the entropy rate [158]. This feature is particularly useful for characterizing a stochastic process where consecutive data points may exhibit some form of relationship.

The signal is quantized into 10 equally spaced levels after normalizing to zero mean and unit variance (i.e. \( \hat{X} = \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_n\} \)). Consequently, segments of size \( L \), containing consecutive points, were coded into a series of integers, such that

\[ \Omega_L = \{\omega_1, \omega_2, ..., \omega_{n-L+1}\} \] (A.17)

where \( \omega_i = \sum_{d=0}^{L-1} (\hat{x}_{i+d} \times 10^d) \). Logarithmic base 10 was used since initially 10 quantization levels were considered. The Shannon entropy was calculated for \( \Omega_L \) as

\[ SE(L) = \sum_{j=0}^{10L-1} p_{\Omega_L}(j) \log_2 p_{\Omega_L}(j) \] (A.18)

where \( p_{\Omega_L}(j) \) is the probability of value \( j \) in \( \Omega_L \), measured by the frequency of the corresponding sample. The entropy was then normalized as

\[ \tilde{SE}(L) = \frac{SE(L) - SE(L-1) + SE(1) \times p'}{SE(1)} \] (A.19)

where \( p' \) is the percentage of the unique occurrences of coded integers in \( \Omega_L \) (i.e. the coded integers that occurred only once).
Consequently, the entropy rate is calculated as the index of regularity as

$$\rho = 1 - \min(\hat{SE}(L)) \quad (A.20)$$

The entropy rate ranges from 0 to 1, indicating maximum randomness to maximum regularity, respectively.

### A.5 AP and SI Combination Features

#### A.5.1 Mutual Information

Mutual information measures the amount of information shared between two signals. Mutual information of value zero indicates that the vibrations in one signal contains no information about the vibrations in the other signal when accelerometry signals are examined. Mutual information is estimated by

$$MI = \sum_i \sum_j p(x_{i}^{AP}, x_{j}^{SI}) \log_2 \frac{p(x_{i}^{AP}, x_{j}^{SI})}{p(x_{i}^{AP})p(x_{j}^{SI})} \quad (A.21)$$

where $p(x_{i}^{AP})$ and $p(x_{j}^{SI})$ are the probability that $X_{AP} = x_{i}^{AP}$ and $X_{SI} = x_{j}^{SI}$ respectively, and $p(x_{i}^{AP}, x_{j}^{SI})$ is the joint probability distribution function of $X^{AP}$ and $X^{SI}$.

#### A.5.2 Cross-Entropy Rate

Cross-entropy rate characterizes the predictability of a data sample in a signal, given a sequence of the current and past data samples of another signal. Inspired by the entropy rate estimation above, the cross-entropy rate between two signals can be measured [159, 160].

After normalizing, quantizing, and coding both signals using similar methodology as described earlier in the entropy rate section, $\hat{X}^{AP} = \{\hat{x}_{1}^{AP}, \hat{x}_{2}^{AP}, ..., \hat{x}_{n}^{AP}\}$ and $\hat{X}^{SI} = \{\hat{x}_{1}^{SI}, \hat{x}_{2}^{SI}, ..., \hat{x}_{n}^{SI}\}$.
\{\hat{x}_{1}^{SI}, \hat{x}_{2}^{SI}, \ldots, \hat{x}_{n}^{SI}\} \text{ are obtained and } \Omega_{L}^{X^{AP}} \text{ and } \Omega_{L}^{X^{SI}} \text{ are measured. Additionally, } \Omega_{L}^{X^{AP}|X^{SI}} = \{\omega_{1}^{X^{AP}|X^{SI}}, \omega_{2}^{X^{AP}|X^{SI}}, \ldots, \omega_{n-L+2}^{X^{AP}|X^{SI}}\} \text{ is calculated as}

\[ \omega_{i}^{X^{AP}|X^{SI}} = \sum_{d=0}^{L-1} \left( \hat{x}_{i+d}^{AP} \times 10^{d} + \hat{x}_{i+d}^{SI} \times 10^{d} \right) \] (A.22)

Consequently, the Shannon entropies \( SE_{X^{AP}|X^{SI}}(L), SE_{X^{SI}}(L), \) and \( SE_{X^{AP}|X^{SI}}(L) \) of \( \Omega_{L}^{X^{AP}}, \Omega_{L}^{X^{SI}}, \) and \( \Omega_{L}^{X^{AP}|X^{SI}} \), are measured, respectively (Equation A.18).

The cross-entropy of \( X^{AP} \) given \( X^{SI} \) is normalized as

\[ \tilde{SE}_{X^{AP}|X^{SI}}(L) = SE_{X^{AP}|X^{SI}}(L) - SE_{X^{SI}}(L-1) + SE_{X^{AP}}(1) \times p'_{X^{AP}|X^{SI}}(L) \] (A.23)

where \( p'_{X^{AP}|X^{SI}}(L) \) is the percentage of the unique occurrences of coded integers in \( \Omega_{L}^{X^{AP}|X^{SI}} \).

Finally, the cross-entropy rate is measured as

\[ \Gamma_{X^{AP}|X^{SI}} = 1 - \min(Z_{X^{AP}|X^{SI}}(L)) \] (A.24)

where \( Z_{X^{AP}|X^{SI}}(L) = \min(\tilde{SE}_{X^{AP}|X^{SI}}(L), \tilde{SE}_{X^{SI}|X^{AP}}(L)) \) is the uncoupling function.

A cross-entropy rate of value 1 indicates high synchronicity between two signals, while a value 0 suggests that the two signals are completely uncoupled.

### A.5.3 Cross-Correlation

Similarity between two signals is measured by the cross-correlation feature. The cross-correlation at zero-lag is calculated by

\[ \hat{R}_{X^{AP},X^{SI}}(0) = \frac{1}{n} \sum_{i=1}^{n} (x_{i}^{AP} x_{i}^{SI}^{SI}) \] (A.25)
Bibliography


[31] Miller, Arthur J. Characteristics of the swallowing reflex induced by peripheral

[32] Steele, Catriona M and Van Lieshout, Pascal. Tongue movements during water
swallowing in healthy young and older adults. *Journal of Speech, Language, and

[33] Shaw, Stephanie M and Martino, Rosemary. The normal swallow: muscular and
956, 2013.

[34] Spiro, Jeffrey, Rendell, Jill K, and Gay, Thomas. Activation and coordina-
tion patterns of the suprahyoid muscles during swallowing. *The Laryngoscope*,


[37] Peladeau-Pigeon, Melanie and Steele, Catriona. Understanding image resolution
and quality in videofluoroscopy. *SIG 13 Perspectives on Swallowing and Swallowing

[38] Swigert, Nancy B. Update on current assessment practices for dysphagia. *Topics

[39] Tabaee, Abtin, Johnson, Paul E, Gartner, Carolyn J, Kalwerisky, Kevin,
Desloge, Rosemary B, and Stewart, Michael G. Patient-controlled comparison
of flexible endoscopic evaluation of swallowing with sensory testing (feesst) and


[78] Willett, LL, Barney, J, Saylors, G, and Dransfield, M. An unusual cause of chronic


[132] Birring, SS, Fleming, T, Matos, S, Raj, AA, Evans, DH, and Pavord, ID. The
Further details of these applications can be found in the references listed:


