CERVICAL AUSCULTATION FOR THE IDENTIFICATION OF SWALLOWING DIFFICULTIES

by

Joshua M Dudik

B.S. in Biomedical Engineering,

Case Western Reserve University, 2011

M.S. in Bioengineering, University of Pittsburgh, 2013

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Joshua M Dudik

It was defended on

November 5, 2015

and approved by

Ervin Sejdic, PhD, Assistant Professor, Department of Electrical and Computer Engineering

Zhi-Hong Mao, PhD, Associate Professor, Department of Electrical and Computer Engineering

Amro El-Jaroudi, PhD, Associate Professor, Department of Electrical and Computer Engineering

Mingui Sun, PhD, Professor of Neurological Surgery, Departments of Bioengineering and Electrical Engineering

James Coyle, PhD, Associate Professor, Department of Communication Science and Disorders

Dissertation Director: Ervin Sejdic, PhD, Assistant Professor, Department of Electrical and Computer Engineering

CERVICAL AUSCULTATION FOR THE IDENTIFICATION OF SWALLOWING DIFFICULTIES

Joshua M Dudik, PhD

University of Pittsburgh, 2015

Swallowing difficulties, commonly referred to as dysphagia, affect thousands of Americans every year. They have a multitude of causes, but in general they are known to increase the risk of aspiration when swallowing in addition to other physiological effects. Cervical auscultation has been recently applied to detect such difficulties non-invasively and various techniques for analysis and processing of the recorded signals have been proposed. We attempted to further this research in three key areas. First, we characterized swallows with regards to a multitude of time, frequency, and time-frequency features while paying special attention to the differences between swallows from healthy adults and safe dysphagic swallows as well as safe and unsafe dysphagic swallows. Second, we attempted to utilize deep belief networks in order to classify these states automatically and without the aid of a concurrent videofluoroscopic examination. Finally, we sought to improve some of the signal processing techniques used in this field. We both implemented the DBSCAN algorithm to better segment our physiological signals as well as applied the matched complex wavelet transform to cervical auscultation data in order to improve its quality for mathematical analysis.

Keywords: cervical auscultation, dysphagia, deep learning, signal analysis, signal features, classification

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 MOTIVATION

1.1.1 Definition

Dysphagia is a term that is used to refer to a multitude of swallowing difficulties and disorders [1]. Typically it is classified into three categories: oropharyngeal dysphagia for causes that originate in or near the patient's pharynx; esophageal dysphagia for causes that originate in the esophagus; and functional dysphagia for those where no cause can be located [2]. Dysphagia can present itself in many ways, including coughing after a swallow, pain during swallowing, or simply difficulty with initiating a swallow [1]. Often times, these symptoms are a sign that the muscles and anatomical structures involved in swallowing are not functioning correctly [1]. The structures that serve to protect the airway during a swallow are of particular interest. However, this airway protection is dependent on many factors including hyolaryngeal displacement and posterior movement of the tongue base, both of which serve to direct the epiglottis into the position necessary to protect the airway, as well as the elevation and anterior-directed tilting of the arytenoids [3]. If any part of this multi-faceted process does not operate correctly during a swallow then food can be allowed to enter the trachea, which can potentially lead to a pulmonary infection and more serious health outcomes [2]. Even if this particular scenario does not occur, dysphagia still typically makes swallowing uncomfortable for the patient [4, 5]. This affects their quality of life and can impact their nutrition as the patient attempts to avoid this unpleasant activity [4, 5].

1.1.2 Incidence and Prevalence

Dysphagia can occur in patients of any age, but is more common among the elderly [6,7]. An estimated 10% of patients over the age of fifty have been formally diagnosed with swallowing disorders [6]. Factoring in the idea that some people do not seek treatment for their condition and that many suffer from 'silent aspirations' and are not properly diagnosed, this number may be as high as 20% or more [6,8]. Regardless, approximately 10 million Americans are formally diagnosed with swallowing disorders every year [6].

Dysphagia is particularly common among people that are admitted to hospitals or medical care facilities. Estimates for the prevalence in these locations ranges from 25% in the general hospitalized population up to 75% in nursing homes or critical care facilities [6]. The numbers given are difficult to calculate with certainty because these populations are also more prone to experiencing 'silent aspirations' than the general population [6,9]. According to some estimates, as many as 80% of patients in critical care facilities have demonstrated this phenomenon [6,9].

Though it is not nearly as common, dysphagia can still occur in the younger populations, particularly infants. The most common source of non-oropharyngeal swallowing disorders in this situation is gastroesophageal reflux disease, which is serious enough to be classified as dysphagia in approximately 8% of infants [7]. It is fairly common for this condition to reverse itself through the course of normal development and ageing, but if ignored it can have a lasting impact on the patient [6]. As infants do not have the same ability to communicate their problems as an adult, this condition carries a noticeable risk of going undiagnosed [6,7].

1.2 ANATOMY AND PHYSIOLOGY

1.2.1 Stages of Deglutition

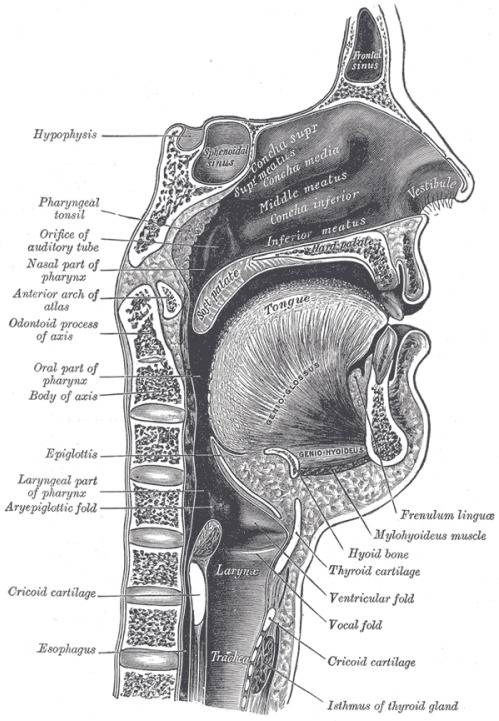


Figure 1: Lateral view of the human head and neck showing major physical structures [10]

A lateral cross-section of the human head and neck can be seen in figure 1. It clearly shows the different sections of the throat that will be referred to later as well as several other notable structures. Meanwhile, figure 2 displays the major muscles and arteries that influence swallowing activity.

The action of swallowing is divided into four stages. The first stage, the preparatory phase, encompasses the activity used to prepare the bolus for a swallow [11]. During this stage the bolus is placed on the dorsum of the tongue while the tip and posterior portions of the tongue prevent premature movement of the bolus [11]. The oral phase begins as the posterior portion of the tongue lowers and the soft palate rises to obstruct the nasopharynx [11,12]. During this stage, the bolus is propelled into the pharynx and hyolaryngeal elevation begins, thereby increasing the volume of the pharynx [11,12]. The third stage of swallowing

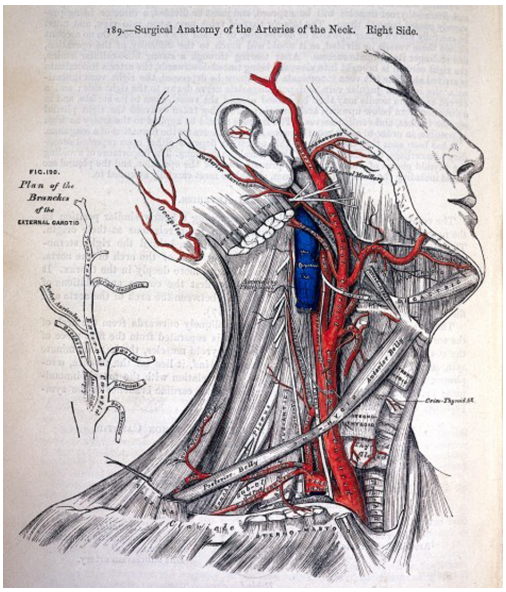


Figure 2: Gross muscle anatomy of the human neck [10].

is the pharyngeal stage [11, 12]. The most common definition of the beginning of this stage is the moment the bolus passes the palatoglossal arch and enters the oropharynx [11, 13]. During this phase, all passages that do not lead the bolus towards the stomach are sealed and actions such as breathing and coughing are inhibited [11,13,14]. The nasopharynx is blocked by the soft palate, the oral cavity is sealed by the tongue, and the larynx is blocked by the superior and anterior displacement of the hyolaryngeal complex which leads to epiglottic inversion to seal the inlet to the airway as well as by adduction of the vocal folds [11,13,15]. Peristalsis of the oropharynx and layngopharynx muscles then actively pushes the bolus past the upper esophageal sphincter [13]. The upper esophageal sphincter also opens during this stage, allowing the bolus a path into the esophagus [11, 13]. This behaviour is the result of multiple actions including relaxation of the sphincter muscles, anterior movement of the larynx and the resulting expansion of the sphincter, and the positive pressure produced by the bolus as it travels towards and through the sphincter opening [11]. The fourth and final stage is the esophageal stage [12]. This stage begins as the bolus passes the upper esophageal sphincter and continues until it reaches the stomach [14]. Peristalsis of the smooth muscles

within the esophagus ensure that the bolus is driven in the correct direction [14]. During this stage, the structures that shifted their position during the pharyngeal stage return to their resting positions and non-swallowing activities are able to resume [14].

1.2.2 Pathology

With 500,000 cases reported every year, neural damage or impairment serves as the most common cause of dysphagia [6]. Usually, it is abnormalities related to the sensory and motor functions mediated by the cranial nerves that cause swallowing disorders, as they directly control the muscles of the throat [4]. While head or neck related trauma can cause damage or paralysis, a stroke is far more often the underlying cause of dysphagia in this regard [4]. The incidence of dysphagia in patients with a stroke is quite high since damage to motor neurons or cranial nuclei can easily disrupt the motor commands that direct proper swallowing function via the cranial nerves [4]. Naturally, if not all of the muscles near the pharynx and larynx are receiving proper neural inputs, the patient will not be able to swallow correctly and may be unable to properly transfer a bolus past the larynx and into the esophagus [4, 16]. Neurological conditions, such as Huntington's or Parkinson's disease, also often result in the patient developing swallowing disorders. Though the nerves are not necessarily damaged it is clear that they and their corresponding muscles do not always function normally in these situations [6]. Often times the muscles related to swallowing will not be activated in the correct sequence or at the proper moments, which can result in issues such as the airway being unprotected as the bolus passes or the upper esophageal sphincter remaining closed as the bolus is pushed towards it [17]. Naturally, either of these issues could result in the patient aspirating and would likely cause a noticeable amount of discomfort [17].

In addition to neural sources, there are several more anatomically-related sources of dysphagia. Conditions that result in an inflamed and swollen esophagus, such as eosinophilic esophagitis or gastroesophageal reflux, can make it difficult for the patient to transfer a bolus into and down the esophagus [2]. This can often lead to the feeling of food becoming "stuck" in the throat if not actually cause the bolus to become trapped [2]. Depending on how the

timing of various muscle components is altered this may cause the airway to open while a bolus is still in the pharynx and put the patient at risk of aspirating [2]. Various abnormal growths, such as tumours, swollen lymph nodes, or esophageal webs, can extend into the digestive tract and obstruct the path of a bolus as well, leading to similar feelings of food becoming trapped in the throat and similar risks of aspiration [18]. Finally, damage to the muscles and structures of the throat can also result in swallowing difficulties [18]. A surgical procedure or radiation therapy used to manage a different medical condition can cause noticeable damage to the patient's body, in addition to other sources of physical trauma. Depending on the extent of damage that the muscles involved in deglutition sustain they may begin to behave incorrectly or too weakly, much like in the neurological case, and put the patient at risk of aspirating [18].

1.2.3 Methods of Swallowing Assessment

There are two overall categories that swallowing assessment techniques are sorted into. The first category, screening, consists of relatively simple pass-fail tests [19–21]. The goal of screening is to very generally characterize a patient's condition, often when the patient is not even expressing the symptoms of the disorder, in order to determine the need for a more thorough diagnostic exam or allow for early intervention [19–21]. The second category, a diagnostic test, attempts to fully characterize the patient's condition [19–21]. While a screen may only be able to detect the presence or absence of a condition, a diagnostic test is able to identify the exact mechanisms through which the condition is expressed along with the underlying causes [19–21]. However, due to the thorough nature of a diagnostic test, the procedure is often more complex and time-consuming than a simple screening.

The most widely accepted method of describing patterns of dysphagia is the videofluoroscopic diagnostic examination [1, 22]. During this test, the patient is asked to swallow small amounts of food or liquid that has been mixed with a contrast agent, typically barium sulphate [17, 22]. The x-ray equipment is aligned to produce a sagittal view of the oropharynx, pharynx, and upper esophagus which contain all of the major structures involved in swallowing activity [22]. With this set-up, a radiologist, speech-language pathologist, or another

specialist can observe the physiologic events that produce bolus movement in real time [22]. Depending on the path the bolus takes and the speed at which it travels, the examiner can determine which aspects of the patient's swallowing are not functioning properly and how significantly the patient aspirates if at all [22]. As all of these factors combine to form a comprehensive and accurate assessment of a subject's swallowing function, and have led to the widespread adoption of videofluoroscopy as a diagnostic test [1, 17, 22].

Though videofluoroscopy is currently the most common and widely accepted method of diagnosing dysphagia, it is not the only diagnostic technique used in a clinical setting. Fiberoptic endoscopy, for example, is also used to assess swallowing disorders in some situations [23–25]. This method is used in much the same way as videofluoroscopy. However, rather than use an x-ray imaging machine, a small camera is directed into the oropharynx or beyond which the speech-language pathologist uses to observe the swallow [23–25]. The advantage of this method is that the examiner can directly observe the patient's anatomy [23,24]. It also allows for examination of much smaller details as well as the colour of surrounding tissues, which can provide important diagnostic information [23, 24]. However, this method has two key drawbacks relative to a videofluoroscopy exam. The first is that only a small section of the anatomy is visible at one time, as the camera has a limited field of view and cannot observe the all parts of the subject's throat simultaneously [24, 26]. Furthermore, the areas that the camera can be placed during a swallow is somewhat limited, as endoscopy cannot meaningfully assess the action of the oral cavity or esophagus during a swallow, which further restricts the information provided by an exam [24]. Second, as the camera utilized for this method uses natural light, this technique cannot view the region under observation for the entire duration of the swallow [23, 24, 26]. When the bolus passes the camera and peristalsis of the relevant muscles begins, the camera's view is physically obstructed until the bolus passes and the muscles relax [23, 24, 26].

Two instrumental methods of screening for swallowing disorders that have been used in the clinical environment in the past are electromyography and cervical auscultation. Electromyography involves placing electrodes on the patient's neck and recording the electrical activity of the muscles during a swallow [27, 28]. The theory is that if the nerves or muscles that are involved in swallowing do not work correctly the signal will change in some

clinically significant way when compared to a recording from a healthy patient [27, 28]. While previously utilized both for swallowing assessment and as a biofeedback technique, electromyography can only indirectly describe a swallow since it is limited to monitoring muscle activation exclusively [27,28]. As a result, this technique remains mostly experimental. Cervical auscultation, on the other hand, records vibratory and acoustic correlates of the physiologic events occurring during a swallow [29]. A sound recording device, usually a stethoscope, is placed over the patient's thyroid cartilage and the examiner listens to the various noises produced as the patient performs a swallow [30]. The theory behind this, much like the electromyographic screening, is that the signal recorded from a patient with dysphagia will be significantly different than that recorded from a healthy patient [30]. In its current condition, this test is simple to perform, but its accuracy and reproducibility is questionable and its ability to identify specific physiologic events has not been established [29,30].

Non-instrumental methods of swallowing screening are highly varied. Methods such as the 3-oz water swallow test, the Toronto Bedside Swallowing Screening test, and the Standardized Swallowing Assessment among others have all been successfully implemented in the clinical setting [31–33]. While the exact details vary, they all operate via similar mechanics. The patient is asked to swallow a certain amount of liquid by a trained examiner, who then observes the patient for any physical sign of coughing, choking, or dysphonia after the swallow [31–33]. Some of these screening tests also take into account patient surveys or tongue and lip motility in order to provide an assessment of the patient [32–34]. Despite their widespread use in the clinic, non-instrumental screening methods have limited accuracy and often significantly over-estimate the rate of patients with dysphagia [33]. However, they are comparatively simple to administer and serve as a reasonable means to determine which patients require a full diagnostic exam [33].

1.3 DIRECTIONS AND GOALS

Considering how common dysphagia is in the general population, it is clear that early identification of this condition is of great importance. In some cases, however, it may not be possible to utilize a full diagnostic test to assess a particular patient's condition. As stated previously, dysphagia is common among patients with neurological disorders or those otherwise admitted to acute care facilities [4, 6, 9]. These patients may not have the physical or mental ability to complete a diagnostic procedure, such as videofluoroscopy, that requires a high level of patient compliance [24, 25]. Diagnostic tests also are not typically used for daily or repeated monitoring of a patient, whether it is because the test cannot be completed at a patient's bedside, repeated radiation exposure, or simply due to time and scheduling constraints. Furthermore, since they are intended to fully characterize a patient's disorder, diagnostic tests are not always the best choice if a quick decision must be made concerning a patient's level of risk or treatment. In all of these situations, it is more common to implement a swallowing screening test.

Common non-instrumental screening methods implemented in the clinical setting demonstrate some success, but have a number of drawbacks [33]. In particular, both the sensitivity and specificity of these tests leave room for improvement [31–34]. This is especially true when many of these tests expect the patient to physically react when they aspirate, despite silent aspirations occurring in up to 30% of patients with dysphagia [8]. Without an accurate screening, these patients may have proper treatment delayed or may be recommended for additional, unnecessary diagnostic procedures. In addition, these screening procedures are not always consistent under otherwise identical conditions [35, 36]. Depending on the screening procedure used and the particular examiner administering the test, components of a patient's condition may be weighted differently during the exam and the conclusions drawn may not be identical [35, 36]. This could cause delays in obtaining the proper care for an individual and may bias a follow-up diagnostic procedure.

Because of these drawbacks, an alternate or more precise method of screening for dysphagia would be beneficial to the general public. Ideally, such a technique would be non-invasive and use equipment that is both inexpensive and portable enough to be easily transported

to the patient's bedside rather than the opposite. Furthermore, such a technique should be easily digitized and automated so as to provide the examiner with an objective assessment of the likelihood that the patient has a condition warranting further objective evaluation. At the moment, we feel that cervical auscultation has the potential to serve as this new, alternate screening method. Microphones and accelerometers are well known for being able to transduce physical signals, such as those of interest for cervical auscultation, into digital form. Furthermore, their prevalence in modern society and the age of the technology has ensured the extremely low cost and small footprint of the associated hardware. In addition, the two-dimensional audio signals produced by these devices are much simpler to process and analyse than the four-dimensional signals produced through videofluoroscopy. Some preliminary research as well as the prevalent use of stethoscopes with existing, non-instrumental screening methods [33,35,36] has demonstrated that cervical auscultation can be successfully used to analyse swallowing signals, but research is far from complete. Some of the topics that still require investigation include finding the best methods to process and analyse the digital signals, correlating specific sounds or vibrations with physiological events, and determining how the recorded signals differ in healthy patients and those with dysphagia.

1.4 DISSERTATION SCOPE

One key area that requires further research is determining how various swallows differ. Work has been done with healthy swallows, investigating how a patient's gender or head position can affect cervical auscultation signals [37, 38]. However, the ways in which these variables affect swallows made by patients with dysphagia is still an open question. Likewise, the ways in which cervical auscultation signals differ between healthy subjects and patients with dysphagia as well as between safe and unsafe dysphagic swallows has not been fully explored. Some researchers have attempted to differentiate these groups in order to provide a simple, lower-cost screening method [39–48], but they did not always operate with a full understanding of how their signals differed. By fully characterizing the attributes of these swallows, we will be better equipped to differentiate them and improve existing cervical

auscultation-based screening techniques. Assuming that these difference exist, it would then be natural to expand upon this issue and develop the means to automatically classify these states.

In order to best analyse and differentiate swallows, however, we must first ensure that we are obtaining the proper data. Though some researchers have begun exploring various features of swallowing vibrations and sounds they still often rely on a videofluoroscopy examination in order to differentiate periods where swallows are or are not occurring. Several attempts have been made to perform this task automatically, but results have been mixed with data from dysphagic subjects proving particularly troublesome [49–54]. In addition, as a physiological event, swallowing signals contain high levels of noise and artefacts. Many researchers have developed their own techniques for filtering and processing this data, but none have been able to determine the best method. In summary, in order to best investigate the differences between healthy and unhealthy swallows we must first ensure that our data is processed correctly. Improving these filtering and segmentation techniques would be of great benefit to the field.

1.5 MAIN CONTRIBUTIONS

We hypothesize that it is possible to automatically identify swallows made by subjects with dysphagia as well as swallows that resulted in unsafe movement of swallowed material with the use of cervical auscultation. We further suggest that this process can be automated and implemented either alongside or independent of existing diagnostic techniques. In an attempt to accomplish this task, we have formulated several key topics in need of further investigation. These points are described below.

- Develop an algorithm to automatically identify periods of swallowing activity based solely on a recording of swallowing vibrations.
- Design FIR filters that are optimized for work with cervical auscultation and implement them with the complex wavelet transform to improve existing noise reduction methods.

- Characterize the differences between swallows made by healthy subjects and swallows made by patients with dysphagia that did not result in a significant amount of penetration into the larynx based on a broad selection of time, frequency, and time-frequency cervical auscultation features.
- Characterize the differences between swallows made by subjects with dysphagia that resulted in significant amounts of penetration into the larynx and those that did not based on a broad selection of time, frequency, and time-frequency cervical auscultation features.
- Develop an algorithm to automatically classify a swallow that did not result in a significant amount of laryngeal penetration as originating from a healthy person or a person suffering from dysphagia based on known distributions of cervical auscultation features.
- Develop an algorithm to automatically classify a swallow from a dysphagic subject as one that resulted in a significant amount of laryngeal penetration or not based on known distributions of cervical auscultation features.

1.6 DISSERTATION ORGANIZATION

Chapter 2 explores the cervical auscultation field, including the current hardware and signal processing standards. Chapter 3 provides an overview of the topics covered in this manuscript. Chapter 4 discusses our attempt to provide an improved method of automatically segmenting swallowing vibrations. Chapter 5 presents an improved method for denoising swallowing vibrations that uses a matched wavelet denoising technique. Chapter 6 characterizes both swallows from healthy adults and swallows that did not result in significant laryngeal penetration from adults with dysphagia and identifies the mathematical differences between their cervical auscultation signals. Chapter 7 offers similar material, but with regards to differentiating swallows which either did or did not result in significant amounts of laryngeal penetration, both made by adults with dysphagia. Chapters 8 and 9

are organized in a similar format to chapters 6 and 7, but instead present the use of deep belief networks to classify the relevant groups. Chapter 10 concludes our research and suggests possible avenues of research for future studies into this field.

2.0 BACKGROUND

The majority of this chapter has been previously published in and reprinted with permission from [55]. ©2015 IEEE. Dudik, J. M.; Coyle, J. L. & Sejdić, E.. Dysphagia Screening: Contributions of Cervical Auscultation Signals and Modern Signal Processing Techniques. *IEEE Transactions on Human-Machine Systems*, 2015, vol. 45, no. 4, 465-477. DOI: http://dx.doi.org/10.1109/THMS.2015.2408615

2.1 DEFINITION AND REASONS FOR INVESTIGATION

Auscultation is defined as using sounds to interpret the activity and function of the internal workings of the body. Cervical auscultation, then, is the application of this technique to the neck and throat and is most commonly used to monitor swallowing activity. This technique has been used for centuries since the invention of the stethoscope and has continued to this day. However, much work has been done recently to replace stethoscopes with digital microphones. At least in theory, a microphone is able to record the same signals as a stethoscope while providing several advantages. Since a microphone is a digital device, the analysis of sounds is no longer limited by the range of human hearing. Furthermore, the signal can be processed and analysed using various signal processing techniques which can extract more useful information than can be obtained by a human. Accelerometers have also been of interest in this field. Though they record vibrations rather than sounds they provide the same advantages as microphones can provide through digitization of the recorded signal. As vibrations and sounds are so closely linked on the physiological level it is logical to acknowledge such devices when discussing cervical auscultation.

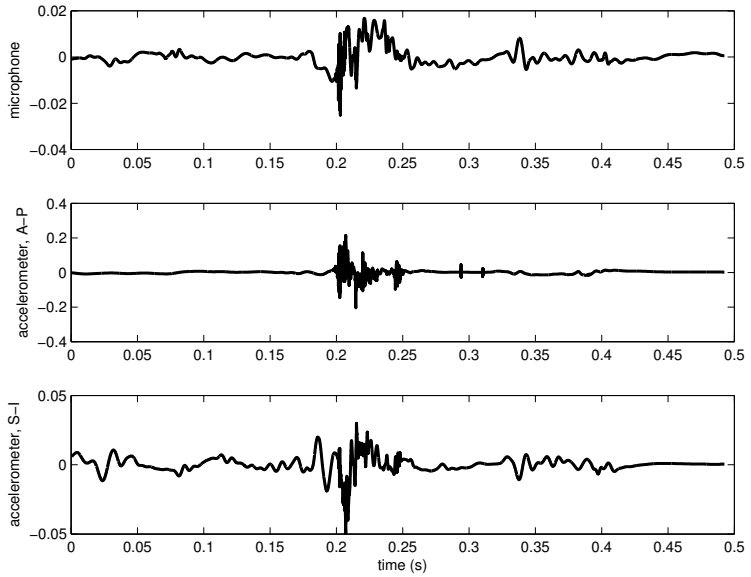
Cervical auscultation as a field of research has developed recently for a number of reasons. The most obvious is that, despite years of use, we still do not clearly understand the links between the action of swallowing and what is observed via a stethoscope. While some researchers have attempted to link 'notable clicks and pops' that occur during a swallow with specific physiological events their work has not been conclusive [56]. Because of this and the difficulty of reproducing these results, the ability of a human with a stethoscope to correctly identify swallowing difficulties is questionable. By implementing these digital transducers, it is possible to more closely investigate the link between physiological events and the recorded sounds and vibrations. One could then use that knowledge to improve the current bedside screening methods and determine if digitizing the process could offer further accuracy improvements. Second, the process of screening for swallowing-related aspiration, or the presence of foreign matter in the airway, has become ubiquitous in hospitals and other health care facilities. Because of the negative impact of aspiration in conditions like stroke, neurodegenerative diseases, and many other conditions, these screening procedures have been universally recognized as essential [32,57–59]. Significant reductions of health care associated pneumonia rates have been demonstrated in facilities instituting formal screening protocols [58] and using more objective screening tactics provides the potential to greatly reduce public health cost and human morbidity and mortality associated with swallowing disorders. Since cervical auscultation is extremely accessible, inexpensive, and easy to implement in the screening process, it has the potential to add needed objective data that could increase screening accuracy and early detection of risk. This outcome could provide significant savings to the health care industry if widely adopted by increasing the precision of screening for dysphagia and predicting which patients admitted to hospitals are at risk for aspiration-related pneumonia before that risk manifests itself when at-risk patients begin eating and drinking.

2.2 TRANSDUCERS

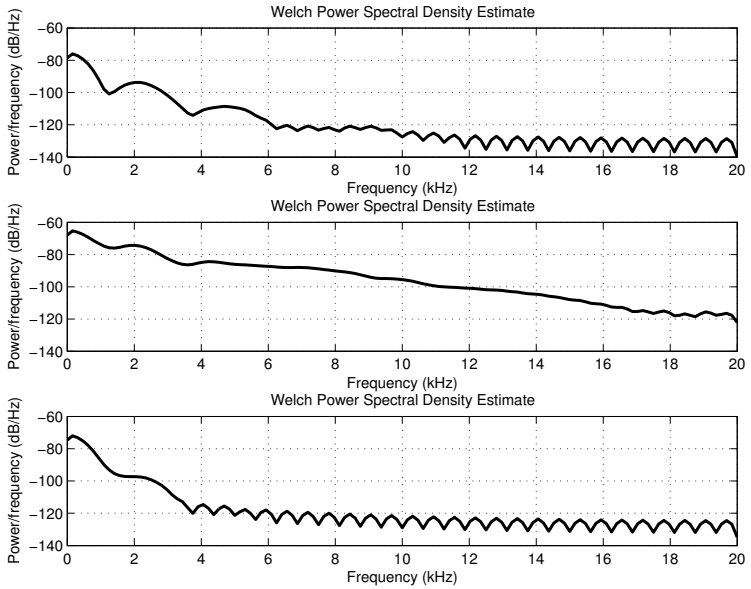
The stethoscope, being the first transducer used in this field, is still often used for cervical auscultation research [5, 29, 56, 60]. Most often, these devices are used in studies alongside imaging modalities such as videofluoroscopy or endoscopy. By recording both sounds and a video of the patient's physiology, researchers have attempted to link the observed sounds with specific physiological events. The studies done by Borr, et al., Zenner, et al., and Marrara, et al. are prime examples. Clinical experts were asked to determine if several features of patents' swallows were normal or abnormal based on a non-concurrent comparison of videofluoroscopic imaging data to observations made at the bedside using a stethoscope [29,60,61]. The feature assessments of the two methodologies were then compared in order to relate the findings of the stethoscope evaluation to an accepted standard [29, 60, 61]. In general, these studies found a high rate of agreement between the bedside evaluation and a videofluoroscopic test [29, 60, 61]. While these studies had a number of design flaws, including a lack of proper blinding of the examiners and non-concurrent video and sound recordings, they did agree that cervical auscultation is a beneficial supplement to the videofluoroscopic evaluation [29,60,61].

In recent years, using microphones and accelerometers for cervical auscultation research has become increasingly common. Though they are not particularly common in the clinical setting, they offer the advantage of digitizing the recorded signal. This opens up a number of possibilities for advanced signal processing and analysis that are not available when using a subjective acoustic analysis via a stethoscope. Figure 3 provides a visual example of the signals obtained in this manner from a single participant. Unfortunately and despite their different methods of operation, few studies have investigated how these transducers differ when it comes to cervical auscultation. Some have even gone so far as to imply that the two are equivalent [62–66]. Of those that did investigate the two transducers concurrently, they concluded that they are not equivalent in general and that the superior device depends on what signal features are of interest to the researcher [37, 63, 67, 68].

The transducer itself is not the only topic debated in the field, but so is the location the transducer uses to record its signal. Only a few studies have focused their recordings on transducer placements superior to the thyroid cartilage. Experiments done by Selley, et al.



(a) Time domain recording. Top figure: Microphone, Middle figure: Anterior-Posterior direction of the accelerometer, Bottom figure: Superior-Inferior direction of the accelerometer. This figure has been previously published by BioMed Central in [37].



(b) Corresponding power spectral density estimates of (3a)

Figure 3: A swallow recorded with a microphone and dual axis accelerometer [37].

as well as Klahn, et al., Perlman, et al., Pinnington, et al., Roubeau, et al., and Smith, et al. made use of the "Exeter Dysphagia Assessment Technique", which records data from the vicinity of the jawline and hyoid bone [69–74], whereas Passler, et al., Shirazi, et al., Sazonov, et al., and Firmin, et al. tested several in-ear, bone-conduction microphones [75–78]. There have also been some experiments which placed the transducer as low as the suprasternal notch [52,65,79,80]. However, the most common recording location by far has been the spot over the cricoid cartilage and its immediate neighbours [37,40,60,62,67,77,78,81–87]. This is logical as the cricoid cartilage is at the same level as many of the anatomical structures that are active during swallowing. The results of a study by Takahashi, et al, which has been cited quite often since publication, suggested that the signal with the greatest peak signal-to-noise ratio could be recorded by a transducer placed either directly on, immediately inferior to, or immediately lateral to the cricoid cartilage [63]. This conclusion served as inspiration, either directly or indirectly, for many of these later studies.

2.3 PROCESSING TECHNIQUES

Arguably, the greatest advantage that research into the field of cervical auscultation can provide is the ability to process the recorded signals before analysis. The simplest implementation of signal processing techniques, applying an analogue bandpass filter to the transducer's output, has also been the most common in this field. Once again Takahashi, et al.'s work [63], which was later supported by Youmans, et al. [62], is cited often because their study characterized the frequency range of swallowing accelerometry signals. They found that the majority of energy contained in swallowing vibrations is at or below 3.5 kHz, and so many studies followed their example and placed the upper notch of their filter at this frequency to eliminate a degree of noise from their signals [5,38–41,43,44,46,48,51,52,56,75,76,80–83,85,88–102]. A similar lower bound to the energy of swallowing vibrations has not been found, however, and so the location of the lower filter notch varies with the researcher. Some filter out frequencies below 50 Hz to ensure that sources of low frequency noise, such as head movements, are eliminated [39,48,76,79,80,91,102]. On the other hand, other researchers

place the lower notch as low as 0.1 Hz in order to minimize any alterations of the recorded signal [37, 38, 40, 54, 85, 93, 103, 104]. Since no similar bandlimits for swallowing sounds have been found, researchers that utilize microphones bandpass filter their signals to either the human audible range [37, 38, 42, 44, 45, 49, 50, 69, 70, 74, 77, 82, 84, 105–117] or the range of common stethoscopes used in bedside assessments [5, 30, 51, 56, 68, 76, 89, 91, 101, 118] in order to provide some level of noise reduction. For many studies, this analogue filtering step is the only signal processing done due to one of two reasons. First, some studies attempt to mimic the signal recorded during a bedside assessment with a stethoscope so as to draw parallels between the two methods [5, 30, 46, 66, 68, 82, 84, 116, 118, 119]. In this situation, the researchers clearly want to minimize any alterations to the recorded signal. Second, some studies simply investigate the most basic signal features such as the number of swallows over time, the timing of swallowing phases, or simply the onset/offset of a swallowing event [29, 60, 62, 63, 67, 69–71, 73–77, 86–92, 94, 108, 113–115, 120–126]. In this situation, extra analytical accuracy is unnecessary and the features can be determined from noisy data.

It is important to note that not all filtering of the swallowing signal is intentional. Ideally, the transducer used would have a flat frequency response and would not change the amplitude of a recorded signal based on the signal's frequency. This is not always the case, however, and each specific transducer model has a unique frequency response curve. Depending on the researcher's available funds, personal preference, or a legitimate design choice they may utilize a model that amplifies some frequency components more than others. This could potentially affect the quality and accuracy of the recorded signal and thereby affect the results of the experiment. The way a transducer is used can also alter the signal. For example, if an omni-directional microphone is placed too close to the source of the sound the 'proximity effect' causes the lower frequency components of the signal to be amplified. It is possible to use filters to compensate for these imperfect frequency responses, but not all researchers do and so care must be taken when analysing the presented results.

Few researchers have applied more complex signal processing techniques, but those who have approached the issue from multiple directions. The first involves the wavelet denoising of the recorded signal. While analogue filtering can reduce or eliminate the impact of signals in specific frequency ranges it does not do as much to eliminate broad-spectrum noise,

such as white-Gaussian noise [127]. By performing the wavelet decomposition of a time domain signal, eliminating any components that are below a specific threshold value, and reconstructing the signal it is possible to reduce the impact of this broadband noise [127]. Given the low signal to noise ratio of swallowing signals and wide bandwidth, this technique is a good match for cervical auscultation as demonstrated in a number of studies [37, 38, 40, 41, 43, 46, 64, 83, 93, 104, 106].

In addition to general noise reduction, there have also been techniques developed to remove specific unwanted signals from swallowing data. Due to the large trends it can induce in a signal, particularly those recorded by an accelerometer, head movements are one such example of noise data. Sejdić, et al., et al has demonstrated a method of eliminating this unwanted signal which uses splines. A signal can be written in terms of splines as

$$x(k) = b^p(k) * c(k) \tag{2.1}$$

where $c(k)$ is an L_2 sequence of real numbers and $b^p(k)$ is the p^{th} order indirect spline filter, also known as a B-spline. This filter is defined as

$$b^p(k) = \frac{1}{m^n} \sum_{j=0}^{p+1} \frac{(-1)^j}{p!} \binom{p+1}{j} (k - jm)^p u(k - jm) \tag{2.2}$$

where u is a step function and m is a time scaling factor. It was found that, in order to minimize the mean square error of the noise approximation, $c(k)$ must be equal to equation 2.3 [103].

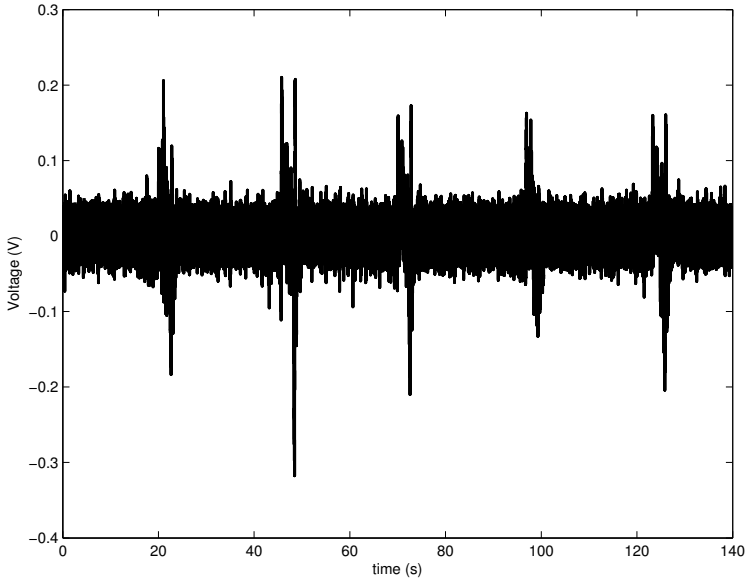
$$c(k) = (b^p * b^p)^{-1} * (b^p * x)(k) \tag{2.3}$$

This technique is able to fit a low frequency (< 2 Hz) trend to the time domain signal that results from head motion during recording, and eliminates the effects of head motion on the signal before the data is analysed [37, 38, 40, 41, 43, 103, 128]. Patient vocalizations, such as speaking or coughing, can also introduce very high-amplitude noise to swallowing signals. The same group was able to utilize the Robust Algorithm for Pitch Tracking (RAPT) to find the areas of highest correlation coefficient of a signal, which correspond to vocalizations, and eliminate them from later analysis [129, 130]. As a swallow and a vocalization cannot

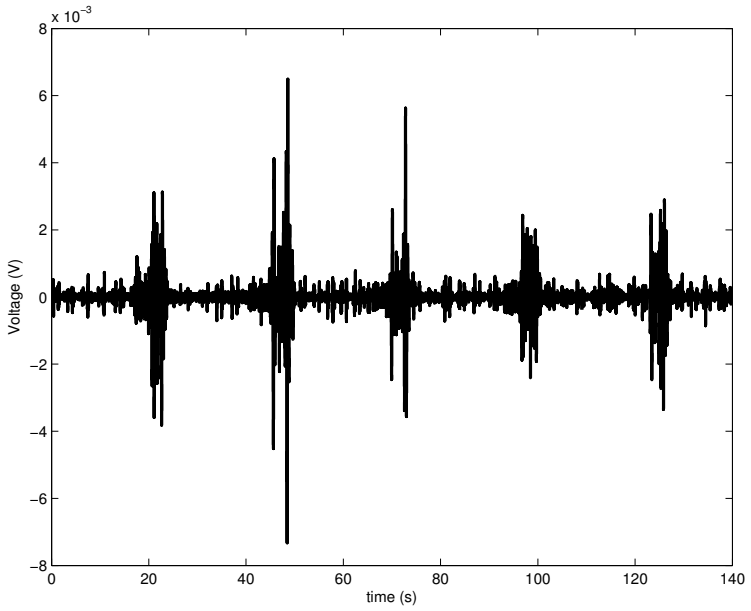
physically occur at the same time, these periods can be freely eliminated without affecting the desired signal [15]. The researchers reported that eliminating this source of noise greatly improved their ability to analytically determine the onset and offset of individual swallows [130]. These researchers have also experimented with techniques to remove noise created by the recording device itself. By fitting an autoregressive model to the output of the transducer, they were able to generate an infinite-impulse response filter to remove noise inherent in the device and improve their signal quality [37,38,41,43,93,104]. As an example, Figure 4 compares a raw transducer signal with one that has undergone all of the conditioning techniques included in [37]. The improvement in the signal-to-noise ratio is visually apparent and should demonstrate how later analysis techniques can benefit from various forms of signal conditioning.

Finally, after removing the unwanted components of swallowing signals, some researchers have attempted to locate and extract periods of swallowing automatically. Both swallowing vibrations and sounds can be divided into two parts: periods of low amplitude when no notable activity occurs, and periods of high amplitude when the structures of the throat move and a swallow occurs. The simplest way to identify these periods of high activity is to threshold the time domain signal, but the accuracy and precision of this method has been shown to be questionable [49–51, 93, 95, 117, 131]. At the other end of the spectrum, some researchers have designed neural networks to identify swallows and non-swallowing events in a given signal [48, 49, 95]. In this situation, the signal is windowed and various features, such as root-mean-squared magnitude or average power, are calculated for each segment [48, 49, 95]. The network then classifies each segment according to the behaviour of the signal during that time [48, 49, 95]. The remaining methods that researchers have used to segment swallowing signals take advantage of their non-stationary nature. Lazareck, et al. and Ramanna, et al. divided the swallowing signal into a number of equal length sections and then calculated, as shown in equation 2.4, the waveform fractal dimension of each [39, 52, 132].

$$WD = \frac{\log \frac{L}{a}}{\log \frac{L}{a} + \log \frac{d}{L}} \tag{2.4}$$



(a) Raw output from accelerometer recording system displaying five saliva swallows [37].



(b) Accelerometer recording system after undergoing conditioning procedures detailed in [37].

Figure 4: Effects of applying signal processing techniques to swallowing accelerometry.

Here, L is the length of a given window, a is the step size, and d is the diameter of the waveform [39, 52, 132]. Swallowing was assumed to occur during the periods of high signal variance, and therefore a large waveform fractal dimension value, so a threshold was set to determine the onset and offset of each swallow [39, 52, 132]. Similarly, Sejdić, et al. also developed a method for identifying swallowing segments based on the variance of the recorded signal. However, rather than using waveform fractal dimension as their feature, they used fuzzy means clustering to sort the data based on its variance over time [37, 38, 43, 54]. Described in equations 2.5-2.7, their algorithm separates the signal into swallowing and non-swallowing clusters, indicated by u_{jk}, based on the prototype v_j and the inner product of the prototype with the signal variance, σ [54]. After providing the initial guesses for u_{jk}, the centres of the two clusters, the values of v_j and u_{jk} are repeatedly updated until the change in the location of the cluster centres is sufficiently small [54].

$$v_j = \frac{\sum_{k=1}^{N} u_{jk}^2 \sigma_k}{\sum_{k=1}^{N} u_{jk}^2} \text{ for } j=1,2 \tag{2.5}$$

$$u_{jk} = \left[\sum_{n=1}^{2} \left(\frac{d_{jk}}{d_{nk}} \right)^2 \right]^{-1} \tag{2.6}$$

$$d_{ij}^2 = ||\sigma_j - v_i||^2 \tag{2.7}$$

2.4 ANALYSIS TECHNIQUES

The most simplistic method of analysis involves a trained examiner judging the recorded waveform personally. Whether visually via a time-domain plot or spectrogram or acoustically, researchers have attempted to link certain waveforms present in the signal to physiological events that are noticed through a videofluoroscopy examination [66, 74, 84, 91, 115]. Features such as the duration of a swallow [71, 90, 91, 105, 107, 108, 112, 116, 122, 133] or the number of swallows that occur over a period of time are of particular interest in this

context [77, 108, 114, 123, 125, 134]. In the vast majority of cases, however, they were not able to find a strong connection between a specific waveform and physiological occurrence, but only weak correlations between events. Researchers have also investigated the overall sensitivity of a bedside dysphagia screen when compared to standard techniques such as videofluoroscopy [30, 118]. Their results demonstrated that their chosen technique is useful as a supplement to existing assessment methods but not as a replacement [30, 118].

Researchers have used a very wide range of different signal features to characterize swallowing sounds and vibrations. Since no one has yet been able to determine what the main characteristics of a swallow are, the exact features calculated for a given experiment are chosen subjectively. In the time domain, it is common to investigate the overall duration of the swallow [38, 39, 52, 62, 65, 81, 82, 85, 93, 132, 134, 135], the timing of the different phases of deglutition such as the duration of a pharyngeal delay [29, 132, 136], the magnitude of the recorded signal [39, 40, 46, 48, 52, 63, 81, 83, 93, 94, 97, 106, 132, 136, 137], and the statistical moments of the signal such as variance or kurtosis [38–40, 46, 52, 85, 93, 120, 121, 135, 138]. Many experiments also looked at various frequency domain features of the signal, such as the peak frequency, average power, or other moments, by either visual inspection of the spectrogram or via the fast Fourier transform [38–40, 46–48, 52, 56, 62, 65, 72, 76, 80, 82, 85, 97, 98, 107, 110, 119, 122, 131–133, 135, 137].

Beyond these obvious features of swallowing signals, some researchers have also investigated more complex characteristics such as stationarity and normality. The statistical complexity of swallowing signals has also been investigated in multiple ways. For example, the Lempel-Ziv complexity, shown in equation 2.8 estimates the randomness of a given signal and is a function of signal length n and number of unique sequences in the signal k [37, 38, 40, 85, 93, 135].

$$\text{Lempel-Ziv Complexity} = \frac{k \log_{100} n}{n} \tag{2.8}$$

Alternatively the Shannon entropy, which is used as an estimate of the uncertainty of a random variable, can also be used to characterize a signal's complexity and has been applied to cervical auscultation research [37, 38, 40, 85, 93, 135]. Shown in equation 2.9, Shannon entropy is a function of the probability mass function ρ of the given signal L.

$$\text{Shannon Entropy (L)} = -\sum_{j=0}^{10^L-1} \rho(j)\ln(\rho(j)) \tag{2.9}$$

The waveform fractal dimension, explained in equation 2.4, has also been used to estimate a signal's complexity with some success [65, 99, 132, 137]. Once again, the chosen measure of complexity varies with the researcher as no one feature has been proven to be the most correct.

Finally, several researchers have investigated signal characteristics that lie outside of the typical time and frequency domains. The Mel-scale Fourier transform has been used in some situations in place of the standard Fourier transform [45, 111]. This transform, which maps the Fourier transform onto the log of a Mel-scale, demonstrates clear advantages when dealing with sounds and comparing recorded signals to human perception [45, 111]. There has also been a fair bit of work done with wavelet transforms [37, 38, 40–42, 44–46, 50, 64, 76, 85, 93, 111, 135]. Instead of describing a signal as the sum of scaled sinusoids, which are infinite in length, this transform describes signals as a sum of scaled and shifted pulses, which are of finite duration [64, 85, 135]. Since swallows are non-stationary signals this technique offers distinct advantages. The hermite projection method, which decomposes the signal into hermite polynomials, shares clear similarities to the wavelet transform method, but does not have the same popularity [131]. Lastly, swallowing signals have been investigated using a phase space transformation [79, 101, 102, 139]. By applying the method of delays, it is possible to map the time domain swallowing signal onto a multi-dimensional phase portrait and generate a recurrence plot [79, 101, 102, 139].

As with the more common transforms, a number of different features can be calculated and used to characterize swallowing signals. The percent of the recurrence plot occupied by recurrence points, the percent of points that form lines parallel to the identity line, and the Shannon entropy, given by equation 2.9, of the length of those parallel lines have all been

utilized on several occasions [79,101,139]. These studies have also calculated the correlation dimension of the system, which estimates the minimum number of variables needed to model a process, by the following equation:

$$CD = \frac{log(C(r))}{log(r)} \tag{2.10}$$

where $C(r)$ is the number of pairs of points in the phase space that are no more than r distance apart [79,101,139]. The Lyapunov exponents, which characterize the convergence or divergence of trajectories in phase space, have also been investigated [102]. These features can be found by solving for λ in equation 2.11, which gives the distance between points in phase space as a function of the Lyapunov exponent (λ), the sampling period (δ), the current point (k), and the distance between an origin point and its nearest neighbour (d_0).

$$d(i) = d_0 * e^{k\delta\lambda} \tag{2.11}$$

3.0 AREAS OF INVESTIGATION

Our literature review of cervical auscultation research has noted several deficiencies of past work. We have identified six key areas of investigation which we feel would be of benefit to the field. The following sections identify each of these topics and explain the strategies that will be used to address them as well as the reasons for their importance.

3.1 SWALLOWING SEGMENTATION

3.1.1 Motivation

The ultimate goal of the cervical auscultation field is to be able to assess a patient's swallowing ability using only non-invasive recording devices. Though work has been done to characterize swallowing vibrations and sounds and in some cases differentiate normal and abnormal swallows, it is still common to rely on a videofluoroscopy examination to provide the swallowing beginning and ending times. Some work has been done to automatically determine these time points from the cervical auscultation data exclusively, but most testing has focused only on data from healthy patients. In order to create a fully autonomous method of swallowing assessment, we must first develop a method of automatically locating swallows in a data set that is independent on the condition of the patient being examined.

3.1.2 Plan of action

Past work has included collecting and analysing cervical auscultation data from healthy subjects. It is a simple matter to apply this methodology to a new data set that consists of subjects that are known to have swallowing difficulties as all existing equipment is available. A concurrent videofluoroscopy examination will be added to the research protocol so as to provide an unambiguous set of time points corresponding to each swallow for later comparison. Past research has also developed a number of useful filtering algorithms which we intend to utilize for this task, as removing noise and artefacts from the signal should make it easier to locate the proper endpoints. Three different segmentation algorithms will then be implemented in Matlab and their segmentation performance will be compared. The first two algorithms, one which utilizes k-means clustering and one which utilizes time-domain thresholding and quadratic variation, were developed in previous studies which used swallowing accelerometry data. The third will implement the DBSCAN algorithm and will developed specifically for this task. Measures of sensitivity, precision, and harmonic average will used as a point of comparison alongside the raw differences between the endpoints provided by each algorithm and the videofluoroscopy assessment.

3.2 DENOISING

3.2.1 Motivation

Cervical auscultation signals have distinctively low signal to noise ratios. Swallowing vibrations and sounds are physiological signals and so are subject to a large number of unrelated events that occur within the body in addition to the usual background noise sources. Past research has shown that wavelet denoising techniques are effective at eliminating a much of the effect of these noise sources and enhancing the signal quality. However, the wavelet transform used in this technique is inherently time-varying which can cause the transformed signal to change considerably based on when a swallow occurs. This can result in artefacts

propagating into the reconstructed signal or simply an incomplete denoising of the data. By implementing the complex wavelet transform, this process can be made approximately time-invariant and improve the signal to noise ratio of the processed signal even further.

3.2.2 Plan of Action

Data from past studies is available, which will allow for testing with data from both healthy and dysphagic subjects. The complex wavelet transform is well defined as is the process of wavelet denoising and can be implemented in a straightforward manner. In summary, wavelet denoising involves performing the wavelet decomposition of a signal, thresholding the resulting components to eliminate the presence of noise, and reconstructing the signal. Implementing the complex wavelet transform in this process simply involves performing the decomposition/reconstruction step twice using orthonormal filter banks. The issues that must be addressed are the development of the filters used in the wavelet decomposition as well as the threshold used to denoise the signal. In order to provide the best results, we plan to create a set of filters to suit our chosen task and data set. The complex wavelet transform provides clear guidelines on certain characteristics of the filters and how they relate to each other, specifically detailing that they must be orthonormal and that the filters in one tree are the reverse and half-sample offset of those in the opposite tree. Past research has shown that utilizing 'Q-shift filters' to decompose the signal has produced good results as these filters generate an approximately linear group delay while maintaining the required orthogonality. The best design for these filters, however is still an open question, which we intend to answer for this specific application. Likewise, past research has developed techniques for wavelet denoising, even for cervical auscultation. We hope to test such methods in the context of the complex wavelet transform and adjust the denoising strategies as necessary.

3.3 HEALTHY AND DYSPHAGIC SWALLOW VARIATIONS

3.3.1 Motivation

In order to use cervical ausculation as an automatic and autonomous method of screening for dysphagia, it is important to first understand how swallows differ from each other. By exploring various features in both healthy and dysphagic populations we will create a clearer picture of what ways these two group differ and be able to develop the means to classify them automatically. It is also possible that the values of these features may provide additional information about the recorded signals and how cervical auscultation relates to the underlying physiological activity.

3.3.2 Plan of Action

Data from both healthy and dysphagic subjects has been collected in previous studies and is available for use. Likewise, various algorithms to filter the data and calculate various features in the time, frequency, and time-frequency domains are available to be used for this research. What remains to be done is applying these algorithms to the data and then applying the relevant statistical analysis techniques to determine what features show statistically significant differences. There are many different ways to stratify the data but we intend to focus on functional differences. Factors such as the presence or absence of dysphagia or stroke are of particular interest as differentiating these populations is of primary concern. Additional factors such as the subject's gender, the viscosity of the swallowed bolus, the presence or absence of compensatory manoeuvres or commands by a third party, and other variables can also be investigated as necessary.

3.4 EFFECTS OF UNSAFE SWALLOWING ON CERVICAL AUSCULTATION

3.4.1 Motivation

As with differentiating dysphagic swallows from healthy activity, detecting unsafe swallows is also a desirable goal. While not as direct as screening for dysphagia in a general sense, knowing when a swallow resulted in significant penetration of the bolus into the larynx would still be of great benefit during both screening and clinical rehabilitation activities. In order to develop this classification method, however, we must first determine how our chosen cervical auscultation features differ between safe and unsafe swallows from patients with dysphagia.

3.4.2 Plan of Action

Our plan for this task will, in general, be identical to our analysis regarding healthy swallows healthy swallows. However, rather than attempting to characterize the differences between healthy and dysphagic swallows, we will attempt to characterize the differences between safe and unsafe swallows from patients with dysphagia.

3.5 HEALTHY AND DYSPHAGIC CLASSIFICATION

3.5.1 Motivation

Once it is determined how cervical auscultation signals vary between healthy and dysphagic swallows, it is then necessary to develop a method to automatically differentiate the two classes. If such a method could be found, it could provide greater objectivity and reproducibility to existing dysphagia screening methods.

3.5.2 Plan of Action

Once the various feature values have been calculated and statistical tests run, we should then have a better understanding of how cervical auscultation signals differ under a multitude of conditions. We will then attempt to automatically differentiate healthy from dysphagic swallows. There are many different ways that we can realize such a system including clustering techniques, Bayesian classifiers, neural networks, and many more which are all well documented methods. However, we intend to implement deep learning techniques for this task. It is a comparatively new technique that identifies high-order, non-linear relationships between inputs. This should provide superior classification ability for physiological signals, which are difficult to classify with a small selection of features and tend to not follow linear relationships.

3.6 SAFE AND UNSAFE SWALLOWING CLASSIFICATION

3.6.1 Motivation

The usefulness of automatically determining which swallows are unsafe using only cervical auscultation signals is similar to the usefulness of classifying dysphagic swallows. Doing so would add a level of objectivity to a traditionally subjective examination with a comparatively low added cost. This technique, if properly developed, could also increase the accuracy of a bedside assessment and allow for monitoring of swallowing activity outside of a dedicated examination environment.

3.6.2 Plan of Action

The process for completing this task is virtually identical to the previous case of attempting to automatically classify dysphagic and healthy swallows. The key variation is that it will utilize a different subset of data and classify it based on the presence or absence of significant laryngeal penetration of the bolus by patients with dysphagia. Beyond that, the same benefits of a high-order, non-linear classification technique apply to this situation.

4.0 AUTOMATED SEGMENTATION OF SWALLOWING VIBRATIONS

The majority of this chapter has been previously published in and reprinted with permission from [140]. ©2015 Elsevier. Dudik, J. M.; Kurosu, A.; Coyle, J. L. & Sejdić, E.. A Comparative Analysis of DBSCAN, K-Means, and Quadratic Variation Algorithms for Automatic Identification of Swallows from Swallowing Accelerometry Signals. *Computers in Biology and Medicine*, 2015, vol. 59, 10-18.
DOI: http://dx.doi.org/10.1016/j.compbiomed.2015.01.007

4.1 MOTIVATION

One of the goals of swallowing research is to be able to identify abnormal swallows without the use of videofluoroscopy. However, many studies still rely on this examination to provide the endpoints of each swallow that the patient makes before they attempt to characterize the data. If cervical auscultation is to be used independently of videofluoroscopy it is necessary to find a viable way of locating periods of swallowing activity from cervical auscultation data exclusively. Such a method would be the first step towards automating the dysphagia screening procedure and would allow for much simpler data collection procedures in future studies. Though several methods have been proposed, no one method has been proven to work sufficiently well in all situations. In this study we have created a new method based on the DBSCAN algorithm which attempts to identify these swallowing segments in a signal recorded from a dual axis accelerometer. We also compare its performance to a similar algorithm which utilizes k-means clustering in order to determine if one clustering method is superior.

4.2 METHODS

4.2.1 The Algorithms

There have been several attempts to segment swallowing vibrations into 'swallowing' and 'non-swallowing' segments, but results have been mixed as the field is still in its early stages. One computationally simple technique involves thresholding the time domain signal, but the accuracy of this method has been questionable [50, 51, 93, 95, 131]. With this technique, an amplitude threshold is declared and any part of the signal that lies above that value, the periods with relatively high amplitude, is considered to be a part of a swallow [50, 51, 93, 95]. This technique has also been modified slightly to threshold, not the signal itself, but various time-varying features with variable levels of accuracy [39, 50–52, 93, 95, 131, 132]. The quadratic variation algorithm is one notably successful example [100]. Based on the magnitude of the amplitude changes between successive points, the algorithm calculates the volatility and curvature of the time domain signal [100]. Since the presence of a swallow causes a notable increase in signal activity and a subsequent large increase in the value of both of these features, swallows can be located by thresholding both feature values and taking the intersection of the sets [100]. While this method has high sensitivity, it is unable to differentiate periods of activity corresponding to swallows and periods of high activity corresponding to coughs or other signal artefacts.

A third technique which strikes a balance between computation requirements and accuracy and has been used to automatically segment swallowing vibrations successfully is the k-means clustering technique [37, 38, 43, 54]. Like with other techniques the signal is windowed and several time-varying features are calculated [141]. Unlike other techniques, however, these points are then mapped onto a feature space along with several additional, randomly placed points [141]. These additional points correspond to the centroids of the number of groups that the data is to be divided into, known as clusters. Each point in the dataset is assigned to the closest centroid, as determined by Euclidean distance or a similar measure, and a cost function is calculated based on the distance between each point and its centroid [141]. Through iterative techniques the k-means algorithm attempts to

find an arrangement of centroids that minimizes the overall cost function of the system, thereby grouping together points with similar feature values [141]. If the chosen features change in value based on whether or not the patient is swallowing, then it is at least theoretically possible to divide swallowing vibration data into swallowing and non-swallowing segments [43,54,141]. However, there are some issues with this clustering technique that can cause problems when segmenting swallowing signals. First, since every point in the feature space must belong to a cluster this algorithm is strongly affected by noise and outliers in the dataset [141]. As a result, the final positions of the centroids can be skewed and a number of points can be assigned to the incorrect cluster, resulting in incorrect swallowing onset/offset times. Second, as a centroid-based clustering technique, the k-means algorithm works best with data clusters that are approximately circular, spherical, or n-spherical as appropriate in the feature space [141]. Data that is distributed in a different pattern or into clusters of different sizes may again result in skewed placement of the centroids and incorrectly sorted data. Finally, since the initial placement of the cluster centroids is random this technique is non-deterministic and may output results that produce a local minimum of the cost function rather than an absolute minimum [141]. This does not necessarily affect the accuracy of the algorithm, but does pose an interesting challenge for any attempts to implement the algorithm without active human oversight.

The density-based spatial clustering of applications with noise algorithm, usually abbreviated as DBSCAN, is an alternative method for clustering data sets that has been developed recently [142]. Unlike other clustering algorithms that require many parameters of the data set, such as the number of clusters in the set, to be known and defined before computation, the DBSCAN algorithm has only two input parameters: the minimum size of a cluster and the maximum distance between points in a cluster [142]. The algorithm operates by cycling through all points in the data set and calculating the number of neighbours each point has, which is defined as the number of other points that are within the minimum distance of the original point [142]. Any data point that has fewer neighbours than the minimum cluster size parameter is declared to be a noise point that is not associated with any cluster [142]. However, a point that has at least as many neighbours as the minimum cluster size is declared to be the start of a new cluster [142]. The neighbours of the starting point are added

to this cluster as are the neighbours of those points provided that they meet the minimum cluster size requirements [142]. The cluster continues to grow in this manner until no more points can be added and the algorithm proceeds to search for the start of a new cluster among the unsorted points [142].

The DBSCAN algorithm has clear computational similarities to centroid-based clustering techniques such as the k-means clustering method. However, the DBSCAN algorithm utilizes the density of the data points in the feature space to identify clusters rather than the location of the centroids, which provides a few advantages. First, this density-based approach allows for superior identification and separation of clusters that are of different sizes and shapes when compared to centroid-based methods [142,143]. In particular, the DBSCAN algorithm is known for being able to correctly separate convex-shaped data clusters in situations where centroid-based clustering perform very poorly. Second, this algorithm has a built-in concept of noise [142,143]. Rather than forcing every point to belong to a cluster to some degree it can exclude points from being part of any cluster, which reduces the effects of outliers in the data set. Lastly, this algorithm is deterministic [142,143]. Some clustering techniques, such as k-means clustering, randomly select the initial locations of the cluster centroids in the feature space which can cause the algorithm to find a local rather than absolute minimum of its cost function. Since there is no randomness inherent in the DBSCAN algorithm it does not carry a similar risk when implemented in an unsupervised manner.

4.2.2 Data Sets

We used two separate data sets for analysis. The first was a collection of 100 artificial signals intended to test the basic functionality of the algorithm. The idea was to very generally represent real swallowing signals by generating noisy signals with localized bursts of activity. Each signal was composed of ten non-overlapping sinusoids with random start times, durations, and frequencies below 5 kHz added to a stream of Gaussian white noise. The signal to noise ratio was equal to four for all sinusoids.

Our second data set was collected from an experiment conducted at the University of Pittsburgh Medical Center Presbyterian Hospital (Pittsburgh, Pennsylvania). Adult patients were asked to make several thin liquid swallows (Varibar Thin Liquid, < 5 cps consistency, Bracco, Milan, ITA) from a neutral position while a tri-axial accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts) was placed over the cricoid cartilage. The main axes of the accelerometer were aligned approximately parallel to the cervical spine and perpendicular to the coronal plane, respectively, while the third axis was not used for this study. The cricoid cartilage was chosen as the mounting location as it was previously demonstrated to provide a high quality vibratory signal [62]. This experiment collected data from patients that were known to have swallowing difficulties and were undergoing a videofluoroscopy-based evaluation. Patients with a history of head or neck cancer or surgery were excluded from the study as were those with assistive equipment which obstructed our recording location or who were unable to grant informed consent. No other disorders were grounds for exclusion from the study. A total of 23 participants were recruited, 8 with a history of stroke and 15 without. There were 191 swallows generated while the patients had their heads in a neutral position whereas 40 swallows were performed while in a chin tuck position. Bolus consistency was not controlled for in order to assess the general viability of the segmentation technique, but manoeuvres other than the chin tuck, such as head rotations or sequential swallows, were excluded from our analysis. A trained research speech language pathologist observed the video recording only and determined the start and end points of each swallow. The beginning of the swallow was defined as the time at which the presented bolus passed the ramus of the mandible while the end was the time that the hyoid bone completed its motion and returned to a resting position. The protocol for the study was approved by the Institutional Review Board at the University of Pittsburgh.

4.2.3 Data Processing and Analysis

Data recorded with the accelerometer underwent several stages to improve its signal quality. A signal recorded from the device when presented with no input on a previous date was used to generate an auto-regressive model of the device's noise. The coefficients of this model

were then used to generate a finite impulse response filter that was used to remove the device noise from the recorded signal. Afterwards, motion artefacts and other low frequency noise were removed from the signal through the use of least-square splines. Specifically, we used fourth-order splines with a number of knots equal to $\frac{Nf_l}{f_s}$, where N is the number of data points in the sample, fs is the original 10 kHz sampling frequency of our data, and f_l is equal to either 3.77 or 1.67 Hz for the superior-inferior or anterior-posterior direction, respectively. The values for f_l were calculated and optimized in previous studies. Finally, we attempted to minimize the impact of broadband noise on the signal by utilizing wavelet denoising techniques. Specifically, we chose to used tenth-order Meyer wavelets with soft thresholding. The value of our threshold was chosen to equal $\sigma\sqrt{2\log N}$, where N is the number of samples in the data set and σ, the estimated standard deviation of the noise, is defined as the median of the down-sampled wavelet coefficients divided by 0.6745.

Once the data was processed we then divided each signal into multiple segments which would later be sorted by the DBSCAN algorithm. We used a simple rectangle windowing function with a length of 200 ms and allowed for a 50 ms overlap with each adjacent segment. This window size was chosen as it would allow for adequate precision by the segmentation algorithm to determine the endpoints of a swallow while still providing enough data points in each segment for properly representative feature calculations. We then calculated two features for each window to serve as the basis of the DBSCAN's sorting. The first, standard deviation, is easily calculated through the common formula

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{4.1}$$

where N is the number of points in the sequence x and μ is the mean of x. In order to allow for comparison between signals and to avoid technical issues with the algorithm, the standard deviation of each window was normalized by the standard deviation of the entire input signal before windowing. The second feature we calculated was the waveform fractal dimension

$$WD = \frac{\log L}{\log d} \tag{4.2}$$

For ordered sets of points such as a time-varying signal, L is the total length of the waveform, defined as the sum of the distances between successive points, and d is the diameter of the waveform, defined as the maximum distance between the starting point and any other point in the waveform [144]. Both of these features have been used in past research on swallowing segmentation [79, 93, 132]. The basic premise is that the vibration signal will maintain some baseline value when the patient is not swallowing, but will significantly increase in amplitude and frequency while a swallow is occurring. Both standard deviation and waveform fractal dimension should follow a similar pattern where their values are high only during periods of swallowing activity. We utilized both features concurrently because past research, as well as our preliminary tests, showed that the waveform fractal dimension and standard deviation of swallowing vibrations are not perfectly correlated despite their similarities [79, 132]. By making use of both features in our analysis we can differentiate small noise perturbations that only affect one feature's value from actual signals caused by physiological disturbances that should affect both features. This will reduce the number of false positives that would occur when looking at each feature independently and increase the overall specificity of the algorithm.

The DBSCAN algorithm itself was implemented in a custom application in the Matlab environment. The features corresponding to both accelerometer axes were entered into the algorithm concurrently, resulting in a four-dimensional feature space. Once again, by including both signals in our analysis we can differentiate between noise that is present in only one axis and actual physiologically based signals that are visible in both axes. Furthermore, though attempts were made to do so, the accelerometer axes were not always perfectly aligned in the sagital plane. Examining data from both axes concurrently ensures that information is not lost or attenuated when the signal is analysed. All data points were sorted in chronological order by the DBSCAN algorithm for simplicity and reproducibility. We chose to use a minimum cluster size that was one more than the number of dimensions, giving us a minimum cluster size of five points. Through extensive trial and error we found that a value of 0.125 for the maximum distance between points in a cluster provided adequate, non-trivial segmentation of the signal without over-tuning the parameters. Since we have not adequately investigated the differences between swallows and other vibratory disturbances we divided

our segmented data into two categories. The first category consisted of all of the periods of low signal activity with no swallowing or other disturbances and always corresponded to the first cluster found by the DBSCAN algorithm due to our chronological input of data points. The second category consisted of all other clusters found by the algorithm along with the cluster-less noise points, which all corresponded to periods of high activity in the vibratory signal. We then returned this information to the time domain and applied minor corrections to smooth out the waveform. Considering the duration of swallowing vibrations reported in past studies [37,54,145,146], any segments in the second category that were less than 400 ms were considered to be false positives and were eliminated from consideration. Likewise, any similar length segments of the first category that were flanked by valid swallowing segments were assumed to be part of the swallow for similar reasons.

After clustering the data was sorted into four categories. A correctly segmented swallow consisted of segments that contained exactly one swallow as defined by the speech-language pathologist. A 'missed' swallow was defined as a swallow that was not segmented as such by the algorithm. A 'false positive' occurred when the algorithm produced a swallowing segment that did not contain an actual swallow or a segment that contained only a fraction of the true swallow as defined by the speech-language pathologist. The last category consisted of segments produced when the x-ray camera was not recording and which were not included in our analysis. Though our accelerometer was active and recording data for up to one minute at a time, the actual videofluoroscopy examination occurred in small bursts as a bolus was presented then swallowed. Any segments produced by the algorithm that were found more than five seconds outside the times that the camera was recording were ignored for our study since any swallows which could have occurred during those times were not part of the examination.

We utilized several measures to assess the performance of our segmentation algorithms. The true positive rate is the ratio of swallows that were correctly segmented by the algorithm to the total number of segments produced, the false negative rate is the ratio of swallows that were missed to the total number of segments produced, and the false positive rate is the ratio of false positives to the total number of segments produced. From there, we were able to calculate the sensitivity of the algorithm as the ratio of true positives to the total

number of swallows presented, the precision as the ratio of true positives to the total number of segments produced by the algorithm, and the harmonic average as twice the product of the sensitivity and precision divided by their sum.

4.3 RESULTS

To confirm that the DBSCAN algorithm functioned as intended, we segmented our first data set consisting of 100 sets of 10 noisy sinusoids. Figure 5 is an example of such a waveform and the results of our test. In all cases the algorithm correctly identified the presence of increased signal activity and provided ten continuous segments. However, it typically over-estimated the duration of each segment. On average, the reported beginning and end of each segment was approximately 130 ms before or after the true start or end, respectively. Considering that the error is significantly less than the length of our windowing function this was considered to be acceptable performance of the algorithm for our purposes. This is further supported by the fact that, for the same data set, the k-means algorithm was less accurate and produced endpoints that were approximately 370 ms before the true start of the artificial swallow and 560 ms after the true end. The quadratic variation algorithm performed the best on this artificial data set with endpoints that were only 60 ms greater than the true endpoints.

We then compared the performance of the three algorithms when segmenting swallowing vibrations from patients with swallowing difficulties. Figure 6 provides an example of this data and the corresponding output from the DBSCAN algorithm for illustrative purposes. Table 1 presents the number of swallows that were sorted into each of our output categories while Table 2 presents these same results with statistical measures.

Since only segments that contained the entire swallow were classified as being correct, all segments produced by all three algorithms were longer than the duration provided by the Speech-Language Pathologist. For swallows produced in a neutral head position, the DBSCAN algorithm provided endpoints that were approximately 0.85 s before and after the true endpoints of the swallow. This distance increased to approximately 1.05 s for chin

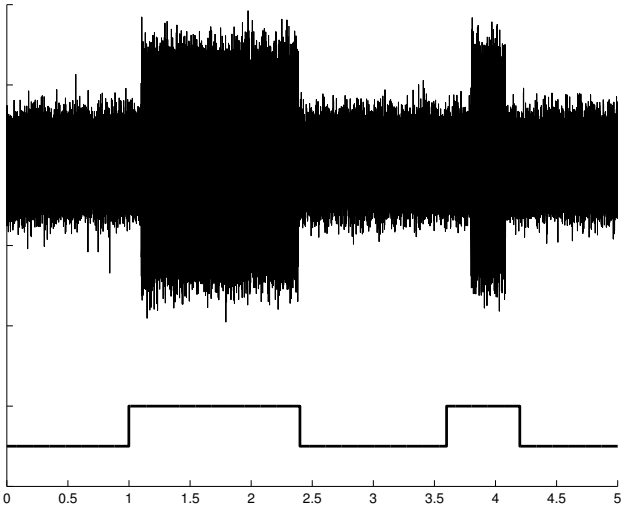


Figure 5: The upper signal is our artificial test signal used to test out sorting algorithm. Below is the indicator sequence of the DBSCAN algorithm which is high when a period of high signal activity is detected and is low otherwise.

Table 1: Raw Segmentation Algorithm Performance

	DBSCAN		K-Means		Quadratic Variation	
	Neutral	Chin Tuck	Neutral	Chin Tuck	Neutral	Chin Tuck
Correct	143	28	103	22	148	32
Missed	48	12	87	18	43	8
False Positive	91	23	101	27	119	35

tuck swallows with a slight bias towards the beginning of the swallow corresponding to the patient's head movement. The k-means algorithm likewise produced endpoints that were 1.20 s before and after the true endpoints of normal swallows and 1.50 s for chin tuck swallows. Finally, endpoints provided by the quadratic variation algorithm were offset by 0.70 s and 0.85 s for normal and chin-tuck swallows, respectively.

Table 2: Summary of Segmentation Algorithm Performance

	DBSCAN		K-Means		Quadratic Variation	
	Neutral	Chin Tuck	Neutral	Chin Tuck	Neutral	Chin Tuck
Sensitivity	0.753	0.700	0.542	0.550	0.775	0.800
Precision	0.598	0.538	0.505	0.449	0.554	0.478
Harmonic Average	0.667	0.608	0.523	0.494	0.646	0.598

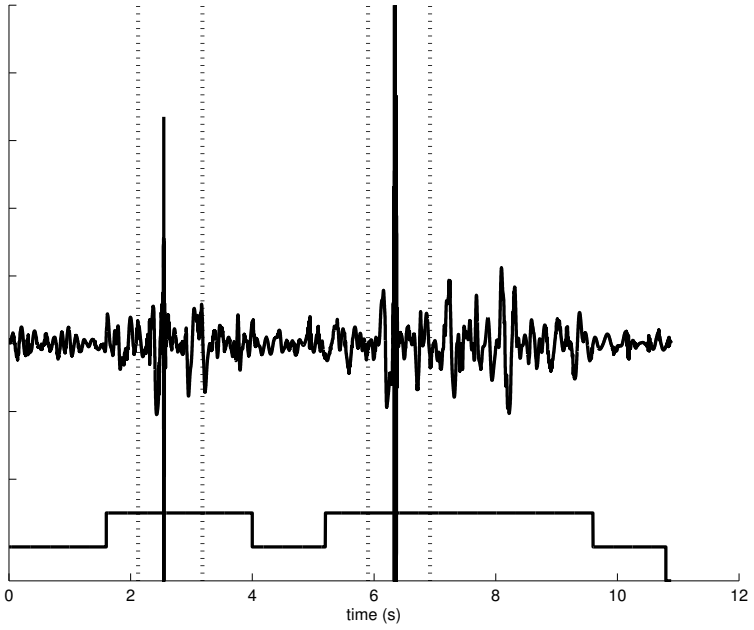


Figure 6: The upper signal shows two swallows made by a patient suffering from swallowing difficulties as recorded by the accelerometer channel aligned in the anterior-posterior direction. The lower signal consists of the output of the DBSCAN segmenting algorithm, which goes high when a swallow is detected and remains low when it has not found a swallow. The dotted vertical lines indicate the endpoints of each swallow as indicated by the speech-language pathologist.

4.4 DISCUSSION

The performance of the quadratic variation algorithm, specifically Table 2, are comparable to the results presented in a previous study with non-healthy swallows [100]. As such, we can safely assume that our data set and testing procedure is valid. The performance of all three algorithms, however, was noticeably worse than its performance with healthy data [54,100]. This could be the result of multiple issues, but we feel that it is chiefly the result of a lower signal to noise ratio. During this experiment, it was noted that the vibrations corresponding to swallows were lower in magnitude than those signals recorded from healthy patients in a previous study [37]. Even after implementing our signal processing strategies the signal to noise ratio did not achieve the same level that it did in the healthy patient experiment. All three algorithms operate on the assumption that the signal being segmented consists of low-amplitude background noise punctuated by relatively high-amplitude swallowing signals. By reducing the separation of these two classes it is more difficult to mathematically differentiate them regardless of the precise method for doing so. Despite this complication, all three algorithms correctly segmented at least half of all presented swallows. Combined with the performance of the algorithms in past studies and with more controlled data sets [37,54,100], we can safely conclude that the lower performance reported in this study is a result of the nature of our data set rather than the implementation of the algorithms.

Looking at Table 2, we can conclude that the k-means algorithm did not perform as well as the alternatives. We believe that this is a result of outliers skewing the data set, which was mentioned previously. We collected data passively during routine videofluoroscopy examinations, which resulted in somewhat more noise to be present in our data than there would have been in a more controlled testing environment. This resulted in all algorithms producing more than the expected number of false positives, as coughing and other artefacts have local variances that are significantly higher than swallows. This is particularly troublesome for the k-means algorithm as the membership of a point to a cluster is dependent on the location of all other points that are already sorted into the cluster. If a number of extreme outliers are added to the data set, the cluster those outliers belong to will change its location and some points will change their membership as a result. In our case, since arte-

facts have high local variances the points associated with these periods are incorrectly sorted as swallows. This causes some points which do contain valid swallows but have smaller local variances, to be missed by the k-means algorithm and sorted as non-swallowing segments. The DBSCAN algorithm does not have this problem. Though it too classified some of these outliers as swallows, producing false positives, the sorting of actual swallows is unaffected. Points are sorted based on their similarity to nearby points only rather than the data set as a whole such as with centroid-based clustering. Likewise, the quadratic variation algorithm sorts data by thresholding the data over time, and so future classification is unaffected by the classification of other points. In summary, the sorting of proper swallowing points is unaffected by the presence of false positive artefacts when using the DBSCAN or quadratic variation algorithms whereas the k-means algorithm suffers from lower sensitivity under noisy conditions.

From our results, we can see that the DBSCAN algorithm and the quadratic variation algorithm performed equally well on our data set. The quadratic variation algorithm demonstrated slightly better sensitivity whereas the DBSCAN algorithm had better precision, but overall they had similar performance. One advantage that the quadratic variation algorithm demonstrated was that the endpoints it provided more closely matched those found through the videofluoroscopic exam by a small amount. However, the DBSCAN algorithm has two key benefits to offset this reduced accuracy. First, it has a notably faster run time. The quadratic variation algorithm does have a complexity of $O(n)$, but this is found assuming that the number of data points extends to infinity [100]. Using our high sampling rate of 20 kHz and recording continuously for several minutes only results in a number of data points on the order of 10^6, which few would consider particularly large with regards to modern computer systems. For practical implementations, this algorithm requires closer to an n^2 number of calculations due to the kernel smoother used in the volatility equation [100]. The DBSCAN algorithm also requires an n^2 number of calculations, but only in the extreme case where no clusters can be found. Our implementation of the algorithm requires far fewer calculations because sorting points into a cluster reduces the number of distance calculations

that must be made when sorting other points. Furthermore, since the DBSCAN algorithm windows the signal to produce its feature space, the clustering algorithm itself operates on an order of magnitude fewer data points.

The second key advantage of the DBSCAN algorithm is the consistency of its performance. While the results varied with the quality of the signal and the patient's actions, the DBSCAN algorithm produced a mixture of true and false positives for a given patient's data. The quadratic variation algorithm, on the other hand, produced significantly different results when presented with signals from different patients even though the overall performance was similar to that of the DBSCAN algorithm. The data from one patient may have been segmented perfectly, but the second data set would produce multiple false positive segments for every true positive while the algorithm was unable to find any segments in a third data set. We believe that this is a result of the difference in the features chosen for each algorithm as well as the nature of clustering and thresholding based classification schemes. The features used in the quadratic variation algorithm, volatility and curvature, are calculated directly from the quadratic variation of the signal. The issue is that the quadratic variation is not a relative measure of a signal's activity, but the raw cumulative sum of its amplitude. The magnitude of these features simply cannot be reliably compared between patients in an uncontrolled environment. As a result, the features used in the quadratic variation algorithm can vary in magnitude significantly between patients and the threshold used for one data set may not produce useful results for another. The DBSCAN algorithm corrects for this issue by using relative values of features. The standard deviation of each data point is not its raw value, but its magnitude relative to the overall signal's standard deviation whereas the waveform fractal dimension is an inherently relative measure for a constant sampling rate. Though these values are not perfectly comparable between patients, large, non-reproducible deviations in the signal such as coughs have less of an effect and the features can be expected to fall within a certain range. The parameters optimized for data from one patient are then translatable to other similar data sets. In addition the these feature differences, the clustering technique used by the DBSCAN algorithm does not have as strict limits on the values of its features. A data point with feature values of $+1$ and $+1$ is functionally equivalent to a point with feature values of $\sqrt{2}$ and 0 if the cluster is located at the origin. This allows the

DBSCAN algorithm to sort a point correctly despite local fluctuations in the signal. The quadratic variation algorithm instead uses hard threshold values for its features. If the signal fluctuates enough so that even one feature does not meet the threshold requirements, then that data point will be sorted as a non-swallowing point. Combining hard thresholding with absolute feature values, as is the case with the quadratic variation algorithm, can strongly impair the algorithm's consistency. On the other hand, the DBSCAN algorithm's use of relative feature values and clustering allow it to better handle unexpected variations in a signal and so can perform more consistently between different patients.

The DBSCAN algorithm also has a few advantages related to usability when compared to the quadratic variation algorithm. First, the DBSCAN algorithm has only one input parameter, the distance between neighbouring points, that must be adjusted in any significant capacity while the others can be simply chosen to suit the task [142]. The quadratic variation algorithm instead relies on three parameters, the thresholds for volatility and curvature as well as the sub-sampling factor k, which must be explicitly calculated or adjusted concurrently in order to segment the data set [100]. This makes the DBSCAN algorithm simpler to implement and modify for a given task. In addition, the segment durations provided by the DBSCAN algorithm are not as closely associated with its input parameter values. The quadratic variation algorithm operates by thresholding the volatility and curvature of a signal over time [100]. Since these values are continuous, increasing or decreasing the threshold magnitudes will correspondingly decrease or increase the length of the segment. This means that the false positive rate and the rate that the algorithm misses swallows are interlocked and one cannot be improved without sacrificing the other. Conversely, the DBSCAN algorithm does not rely on hard thresholding and instead utilizes windowing and clustering techniques. Just as these attributes can somewhat account for large signal changes over time, as described previously, they can also minimize the effects of changing the input parameters. This allows for individual performance metrics of the DBSCAN algorithm to be adjusted independently without additional classification methods or reduced performances in other areas.

The largest obstacle to the implementation of the DBSCAN algorithm for swallowing vibration segmentation is the density of the points in the feature space. One of the general requirements of the DBSCAN algorithm is that each cluster should have a similar feature density. Unfortunately, swallowing vibration signals do not follow this requirement. As described previously, swallowing vibrations are bursts of high amplitude added over a low amplitude background noise. These bursts of activity, however, are not identical. Furthermore, swallowing is very fast compared to the total length of the recorded signal. In our feature space, this results in a large number of points crowded into the low standard deviation and low waveform fractal dimension quadrant consisting of background noise segments and only a few other points spread around the remainder of the feature space that form the segments that contain swallowing activity. There are two ways to solve this. Our chosen method was to simply turn the multi-cluster sorting into a binary sorting, where any point that is not part of the cluster containing background noise was assumed to be part of a swallowing segment. As our results showed, this method has clear problems with generating false positives since it does not differentiate between swallows or other disturbances. Though it could eventually be possible to automatically differentiate swallowing vibrations from coughing, breathing, or other disturbances we do not currently have that knowledge and so little can be done at the moment to correct the issue. The second possible solution is to obtain more data by having the patient initiate a greater number of swallows, thereby more densely populating the area of feature space that contains points associated with swallowing activity. Though a good idea in theory, this is likely not feasible to accomplish on an individual basis. Patient fatigue and safety, particularly with regards to the target population of dysphagic patients, would likely become an issue before an adequate number of swallows were recorded for this solution. One could pool the data from multiple participants, but this would introduce a number of issues not necessarily related to segmenting a given signal and is outside the scope of our current research.

4.5 CONCLUSION

Our goal in this study was to segment swallowing accelerometry data with three different algorithms, one based on k-means, one based on quadratic variation, and a new algorithm that utilized the DBSCAN method, and compare the performance of each. Data was taken from patients with swallowing difficulties and the algorithms were assessed based on the number of swallows found and how closely the calculated endpoints matched those provided by a concurrent videofluoroscopy evaluation. In summary, we found that the k-means algorithm was objectively inferior in all respects, but the DBSCAN and the quadratic variation algorithm had similar results for our chosen performance metrics. We still feel that the DBSCAN algorithm is the superior option, however, because it offers several usability and consistency improvements while providing similar performance to the quadratic variation algorithm. In spite of these advantages, there is still room for improvement when it comes to automatically segmenting swallowing data in an unsupervised manner.

5.0 A MATCHED WAVELET FOR DENOISING SWALLOWING VIBRATIONS

The content of this chapter is currently under review with Biomedical Signal Processing and Control. Dudik, J. M.; Coyle, J. L.; El-Jaroudi, A.; Sun, M. & Sejdić, E.. A Matched Dual-Tree Wavelet Denoising for Tri-Axial Swallowing Vibrations. *Biomedical Signal Processing and Control*, submitted August 2015.

5.1 MOTIVATION

Swallowing vibrations are signals produced by numerous, temporally-overlapping physiological events that co-occur with other non-swallowing physiological events. As a result, they tend to have relatively low signal to noise ratios. Past studies have often implemented wavelet denoising techniques to correct this issue, but the standard, single tree wavelet transform is time varying and the decomposition depends on when the event occurs. Though the decomposition and reconstruction filters are normally balanced so as to correct for this aliasing, denoising the signal by thresholding coefficients upsets this balance and can cause reconstruction errors. In addition, past studies often use one of several, general purpose wavelets. These waveforms are not always best suited to decompose a given signal and can lead to inefficient or ineffective filtering of noise components. In this study, we create a match wavelet that is better suited to decomposing swallowing vibrations than the general purpose wavelets. In addition, we implement this wavelet in a dual-tree decomposition structure so as to minimize the effects of aliasing and maximize the potential signal-to-noise ratio improvement.

5.2 METHODS

5.2.1 Dual-Tree Wavelet Decomposition

The dual-tree complex wavelet transform seeks to eliminate the aliasing issues generated by the single-tree decomposition. In the simplest terms, it decomposes the signal with two parallel trees to create an analytical representation of the input [147]. This effectively doubles the amount of data available for analysis and allows for down sampling without aliasing [147]. This turns the wavelet decomposition into a time invariant process that maintains perfect reconstruction properties regardless of how the decomposed signal of interest is thresholded or otherwise modified [147]. While there are other methods of achieving a time-invariant denoising process, such as removing the downsampling stages entirely, the dual-tree method has a much lower computational load and does not necessarily require significant modification of the single-tree filters.

The dual-tree complex wavelet transform places additional requirements on the properties of its filters when compared to the single tree version. For this study, we chose to utilize the q-shift filter variant as proposed by Kingsbury [147]. In simplest terms, this filter configuration doubles the effective sampling rate of the input signal by using two filter trees with a group delay difference of one half of a sample. This results in the filters of one tree interpolating halfway between the data points analyzed by the other tree. Kingsbury recommends the easiest method of achieving this delay is to first design a single orthonormal filter pair with a group delay of one quarter of a sample [147]. One tree can then decompose the signal with the decomposition filters and reconstruct it with the time-reversed version of those filters, as is typical of the wavelet transform. The other tree uses the reconstruction filters to decompose the signal and the decomposition filters to reconstruct it, thereby adding a quarter sample delay in the opposite direction and producing an overall shift of one half of a sample. The full derivation is provided by Kingsbury in [147], but the relevant portions of their work is included here for convenience.

We first imagine a pair of discrete wavelet trees: tree a and tree b. For illustration purposes, the first two levels of the decomposition can be seen in Figure 7. Each consists of their own decomposition (H_a and H_b) and reconstruction (G_a and G_b) finite impulse response (FIR) filters. Every decomposition filter is followed by a factor of 2 downsampling stage while each reconstruction filter is preceded by a factor of 2 upsampling stage. The overall output of the system is the sum of the outputs of each wavelet tree. The response at a given decomposition level can be more simply presented the cascaded response of all higher level filters. For example, the response of wavelet tree a at the third level to a given input signal can be found by applying the third level decomposition transfer function ($H_{0a}(z)H_{00a}(z^2)H_{001a}(z^4)$) and then the third level reconstruction transfer function ($G_{001a}(z^4)G_{00a}(z^2)G_{0a}(z)$) with all other terms ignored. Here, the number of subscript digits indicates the level of the filter while a value of zero or one in the last position indicates either the scaling or wavelet filter, respectively. Naturally the same process can be applied to the second wavelet tree. If it can be shown that each level of the wavelet decomposition is shift invariant, then the same property can be assumed of the whole system.

We can define $A(z)$ and $B(z)$ as the overall decomposition transfer functions corresponding to a given decomposition level in trees a and b, respectively. Likewise, $C(z)$ and $D(z)$ can be used to represent the complementary reconstruction transfer functions. Finally, it is well known that down sampling and then up sampling a signal $U(z)$ by the same factor provides the following result:

$$V(z) = \frac{1}{M}\sum_{k=0}^{M-1} U(W^k z) \tag{5.1}$$

where M is the re-sampling factor and the scaling term W is equal to $e^{j2\pi/M}$. Combining these choices of notation, we can easily write the output of the dual tree system as follows:

$$Y_{overall}(z) = Y_a(z) + Y_b(z) = \frac{1}{M}\sum_{k=0}^{M-1} X(W^k z)[A(W^k z)C(z) + B(W^k z)D(z)] \tag{5.2}$$

where $Y_{a/b}$ is the output of the respective wavelet tree, X is the system input, and all other terms are defined. We see then, that for $k = 0$, the system behaves as expected with $W^k = 1$. However, for non-zero values of k there is potentially a non-zero response.

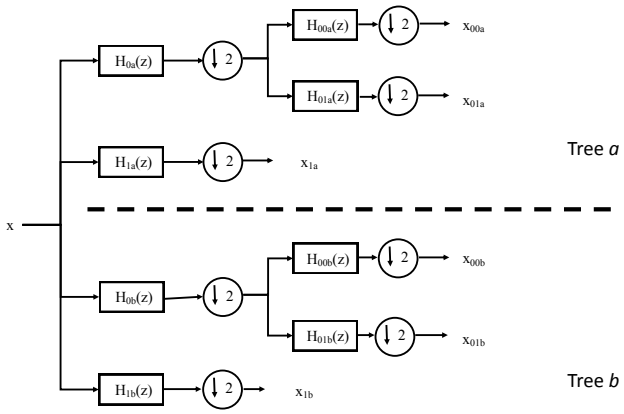


Figure 7: The first two stages of a dual-tree wavelet decomposition. The number of subscripts indicate the level of the filter, H, while the final digit indicates whether it is a scaling or wavelet filter (0 or 1, respectively). The same notation is used for the outputs of the filters. As with the single tree configuration, the reconstruction tree simply inverts this system with reconstruction filters $G(z)$.

Small values of k result in significant overlap of the frequency responses of the decomposition and reconstruction filters. This indicates aliasing of the signal and results in time-varying properties. There are two possible ways to correct for this issue. First, if the alias terms $A(W^k z)C(z)$ and $B(W^k z)D(z)$ can be made sufficiently small, the overall system can be made approximately time invariant. In situations where this is not possible or is too difficult, such as when the filters have less than ideal transition bandwidths, the alias terms can be designed to negate each other.

For values of k greater than or equal to 2, the aliasing terms of our scaling functions can be made insignificant. The additional amount of shift in the frequency domain ensures a minimal amount of overlap between the passbands of the decomposition and reconstruction filters for most filter designs. However, this is not true when $k = 1$ and so we must design our system so that these aliasing terms cancel one another for odd k values.

$$(-1)^k A(W^k z)C(z) = B(W^k z)D(z) \tag{5.3}$$

A valid solution to this system can be found if the following relations are held.

$$B(z) = z^{\pm M/2}A(z) \text{ and } D(z) = z^{\mp M/2}C(z) \tag{5.4}$$

Observing the frequency response of the complementary wavelet filters, we can show that non-zero values of k introduce aliasing terms, as expected. However, these terms are brought about by the overlap of opposite frequency passbands. That is to say, that the positive frequency component of the unshifted filter, $C(z)$ or $D(z)$, will partially overlap with the negative frequency component of the shifted filter, $A(W^k z)$ or $B(W^k z)$, or vice versa. This suggests that the aliasing terms that occur due to the shifted wavelet filters can be cancelled by reversing the positive and negative passband polarities in one of the wavelet trees. If we define $P(z)$ and $Q(z)$ as the positive passbands of the wavelet filters of $A(z)$ and $C(z)$ and the conjugates of $P(z)$ and $Q(z)$ as the negative passbands, our dual tree system can be described as follows

$$\begin{aligned} A(z) &= 2Real[P(z)] = P(z) + P*(z) \\ B(z) &= 2Imag[P(z)] = j[P(z) - P*(z)] \\ C(z) &= 2Real[Q(z)] = Q(z) + Q*(z) \\ D(z) &= -2Imag[Q(z)] = -j[Q(z) - Q*(z)] \end{aligned} \tag{5.5}$$

In this form, we see that the wavelet filters of $B(z)$ and $D(z)$ are the Hilbert transforms of those in $A(z)$ and $C(z)$. When utilized in the dual-tree structure, they can be thought of as the real and imaginary parts of the system's response to an input signal. This analytic representation results in an overall system that has only a positive frequency response, resulting in half of the effective bandwidth of the single tree structure, and is approximately time invariant.

5.2.2 Filter Design

There are several valid designs for these sets of filters. For this project, we have chosen to utilize Kingsbury's q-shift filter relations [147]. The dual set of orthonormal wavelet filters can be calculated from a single, even-length, lowpass filter with a quarter sample group delay $H_L(z)$ as follows [147]

$$\begin{aligned} H_{00a} &= z^{-1} H_L(z^{-1}) \\ H_{01a} &= H_L(-z) \\ H_{00b} &= H_L(z) \\ H_{01b} &= z^{-1} H_L(-z^{-1}) \end{aligned} \tag{5.6}$$

These equations ensure that the two decomposition trees allow for perfect reconstruction of the original signal while orthogonal relationships are maintained between both the scaling and wavelet filters of each tree as well as between the filters of each individual tree. The quarter sample group delay of the scaling filter ensures that its conjugate will have a delay of three quarter samples, thereby resulting in a net half sample delay between the responses of each decomposition tree as necessary. Rather than utilize Kingsbury's method of designing this filter, we chose to create a new wavelet that is optimized for use with cervical auscultation signals.

5.2.3 Wavelet Matching

We followed the procedure outlined by Chapa and Rao to create a wavelet matched to a cervical auscultation signal [148]. Rather than attempting to create multiple wavelets for specific conditions, we attempted to create a generalized wavelet that would be useful for all cervical auscultation applications by matching the wavelet spectrum to the average frequency spectrum of swallowing vibrations. We used data gathered from previous studies via a tria-axial accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts) mounted over the subject's cricoid cartilage [63] to do this. A total of 76 patients with known swallowing difficulties from the University of Pittsburgh Medical Center Presbyterian Hospital

(Pittsburgh, PA) and 55 healthy participants performed 2842 individual swallows. These included swallows in both a neutral and chin-tuck head position and the consistency of the swallowed material ranged from thin ($< 5cps$) fluids to solid food.

Once recorded, the accelerometer signals underwent two filtering steps. First, an autoregressive model of the device's inherent noise profile was created by recording its output with no input signal. The coefficients of this model were then used to generate a finite impulse response filter that was used to remove the device noise from the recorded signal [85]. Afterwards, motion artifacts and other low frequency noise were removed from the signal through the use of least-square splines, as in previous studies [103]. Endpoints for each individual swallow were extracted based on either a custom, clustering-based algorithm (for swallows from healthy subjects) or by a concurrent videofluoroscopy exam (for swallows for dysphagic patients. Each filtered signal's spectrum was then calculated using Matlab's built-in FFT algorithm and the average swallowing vibration spectrum was calculated. This spectrum was smoothed by using a weighted linear least squares method of linear regression with a window size of 50 Hz (2.5% of the signal length) so as to minimize the later amplitude matching error.

With this ideal amplitude spectrum, we then followed Chapa and Rao's work [148] to create a matched wavelet. As with the dual-tree wavelet transform, we reproduce the relevant portions of the original work [148] here for convenience.

While creating a matched wavelet function is relatively simple, ensuring that the corresponding matched scaling function maintains the requirements of an orthonormal multiresolution analysis is more difficult. Utilizing the well known frequency representation of the wavelet and scaling functions

$$\begin{aligned} \Phi(\omega) &= H(\frac{\omega}{2})\Phi(\frac{\omega}{2}) \\ \Psi(\omega) &= G(\frac{\omega}{2})\Psi(\frac{\omega}{2}) \end{aligned} \tag{5.7}$$

as well as the similarly well-defined requirements of the conjugate quadrature filters $H(\omega)$ and $G(\omega)$

$$|H(\omega)|^2 + |G(\omega)|^2 = 1 \tag{5.8}$$

$$H(\omega)\overline{H(\omega+\pi)} + G(\omega)\overline{G(\omega+\pi)} = 0 \tag{5.9}$$

we can show that the scaling function power spectrum can be computed as follows.

$$|\Phi(\omega)|^2 = |\Psi(2\omega)|^2 + |\Phi(2\omega)|^2 \tag{5.10}$$

As is typical, Φ represents the scaling function spectrum, Ψ represents the wavelet function spectrum, and ω is the frequency. Generalizing this expression for a wavelet that is infinitely differentiable produces the following non-recursive expression.

$$|\Phi(\omega)|^2 = \sum_{j=1}^{\infty} |\Psi(2^j\omega)|^2 \tag{5.11}$$

This demonstrates that we can create a wavelet amplitude spectrum which matches our signal of interest and then calculate the corresponding scaling amplitude spectrum while maintaining the requirements of a multi-resolution analysis. Substituting equation 5.11 into the well known Poisson summation of the scaling function

$$\sum_{m=-\infty}^{\infty} |\Phi(\omega + 2\pi m)|^2 = 1 \tag{5.12}$$

to produce the following equation

$$\sum_{p=0}^{l} \sum_{m=-\infty}^{\infty} Y(\frac{2^l}{2^p}(k + 2^{l+1}m)) = 1 \tag{5.13}$$

where k is simply the discrete sample number. The absolute value of the argument of Y is bounded by the passband of our signal of interest. For this paper, we utilize a passband of $[\frac{2\pi}{3}\frac{8\pi}{3}]$ radians, which matches the passband of the common Meyer wavelet and is used by Chapa and Rao for the remainder of their work [148]. l is a defined by the sample rate of

the signal of interest $\delta\omega = 2\pi/2^l$ and is equal to 4 for our work. Expanding the summations in equation 5.13, keeping in mind the bounds on the argument, results in the following matrix forms.

$$\sum_{i=1}^{L} a_{ik} Y(k) \tag{5.14}$$

$$AY = 1 \tag{5.15}$$

L is the number of unique samples in the passband between the lower bound and the Nyquist rate, π, and A has dimensions of Lx2^l. Naturally, both a and A can only have values of 0, 1, or 2 depending on how many valid terms can be found for the argument of Y in equation 5.13 for a given value of k. This matrix A is then used to build a matched wavelet spectrum from a given signal.

We define W as the sampled power spectrum of the target signal with a passband of $[\frac{2\pi}{3}\frac{8\pi}{3}]$ while Y is the corresponding power spectrum of the matched wavelet. We wish to ensure that Y closely matches W in a mean squared sense.

$$error = \frac{(W - aY)^T (W - aY)}{W^T W} \tag{5.16}$$

We can rearrange this equation to solve for Y directly.

$$Y = \frac{1}{a} W + A^T (AA^T)^{-1} (n - \frac{1}{a} AW) \tag{5.17}$$

The term a is a scaling factor, which is defined as

$$a = \frac{n^T (AA^T)^{-1} AW}{n^T (AA^T)^{-1} n} \tag{5.18}$$

n is a vector of dimensions 1xL where each element has a singular value while A has been defined previously.

We did not follow the authors' procedure for matching the phase of our wavelet since we only cared to match the amplitude to the average swallowing spectrum. Instead, we applied a simple, quarter sample phase shift to our matched scaling function amplitude spectrum. This ensured that the wavelet would meet the sample delay requirements set forth by Kingsbury's

q-shift dual-tree complex wavelet transform [147]. Inverse transforming this signal produced the time domain representation of the scaling function of our matched wavelet. From this, it is a simple task to implement a set of FIR filters corresponding to the wavelet.

This procedure was implemented a second time in order to match our artificial test signal, presented in section 5.2.4. This is so we could have a wavelet matched to a known signal, rather than an estimated average, and could assess the ideal effectiveness of our denoising strategy.

5.2.4 Comparison with Artificial Signal

Following the example of Sejdić et al [104], we using the following mathematical signal model to test the effectiveness of our systems.

$$f(n) = \begin{cases} f_0(n) + 0.6\cos(210\pi nT) & 8100 \leq n \leq 16430 \\ f_0(n) + 0.5\cos(210\pi nT) & 11400 \leq n \leq 18330 \\ f_0(n) + 0.2\cos(210\pi nT) & 13200 \leq n \leq 25230 \\ f_0(n) + 0.4\cos(210\pi nT) & 12250 \leq n \leq 23400 \end{cases}$$

$$\begin{aligned} f_0(n) = {} & 0.1\sin(8\pi nT) + 0.2\sin(2\pi nT) + 0.15\sin(20\pi nT) \\ & +0.15\sin(6\pi nT) + 0.12\sin(14\pi nT) + 0.1\sin(4\pi nT) \end{aligned} \tag{5.19}$$

Here, $T = 1/10000s$ and the length of the signal is 36000 samples. The resulting waveform can be seen in Figure 8. We also added Gaussian white noise to this signal with a standard deviation of 1.9 and a signal-to-noise ratio between 0.25 and 6 in increments of 0.25. This is meant to mimic the shape, duration, and frequency content of a real cervical auscultation signal, but is not derived from any real data itself. Shifted versions of this signal were created by removing up to 4200 data points in increments of 600 data points from either the leading or lagging end of the signal and placing them at the opposite side. This ensured that the overall signal energy and frequency content remained unchanged while still simulating a significant sampling delay or poorly-aligned segment endpoints. A total of 1000 normal and shifted pairs were created for each signal-to-noise ratio and shift amount combination.

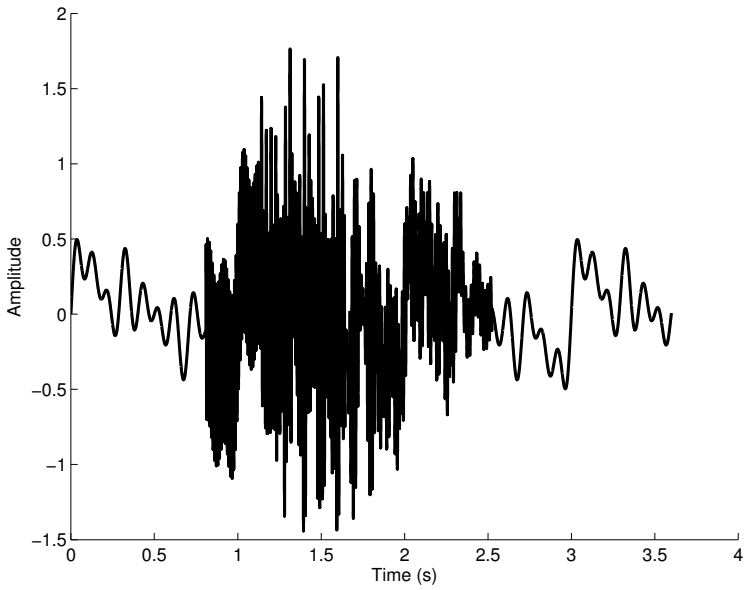


Figure 8: A mathematical imitation of a swallowing vibration signal, originally used in [104].

To provide an upper limit to our denoising algorithm's effectiveness, we created a matched wavelet with respect to this artificial signal. We used the same procedure outlined previously in Section 5.2.3 to create a wavelet with the same frequency response as the 'swallow' portion of our artificial test signal ($8100 \leq n \leq 25230$).

We then used these signals to examine the time invariance and the denoising effectiveness of both the single tree and dual-tree wavelet decompositions. Each signal was decomposed to 5 levels by both the single and dual-tree decomposition trees. Soft thresholding with the optimal threshold as calculated by [104] was used for the denoising tests. In this case, the optimal threshold was found to be $\sigma\sqrt{2\log N}$, where N is the number of samples in the data set and σ, the estimated standard deviation of the noise, is defined as the median of the down-sampled wavelet coefficients divided by 0.6745 [104]. In addition to the wavelets we matched to our artificial signal and our real data, we also utilized the Meyer and 6-tap Daubechies, Coiflet, and Symlet wavelets. As our chosen wavelet matching algorithm [148] combines properties of both the Meyer and Daubechies wavelet families, we feel that these selections will provide a reasonable point of comparison for our new wavelet filter.

Wilcoxon rank-sum tests with a significance value of 0.05 were used to compare the shift invariance and denoising effectiveness for the single and dual-tree systems. The performance of each specific wavelet was also compared against the performance of every other wavelet using the same test to ensure that independent results were obtained.

5.2.5 Comparison with Real Signals

Since the true noise content of a real swallowing signal is unknown we used an alternative method of assessing the denoising effectiveness of our algorithms. Following the example of previous studies [104], we subtracted the approximate signal-to-noise ratio of the original signal from the signal-to-error ratio, as shown in the following equation.

$$\begin{aligned} performance &= 10\log_{10}(\frac{\sigma_x^2 - \sigma_e^2}{r}) - 10\log_{10}(\frac{\sigma_x^2 - \sigma_e^2}{\sigma_e^2}) \\ &= 10\log_{10}(\frac{\sigma_e^2}{r}) \end{aligned} \tag{5.20}$$

The performance is calculated in decibels and indicates the degree to which the signal-to-noise ratio was improved. σ_x^2 is the observed variance of the signal while σ_e^2 is the approximate variance of the noise and is defined as the square of the median of the down-sampled wavelet coefficients divided by 0.6745. The term r is calculated as

$$r = \frac{\sigma_e^2\sqrt{2m}}{N}(\sqrt{2m} + \beta) + d_e - \sigma_e^2 + \frac{2\alpha\sigma_e}{\sqrt{N}}\sqrt{\frac{\alpha^2\sigma_e^2}{N} + d_e - (1 - \frac{m}{N})\frac{\sigma_e^2}{2}} + \frac{2\alpha^2\sigma_e^2}{N} \tag{5.21}$$

where N is the length of the signal, optimal values for α and β were previously calculated to be 1 and 2.5, respectively, m is the number of coefficients that had a value greater than the threshold value, and d_e is calculated as

$$d_e = \frac{1}{N}||x - \hat{f}||^2 \tag{5.22}$$

where x is the original signal and $\hat{f}$ is the output of the denoising algorithm.

5.3 RESULTS

Figure 9 shows the output of the amplitude matching algorithm compared to the average wavelet power spectrum amplitude of real swallowing signals within our desired passband. Figure 10 displays the time domain representation of the scaling function of our real matched wavelet derived from this amplitude match.

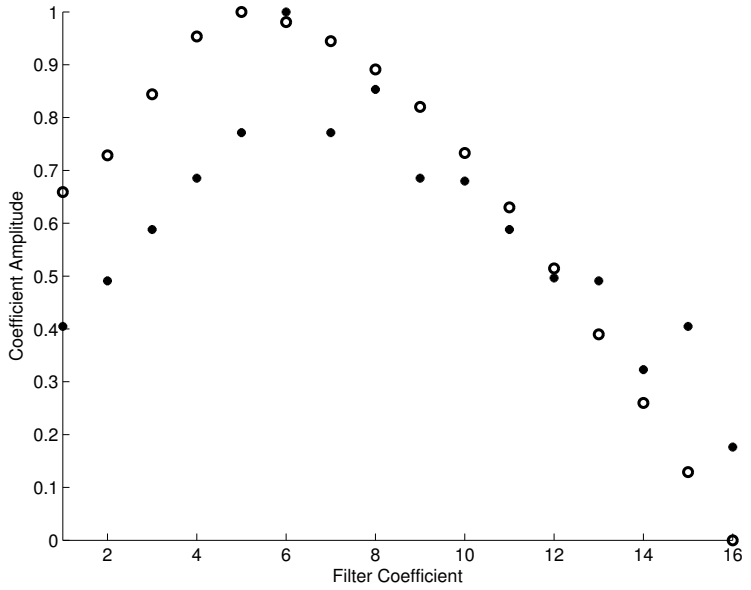


Figure 9: The open circles represent the power of our average cervical auscultation signal at discrete points within our chosen passband. The closed dots indicate the matched wavelet power spectrum as calculated by the procedure in [148].

5.3.1 Shift Invariance

Figures 11 and 12 present the degree to which the single and dual-tree wavelet decompositions changed when conducted with the normal and shifted versions of our artificial signal. Specifically, it presents the total amount of energy that moved between the different levels when the signal was shifted as a percentage of the total signal energy. The single and

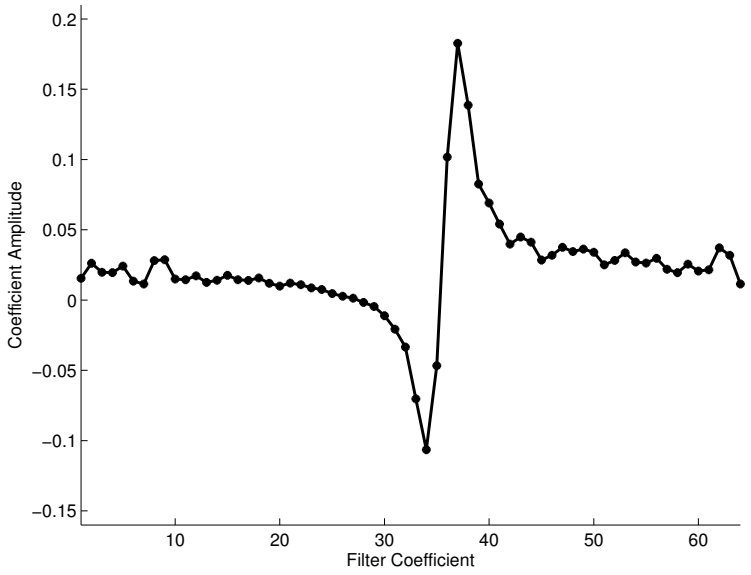


Figure 10: The scaling function derived from the matched wavelet power spectrum in 9 and the procedure described in [148].

dual-tree results were shown to be significantly different $p << 0.001$, and they demonstrate that all of our wavelets possessed greater time invariant properties when used in the dual-tree system. indicating that the decomposition changed less when the input signal was shifted.

5.3.2 Denoising Effectiveness

Figures 13 and 14 display the signal-to-noise ratio of each denoising algorithm output as a function of the input's signal-to-noise ratio for our artificial signal. Overall, the Artificial Matched wavelet eliminated the greatest amount of noise from our artificial test signal. Comparing the single and dual-tree implementations, we see that the Daubechies, Coiflet, and Real Matched wavelets removed a greater amount of the additive noise from the signal while the performance of the Meyer, Symlet, and Artificial Matched wavelet systems degraded to some extent when implemented in the dual-tree configuration. Again, comparing the single and dual-tree outputs resulted in statistically significant outcomes ($p << 0.001$) for all entries.

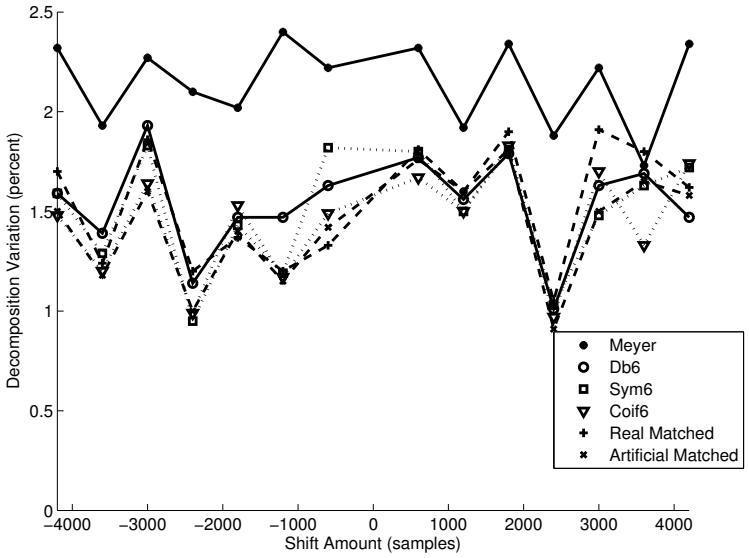


Figure 11: The percent of the signal's energy that changes decomposition levels when the artificial signal is shifted as a function of the shift magnitude. The displayed graph is for the single-tree decomposition.

The results of our analysis with real signals is summarized in tables 3 and 4, where the relevant data is analyzed for each individual axis of vibration. We found that the Real Matched wavelet provided the best denoising ($p << 0.001$) for both the single and dual-tree configurations. We also notice that the dual-tree denoising method was able to remove a statistically greater amount of noise from the signal for all of our chosen wavelets ($p << 0.001$ for all) with the exception of the Symlet waveform.

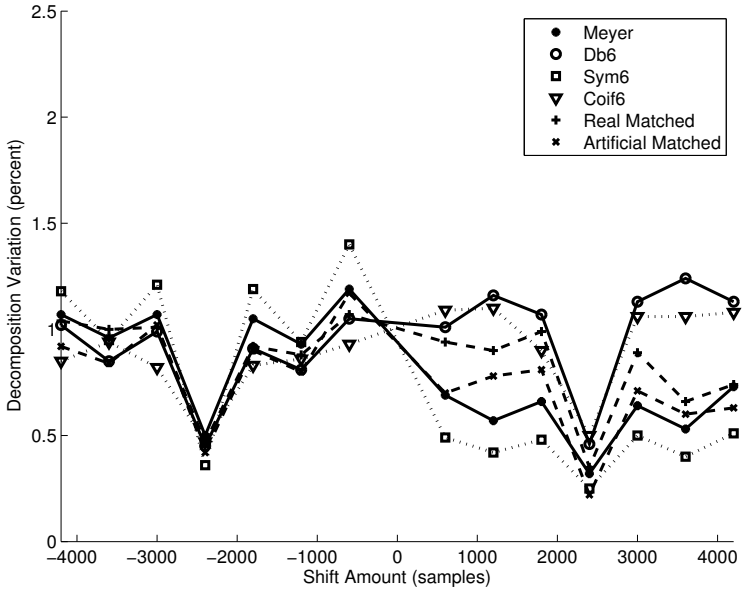


Figure 12: The percent of the signal's energy that changes decomposition levels when the artificial signal is shifted as a function of the shift magnitude. The displayed graph is for the dual-tree decomposition.

5.4 DISCUSSION

5.4.1 Shift Invariance

In this study, we found that the dual-tree wavelet decomposition provided improved time invariant properties for all of our chosen wavelets. However, the degree of improvement varied with the wavelet, with the Meyer wavelet showing the greatest benefit when implemented in the dual-tree system. This suggests that the dual-tree configuration provides tangible benefits for applications which use these wavelets, such as cervical auscultation. Greater time invariance ensures that artefacts introduced by digital sampling or segmentation are minimized and that signal processing techniques can be generalized to any situation.

It is important to note, however, that the overall time varying properties demonstrated by our chosen wavelets were quite small for both decomposition trees. For either tree configuration, we found that the energy contained in a single level of the wavelet decomposition

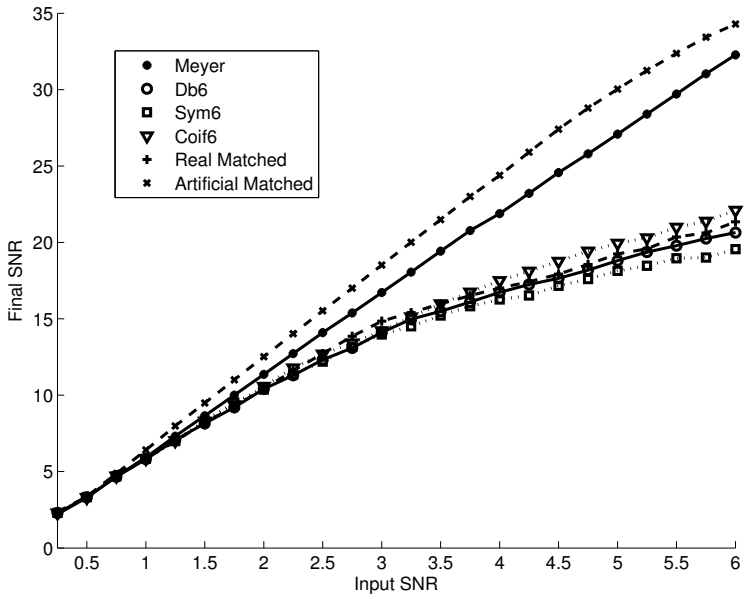


Figure 13: The output signal-to-noise ratio of the single-tree denoising algorithm as a function of the artificial input signal's signal-to-noise ratio.

varied by no more than 2.5% when the input signal was shifted. Often times, this value was less than 3%, as shown by the mean values presented in Figures 11 and 12. Hence, the dual-tree system generally improved these properties but such improvements may be of great interest only for high-precision applications. Real-time applications can utilize the same benefit, but may be willing to accept a greater amount of potential error in order to minimize computation time. In summary, the dual-tree configuration can improve the time invariant properties of a wavelet decomposition of swallowing vibration signals.

5.4.2 Denoising Effectiveness

Similar to our examination of time varying properties, we found that the denoising effectiveness offered by the dual-tree wavelet decomposition varied somewhat with the specific wavelet. The Daubechies, Coiflet, and Real Matched wavelets were all able to eliminate more of the additive Gaussian white noise from the artificial signal when implemented in the dual-

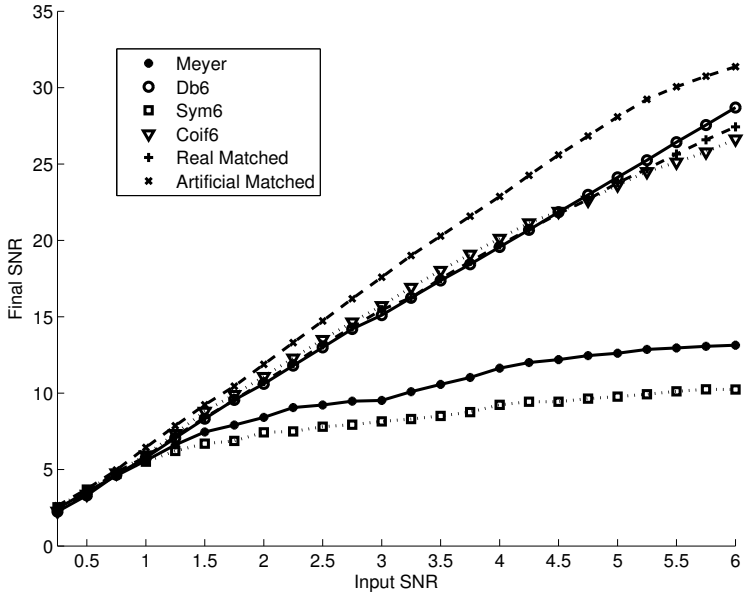


Figure 14: The output signal-to-noise ratio of the dual-tree denoising algorithm as a function of the artificial input signal's signal-to-noise ratio.

Table 3: Single-tree denoising effectiveness with real swallowing signals

	Anterior-Posterior	Superior-Inferior	Medial-Lateral
Meyer	10.22 ± 1.49	10.25 ± 1.50	10.20 ± 1.49
Daubechies6	9.47 ± 1.47	9.48 ± 1.46	9.47 ± 1.46
Symlet6	9.80 ± 1.49	9.77 ± 1.50	9.79 ± 1.50
Coiflet6	8.89 ± 1.45	8.90 ± 1.44	8.89 ± 1.44
Artificial Matched	9.40 ± 1.41	9.41 ± 1.41	9.42 ± 1.40
Real Matched	10.30 ± 1.39	10.32 ± 1.40	10.29 ± 1.40

tree configuration. The Artificial Matched wavelet, on the other hand, performed slightly better in the single-tree configuration while the Meyer and Symlet wavelets performed relatively poorly in the dual-tree configuration. The performance of the Symlet, Meyer, and

Table 4: Dual-tree denoising effectiveness with real swallowing signals

	Anterior-Posterior	Superior-Inferior	Medial-Lateral
Meyer	10.26 ± 0.80	10.25 ± 0.79	10.25 ± 0.79
Daubechies6	10.19 ± 0.80	10.20 ± 0.80	10.19 ± 0.79
Symlet6	9.00 ± 0.80	8.92 ± 0.79	8.96 ± 0.80
Coiflet6	10.21 ± 0.80	10.20 ± 0.80	10.20 ± 0.79
Artificial Matched	10.17 ± 0.78	10.17 ± 0.81	10.19 ± 0.80
Real Matched	10.46 ± 0.82	10.47 ± 0.82	10.46 ± 0.81

Artificial Matched wavelets offer a common trade-off when implemented in either a single or dual-tree system. The first system offers better denoising properties, while the second provides greater time invariance for these specific wavelets. The remaining wavelets perform better overall in the dual-tree system. Judging by Figures 13 and 14, this discrepancy seems to be related to the symmetry of chosen wavelet, with the more asymmetrical wavelets performing better in the dual-tree configuration. However, this may simply be related to our specific artificial test signal and not the underlying decomposition, and so more research would be necessary to provide a conclusive analysis. Regardless, we see that the dual-tree wavelet decomposition does offer potential advantages with regard to the removal of additive noise from cervical auscultation-like signals.

The benefits of the dual-tree configuration are geared towards high-precision applications. The single-tree system is able to remove at least 80% of the additive noise using any wavelet, and the dual-tree system can remove approximately 5% more for those wavelets that have improved performance. This offers a trade-off between a longer computation time and the removal of more noise, or the removal of slightly less noise in less time.

Our results for the denoising of real signals are similar to what was found with regards to the artificial signal. Again, the Symlet wavelet performed poorly in the dual-tree configuration while the Meyer wavelet performed nearly the same and the remaining wavelets showed

varying amounts of improvements. We have observed a maximum increase of approximately 1.5 dB. As with the artificial signal, we see that the dual-tree decomposition offers somewhat greater denoising ability with regards to cervical auscultation signals, but at the cost of greater computational time and complexity. An additional benefit of the dual-tree configuration is the consistency of the results. The amount of noise removed from the real signal was nearly independent of the choice of wavelet when using this method. Furthermore, there was less variance of the algorithm's performance, indicating that the amount of noise removed from each signal was similar. This is likely the result of the final step of the dual-tree system which combines the output of each tree, thereby minimizing the effects of outliers or other significant deviations that occur in a single decomposition tree. The benefits offered by the dual-tree decomposition are still an important consideration in high-precision applications and are worthy of consideration.

5.4.3 Matched Wavelet Effectiveness

Ultimately, we demonstrated that our Real Matched wavelet does offer certain advantages over existing wavelets. Its time invariant properties and denoising effectiveness with regards to our artificial signal are comparable to the more widely implemented wavelets included in this study. This was true for the typical single tree wavelet decomposition in addition to the less common dual-tree decomposition structure. The performance of our Artificial Matched wavelet in these same tasks provides an upper limit to the usefulness of this wavelet matching technique and demonstrates its potential benefit. However, the goal of this Real Matched wavelet was to provide a more sparse decomposition of swallowing vibrations so as to allow for superior noise removal. In that application, it was successful and was able to remove the greatest amount of noise from our real data. It must be noted, though, that this benefit is very small in magnitude when compared to existing wavelets and comes with a notable drawback. The Daubechies and related wavelet families can be accurately represented by a very small number of coefficients. This study, for example, used only 6. The Meyer wavelet, and by extension our Real Matched wavelet as they share certain formulation similarities [148], requires an order of magnitude more coefficients to provide a reasonable

FIR filter approximation. Similarly, though the amplitude spectrum of our Real Matched wavelet much more closely resembles the amplitude spectrum of a swallowing vibration, it does not share all of the useful properties of the Meyer wavelet such as infinite differentiability or symmetry. Therefore, we conclude that the Real Matched wavelet offers some advantages with respect to the removal of noise from cervical auscultation signals, but with additional computational and analytical overhead.

5.5 CONCLUSION

In this study, we sought to improve the wavelet denoising methods traditionally implemented with regards to cervical auscultation signals. We attempted to create a wavelet whose frequency response was matched to that of real cervical auscultation signals in order to allow for a stricter threshold and greater noise removal. We also implemented the dual-tree complex wavelet transform in order to make the process more time invariant. We were able to achieve both objectives. However, the advantages supplied by these methods are very small in magnitude when compared to existing wavelets and the common, single tree wavelet transform. We conclude that our methods offer tangible advantages only to high-precision applications that can spare the additional computation time and complexity.

6.0 COMPARISON OF HEALTHY AND NON-PENETRATING DYSPHAGIC SWALLOWS

The content of this chapter is currently under review with IEEE Transactions on Neural Systems and Rehabilitation Engineering. Dudik, J. M.; Kurosu, A.; Coyle, J. L. & Sejdić, E.. Statistical Variation of Swallowing Sounds and Vibrations Due to Dysphagia in Adults. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, submitted May 2015.

6.1 MOTIVATION

Being able to identify swallowing disorders is an important issue in healthcare. Videofluoroscopy is the most common and most trusted technique implemented for this task. However, the equipment required to perform the exam is expensive and non-mobile, which can limit the availability of proper examinations for patients. Considering the high rate of neurological disorders among the dysphagic population, patient compliance and ability to complete a full diagnostic examination is not always guaranteed. Cervical auscultation has received attention recently as a way to address some of these issues. Particularly, the goal is to develop a method that can automatically and objectively screen for swallowing difficulties in order to enhance the accuracy of the expert's assessment at the patient's bedside or during a videofluoroscopy examination. However, in order to develop this cervical auscultation technique, we must first characterize how swallows made by a patient with dysphagia differ from those made by healthy people.

6.2 METHODS

6.2.1 Data Collection

Our recording equipment consisted of a dual-axis accelerometer and a contact microphone attached to the participant's neck with double-sided tape. The accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts) was mounted in a custom plastic case, and affixed over the cricoid cartilage in order to provide the highest signal quality. The main accelerometer axes were aligned approximately parallel to the cervical spine and perpendicular to the coronal plane and will be referred to as the superior-inferior and anterior-posterior axes, respectively. The third axis was not used for this study. The sensor was powered by a power supply (model 1504, BK Precision, Yorba Linda, California) with a 3V output, and the resulting signals were bandpass filtered from 0.1 to 3000 Hz with ten times amplification (model P55, Grass Technologies, Warwick, Rhode Island). The voltage signals for each axis of the accelerometer were both fed into a National Instruments 6210 DAQ and recorded at 20 kHz by the LabView program Signal Express (National Instruments, Austin, Texas). This set-up has been proven to be effective at detecting swallowing activity in previous studies. The microphone (model C 411L, AKG, Vienna, Austria) was placed below the accelerometer and slightly towards the right lateral side of the trachea so as to avoid contact between the two sensors but record events from approximately the same location. This location has previously been described to be appropriate for collecting swallowing sound signals. The microphone was powered by a power supply (model B29L, AKG, Vienna, Austria) and set to 'line' impedance with a volume of '9' while the resulting voltage signal was sent to the previously mentioned DAQ. This signal was left unfiltered, as an upper limit to the bandwidth of swallowing sounds has not yet been found. Again, the signal was sampled by Signal Express at 20 kHz. For the non-healthy patients, concurrent videofluoroscopy images were available for recording. For these participants, the images output by the x-ray machine (Ultimax system, Toshiba, Tustin, CA) were input to a video capture card (AccuStream Express HD, Foresight Imaging, Chelmsford, MA) and recorded with the same Labview program.

The protocol for the study was approved by the Institutional Review Board at the University of Pittsburgh. The data corresponding to healthy subjects has been published in a previous study [38]. This study used the 'water' and 'nectar-thick' swallows as the data from healthy subjects [38]. A total of 53 non-healthy participants were recruited from the population of patients that were scheduled to undergo a videofluoroscopic evaluation at the University of Pittsburgh Medical Center Presbyterian Hospital (Pittsburgh, Pennsylvania). Thirteen of these patients (10 men, 3 women) had a past history of stroke while the remaining 40 (24 men, 16 women) had no such history. Any patient that was scheduled for this exam was confirmed to have a history of swallowing difficulties. Those patients that had a history of head or neck surgery, were equipped assistive devices that obstructed the anterior neck, or were not in a condition to consent were not included in the study, but no other conditions were excluded. Patients with dysphagia did not undergo a standardized data collection procedure, as the videofluoroscopy examination is modified to suit the individual patient, but analysed swallows were limited to those made while in a neutral head position. The liquids swallowed during the examination included chilled ($5°C$) Varibar Thin Liquid and Varibar Nectar (Bracco, Milan, ITA) presented as either a self-administered cup or a 5mL spoon. The consistencies of these two liquids were determined to be sufficiently similar to the liquids presented to healthy participants based on available product information and qualitative guidelines. A total of 550 swallows were recorded from healthy subjects while 64 were recorded from dysphagic patients with stroke and 158 were recorded from dysphagic patients without a history of stroke.

6.2.2 Signal Processing

Data recorded with the accelerometer underwent several stages to improve its signal quality. A signal recorded from the device when presented with no input on a previous date was used to generate an auto-regressive model of the device's noise. The coefficients of this model were then used to generate a finite impulse response filter that was used to remove the device noise from the recorded signal. Afterwards, motion artefacts and other low frequency noise were removed from the signal through the use of least-square splines. Specifically, we

used fourth-order splines with a number of knots equal to $\frac{Nf_l}{f_s}$, where N is the number of data points in the sample, fs is the original 10 kHz sampling frequency of our data, and f_l is equal to either 3.77 or 1.67 Hz for the superior-inferior or anterior-posterior direction, respectively. The values for f_l were calculated and optimized in previous studies. Finally, we attempted to minimize the impact of broadband noise on the signal by utilizing wavelet denoising techniques. Specifically, we chose to used tenth-order Meyer wavelets with soft thresholding. The value of our threshold was chosen to equal $\sigma\sqrt{2\log N}$, where N is the number of samples in the data set and σ, the estimated standard deviation of the noise, is defined as the median of the down-sampled wavelet coefficients divided by 0.6745. We applied the same FIR filtering and wavelet denoising techniques to the microphone signal after re-calculating the appropriate coefficients. No splines or other low-frequency removal techniques were applied to the swallowing sounds because we had not investigated if such frequencies contained important sound information.

A trained speech language pathologist with established accuracy, inter- and intra-rater reliability for detection of physiological swallowing events recorded with videofluoroscopy, observed the video recording only and determined the start and end points of each swallow from subjects with dysphagia while blinded to the accelerometry data. The beginning (onset) of a swallow segment was defined as the time at which the leading edge of the presented bolus intersected with the shadow cast on the x-ray image by the posterior border of the ramus of the mandible while the end (offset) was the time that the hyoid bone completed motion associated with swallowing-related pharyngeal activity and returned to rest. Each swallow was also rated on an accepted clinical penetration-aspiration scale and any swallows with a rating of 3 or lower was included in our analysis [149]. These swallows, which resulted in no more than shallow penetration of the bolus into the upper airway, were considered safe, non-aspirating swallows for our study as these scores are common among elderly patients without dysphagia [149]. Since no concurrent videofluoroscopy images were available, the swallows produced by healthy patients were segmented with a custom algorithm that has been proven to work with sufficient accuracy on healthy data [54]. Our chosen method was a modified version of a segmentation method proposed by Wang and Willett, which

investigated the local variance of a windowed signal [150]. It utilized a proven two-class fuzzy c-means segmentation technique to locate periods of high variance in our signal and produced the endpoints of each swallow from that information [54].

Once the signals were filtered and segmented we calculated several different features in order to characterized each swallow. In the time domain, we investigated the skewness and kurtosis of the signal, which can be calculated with the typical statistical formulas. We also calculated multiple information-theoretic features by following the procedure outlined in previous publications. The signals were normalized to zero mean and unit variance, then divided into ten equally spaced levels, ranging from zero to nine, that contained all recorded signal values. We then calculated the entropy rate feature of the signals. This is found by subtracting the minimum value of the normalized entropy rate of the signal from 1 to produce a value that ranges from zero, for a completely random signal, to one, for a completely regular signal [135]. The normalized entropy rate is calculated as

$$NER(L) = \frac{SE(L) - SE(L-1) + SE(1) * perc(L)}{SE(1)} \tag{6.1}$$

where *perc* is the percent of unique entries in the given sequence L [135]. SE is the Shannon entropy of the sequence and is calculated as

$$SE(L) = -\sum_{j=0}^{10^L-1} \rho(j) \ln(\rho(j)) \tag{6.2}$$

where $\rho(j)$ is the probability mass function of the given sequence. Quantizing the original signal to 100 discrete levels instead of ten allowed us to calculate the Lempel-Ziv complexity as

$$C = \frac{k \log_{100} n}{n} \tag{6.3}$$

where k is the number of unique sequences in the decomposed signal and n is the pattern length [151].

We also investigated several features in the frequency domain. The centre frequency, sometimes referred to as the spectral centroid, was simply calculated by taking the Fourier transform of the signal and finding the weighted average of all the positive frequency components:

$$C = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{6.4}$$

where $x(n)$ is the magnitude of a frequency component and $f(n)$ is the frequency of that component. Similarly, the peak frequency was found to be the Fourier frequency component with the greatest spectral energy. We defined the bandwidth of the signal as the standard deviation of its Fourier transform [135].

Lastly, we characterized our signal in the time-frequency domain. Previous contributions found that swallowing signals are to some degree non-stationary [152], to which wavelet decomposition is better suited than a simple Fourier analysis [93, 153, 154]. We chose to decompose our signal using tenth-order Meyer wavelets because they are continuous, have a known scaling function [155], [156], and more closely resemble swallowing signals in the time domain compared to Gaussian or other common wavelet shapes [104]. The energy in a given decomposition level was defined as

$$E_x = ||x||^2 \tag{6.5}$$

where x represents a vector of the approximation coefficients or one of the vectors representing the detail coefficients. $||*||$ denotes the Euclidean norm [135]. The total energy of the signal is simply the sum of the energy at each decomposition level. From there, we could calculate the wavelet entropy as:

$$WE = -\frac{Er_{a_{10}}}{100}\log_2\frac{Er_{a_{10}}}{100} - \sum_{k=1}^{10}\frac{Er_{d_k}}{100}\log_2\frac{Er_{d_k}}{100} \tag{6.6}$$

where Er is the relative contribution of a given decomposition level to the total energy in the signal and is given as [135]

$$Er_x = \frac{E_x}{E_{total}} * 100\% \tag{6.7}$$

After calculating the relevant features we performed various statistical comparisons on our data set. First, we used the Wilcoxon rank sum test to test for differences with regards to each feature of all three signals for swallows made by healthy people and non-aspirating swallows made by patients with dysphagia but without stroke. In this situation, data was separated based on the consistency of the ingested bolus and a p-value of 0.002 was used to determine significance after applying the Bonferroni correction. This process was repeated to test for differences between dysphagic patients with and without stroke. To mirror the results of our previous studies we performed another set of rank sum tests to examine sex-based differences in the data recorded from the dysphagic population. This data was separated based on the presence or absence of stroke and the Holm-Bonferroni correction was applied with a starting p-value of 0.05. Finally, the effects of bolus viscosity on our data was examined through the use of Wilcoxon signed-rank tests. Again, the data was analysed separately based on the presence or absence of stroke and the Holm-Bonferroni correction was applied.

6.3 RESULTS

Tables 5 through 12 present the mean and standard deviation of each feature of our data set separated by bolus viscosity and the presence or absence of stroke.

Table 5: Time domain features for patients with dysphagia and without stroke performing thin liquid swallows

	A-P	S-I	Sounds
Skewness	0.307 ± 1.800	-0.087 ± 2.396	0.331 ± 5.253
Kurtosis	21.98 ± 29.58	25.18 ± 67.34	342.1 ± 482.3
Entropy Rate	0.986 ± 0.006	0.988 ± 0.004	0.987 ± 0.008
L-Z Complexity	0.065 ± 0.024	0.073 ± 0.027	0.034 ± 0.018

Table 6: Frequency domain features for patients with dysphagia and without stroke performing thin liquid swallows

	A-P	S-I	Sounds
Peak Frequency (Hz)	11.68 ± 27.98	11.74 ± 14.58	304.0 ± 491.0
Centre Frequency (Hz)	73.15 ± 113.9	54.60 ± 84.64	801.9 ± 682.6
Bandwidth (Hz)	134.75 ± 211.4	92.45 ± 106.8	552.1 ± 562.0
Wavelet Entropy	0.905 ± 0.703	1.063 ± 0.707	1.185 ± 0.725

Table 7: Time domain features for patients with dysphagia and without stroke performing viscous swallows

	A-P	S-I	Sounds
Skewness	0.414 ± 1.126	-0.350 ± 1.684	-0.454 ± 6.055
Kurtosis	14.88 ± 28.27	13.80 ± 14.75	426.9 ± 965.2
Entropy Rate	0.988 ± 0.005	0.988 ± 0.005	0.990 ± 0.006
L-Z Complexity	0.068 ± 0.021	0.073 ± 0.025	0.033 ± 0.018

Comparing data from this study collected from patients with dysphagia but without stroke to data collected in a previous study from healthy subjects found many significant differences. For thin liquid swallows, the dysphagic population data demonstrated greater Lempel-Ziv complexity, centre frequency, peak frequency, and bandwidth for all three signals ($p << 0.001$ for all) while demonstrating lower kurtosis, entropy rate, and wavelet entropy ($p << 0.001$ for all). The skewness of the data was mixed. It was lower in magnitude for the anterior-posterior accelerometer signal ($p << 0.001$), but higher in magnitude for the superior-inferior signal as well as the microphone signal ($p << 0.001$ for both) in the dysphagic population. The viscous swallows demonstrated fewer differences between the healthy

Table 8: Frequency domain features for patients with dysphagia and without stroke performing viscous swallows

	A-P	S-I	Sounds
Peak Frequency (Hz)	10.53 ± 22.95	10.02 ± 12.65	64.03 ± 217.5
Centre Frequency (Hz)	93.42 ± 301.3	32.27 ± 23.70	850.4 ± 1289
Bandwidth (Hz)	202.7 ± 557.1	63.60 ± 68.09	615.3 ± 762.6
Wavelet Entropy	0.625 ± 0.637	0.946 ± 0.693	0.908 ± 0.786

Table 9: Time domain features for patients with dysphagia and stroke performing thin liquid swallows

	A-P	S-I	Sounds
Skewness	0.545 ± 2.710	-1.038 ± 1.751	2.082 ± 9.061
Kurtosis	49.86 ± 152.3	22.54 ± 44.13	523.7 ± 978.5
Entropy Rate	0.985 ± 0.009	0.986 ± 0.008	0.989 ± 0.008
L-Z Complexity	0.056 ± 0.020	0.065 ± 0.017	0.028 ± 0.019

and dysphagic populations. As with the thin liquid swallows, the dysphagic population had greater Lempel-Ziv complexity as well as lower entropy rate and wavelet entropy for all three signals ($p << 0.001$ for all). However, only the anterior-posterior accelerometer signal demonstrated greater skewness, centre frequency, and bandwidth as well as lower kurtosis ($p << 0.001$ for all). The superior-inferior accelerometer signal demonstrated a lower centre frequency and bandwidth while the superior-inferior accelerometer and microphone signals demonstrated increased peak frequencies ($p << 0.001$ for all).

Table 10: Frequency domain features for patients with dysphagia and stroke performing thin liquid swallows

	A-P	S-I	Sounds
Peak Frequency (Hz)	34.54 ± 97.57	14.94 ± 54.28	257.0 ± 433.4
Centre Frequency (Hz)	199.3 ± 291.0	90.75 ± 184.6	895.9 ± 899.3
Bandwidth (Hz)	344.2 ± 498.2	136.4 ± 265.6	697.3 ± 704.4
Wavelet Entropy	1.204 ± 0.870	1.171 ± 0.772	1.027 ± 0.776

Table 11: Time domain features for patients with dysphagia and stroke performing viscous swallows

	A-P	S-I	Sounds
Skewness	-0.440 ± 3.314	-0.050 ± 1.078	1.083 ± 2.478
Kurtosis	43.32 ± 147.3	10.42 ± 13.04	281.2 ± 468.1
Entropy Rate	0.988 ± 0.006	0.988 ± 0.005	0.991 ± 0.005
L-Z Complexity	0.060 ± 0.028	0.072 ± 0.024	0.029 ± 0.018

In contrast to these results, comparing dysphagic data with and without the presence of stroke saw few statistical differences. The data from patients with a history of stroke demonstrated greater centre frequency in the anterior-posterior accelerometer signal ($p = 0.006$) along with a greater skewness magnitude in the superior-inferior accelerometer signal ($p = 0.01$) and greater entropy rate in the microphone signal ($p = 0.03$).

The participant's sex also had little impact on the significance of the data. For dysphagic patients without stroke males demonstrated greater skewness magnitude ($p = 0.015$) but lower kurtosis ($p = 0.020$) for the anterior-posterior accelerometer only. For dysphagic patients with stroke, our data showed significantly greater Lempel-Ziv complexity ($p = 0.013$)

Table 12: Frequency domain features for patients with dysphagia and stroke performing viscous swallows

	A-P	S-I	Sounds
Peak Frequency (Hz)	21.07 ± 54.81	19.55 ± 48.17	59.36 ± 199.0
Centre Frequency (Hz)	132.5 ± 315.1	34.34 ± 67.53	788.5 ± 1242
Bandwidth (Hz)	283.6 ± 518.1	82.99 ± 116.7	666.4 ± 824.2
Wavelet Entropy	0.568 ± 0.610	0.719 ± 0.545	0.801 ± 0.794

and bandwidth (0.003) in the anterior-posterior accelerometer signal for male participants while the centre frequency ($p = 0.018$) and wavelet entropy ($p = 0.005$) for the anterior-posterior accelerometer signal and the entropy rate of the superior-inferior accelerometer signal ($p = 0.005$) were lower in the male population.

Lastly, we found a few significant differences between thin liquid and viscous swallows. For non-stroke patients, the higher viscosity bolus produced lower kurtosis ($p = 0.005$) and wavelet entropy ($p = 0.019$) for the anterior-posterior accelerometer signal along with a lower peak frequency ($p = 0.024$) and wavelet entropy ($p = 0.032$) for the microphone signal. For stroke patients, we found that increasing the viscosity decreased the anterior-posterior centre frequency ($p = 0.28$), microphone peak frequency ($p = 0.23$), anterior-posterior wavelet entropy ($p = 0.011$), and superior-inferior wavelet entropy (0.029).

Figures 15 and 16 show the mean and standard deviation of the energy distribution of the wavelet decomposition of all three signals for thin liquid and viscous swallows, respectively. The vibrations demonstrate similar behaviour to our previous studies, with the majority of energy being present in the lowest frequency level [37, 38]. The swallowing sounds, however, demonstrate a large increase in energy in the d8 through d6 bands (corresponding to approximately 40-300 Hz in this study) that was not present in our earlier findings [37, 38].

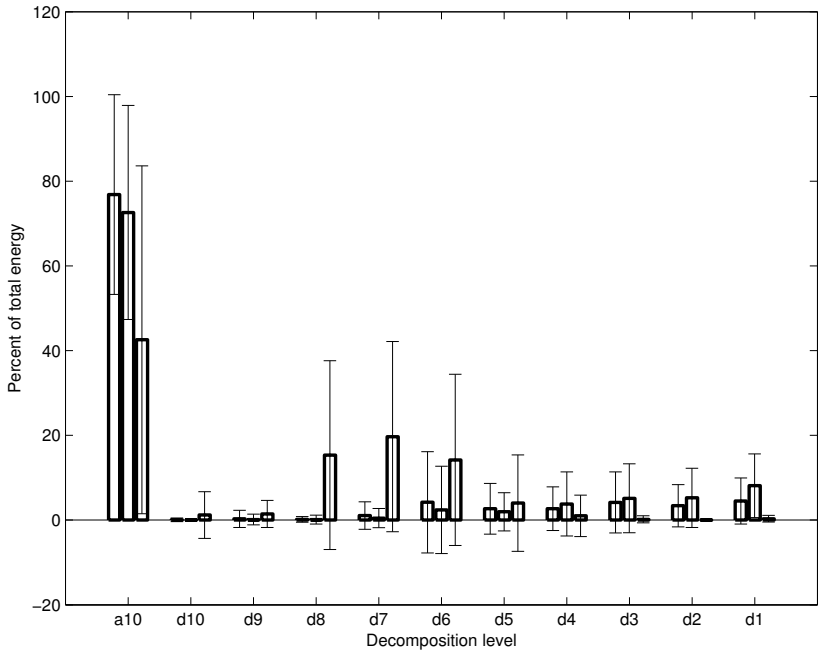


Figure 15: Wavelet energy composition of swallowing vibrations and sounds during thin liquid swallows. From left to right, the bars for each decomposition level correspond to the signals recorded from the anterior-posterior accelerometer, the superior-inferior accelerometer, and the microphone.

6.4 DISCUSSION

In contrast to our previous work on the subject with healthy patients [37, 38], the data gathered from dysphagic patients showed few significant differences with respect to the subject's sex. Also, some of the differences that are present appear to be counter-intuitive, such as how males with stroke showed decreased anterior-posterior centre frequency but greater bandwidth. We feel that this is indicative of the highly variable nature of dysphagia. Even if two patients receive the same diagnosis they may express different symptoms or severity of those symptoms. For example, two patients may experience a stroke and have difficulties swallowing as a result, but the location and size of the lesion will affect their overall nervous and motor functions and can result in a personalized form of dysphagia [1,4,16,57]. The fact

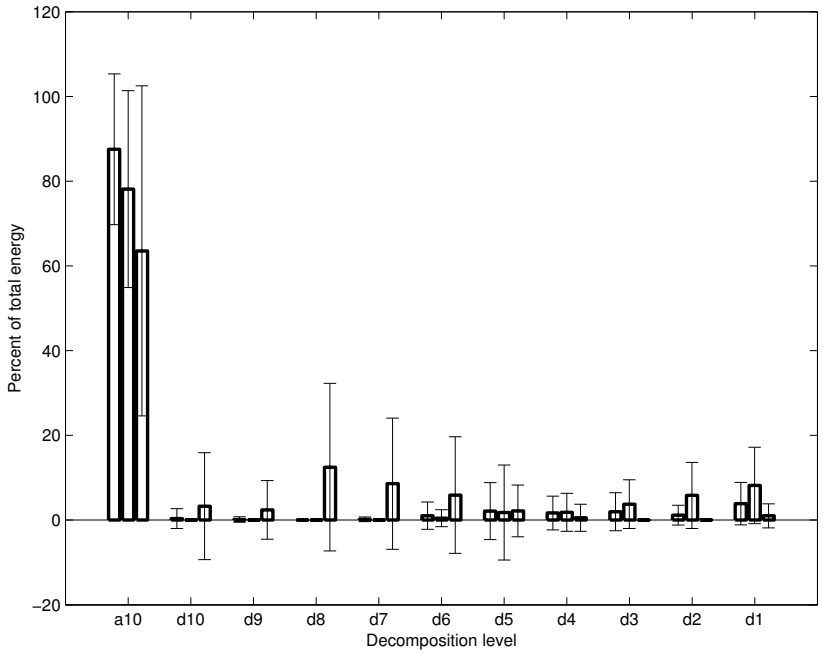


Figure 16: Wavelet energy composition of swallowing vibrations and sounds during viscous swallows. From left to right, the bars for each decomposition level correspond to the signals recorded from the anterior-posterior accelerometer, the superior-inferior accelerometer, and the microphone.

that our data for dysphagic subjects shows wider distributions for all of our features than in the healthy case supports this point [37, 38]. As a result, we conclude that the increased feature variability is greater than and masks any effect that the patient's sex has on the data recorded from dysphagic subjects.

Though we did not find as many differences as we did when analysing data from healthy subjects only [38], our examination of the effects of fluid viscosity are similar to what was expected. For both patients with and without stroke we see that swallowing higher viscosity fluid produced sounds and vibrations with lower frequency, kurtosis, and entropy. Again, for much the same logic as to why we observed fewer effects of the patient's sex, we feel that the lower number of features demonstrating statistical significance is a result of the highly variable nature of dysphagia.

Our study found many differences between healthy and dysphagic patients. When performing thin liquid swallows, subjects with dysphagia demonstrated higher frequency sounds and vibrations with greater Lempel-Ziv complexity, but lower kurtosis, entropy rate, and wavelet entropy. Similar results were found when comparing swallows made with viscous liquid, but in this case the statistical significance of the swallowing sounds and superior-inferior vibrations were lost with respect to skewness, kurtosis, and centre frequency. Together, these factors all indicate a signal that contains more, sudden changes with less predictability. We can rule out the possibility that administering different brands of test liquids caused these results due to the viscosity information gathered (via repeated measures) in both this and our previous study [38]. If the different brands were to blame, then we would expect the data from dysphagic patients to give results opposite to those shown, as the Varibar Thin Liquid is known to be slightly more viscous than ordinary water. However, we are still unsure as to what underlying mechanics did result in this reported variation. Since we did not control for the original cause of dysphagia, with the exception of stroke patients, the signal variations we observed could be the result of many structural, musculoskeletal, or neurological abnormalities. We suspect that our data could be indicative of deficiencies related to the hyolaryngeal movement during a swallow, as this motion contributes significantly to swallowing vibrations, but we have no further proof to support this point [93].

The increase in wavelet energy for swallowing sounds in the 40-300 Hz range may or may not be a significant finding. The initial consideration is that this range corresponds to the frequency of electrical power transmission along with several higher harmonics. Since any x-ray camera will have a significant power draw when it is under operation, it is possible that the wiring for our microphone picked up this radiation and slightly corrupted our signal. However we did not see a similar spike in energy for the signals recorded by the accelerometer, which is not as well shielded from such interference. One also cannot help but notice that this range also corresponds to the reported range of the fundamental frequency of human vocal folds [157, 158]. Many of the swallows included in this study rated a level of 2 or 3 on the penetration-aspiration scale, which indicated that the upper airway was not completely closed and shallow penetration of the bolus into the airway occurred [149]. It is at least theoretically possible that the high hypoharyngeal pressure present during a

swallow, combined with an incomplete sealing of the laryngeal vestibule, forces air across the vocal folds and produces sounds which would never occur in a healthy subject. Again, the lack of similar results for swallowing vibrations is perplexing, but we believe that this issue in particular merits further investigation in the future.

Our study found minimal differences between the sounds and vibrations produced by stroke patients and those produced by patients with other causes of dysphagia. This implies one of two possibilities. The first is that, despite other differences, dysphagia as a symptom of a stroke is functionally equivalent to dysphagia as a symptom of another condition and produces the same sound and vibration pattern. The second, and seemingly more likely, option is that dysphagia as a symptom of stroke does not result in any reliable alterations to our chosen signals. Instead, the feature values of our signals may vary a great deal but not in such a way to make the population distribution significantly higher or lower than the non-stroke population. Judging by the high standard deviation of all of our chosen features and previously described variable nature of dysphagia, we believe that the second option is far more likely.

There are a number of limits to our findings with this study. First, is the chosen population sample. No efforts were made to specifically recruit dysphagic subjects of specific demographic categories, including race, age, and gender. This could potentially bias our results as our population sample may not be an accurate representation of the whole population. This is especially true of subjects with a stroke, as the rate of dysphagia as a result of a stroke in the general population is notably higher than our sample population [4]. Second is our selection of statistical features. While we used a broad selection of features from both time and frequency domains to characterize our swallows, there still has not been a consensus on what features are most important in this regard. Though they certainly provide useful information, it is possible that our chosen features are not the best representation of swallowing sounds and vibrations. The final potential issue of this study is the lack of consistency when performing the modified barium swallow exam. As data was recorded during routine examinations of patients rather than during a dedicated experimental procedure, there were notable differences in the way each patient was examined. Items such as the size of each bolus or the order in which boluses were presented were not strictly controlled and

each patient completed a unique number of swallows as instructed by the examiner. Because of this, it is possible that there is some bias in our data, such as those with more difficulties swallowing being asked to make an overall greater number of swallows, which could have affected our results.

6.5 CONCLUSION

In this study, we attempted to characterize how swallowing sounds and vibrations differ between healthy subjects and subjects with dysphagia. We found that, for swallows with minimal amounts of penetration, the majority of our chosen features did show significant differences between these two groups for both sounds and vibrations. We were also able to confirm our previous findings on the effects of fluid viscosity and the subject's sex on our chosen features. Finally, we found extremely few differences in our chosen signal features between patients with dysphagia as a symptom of stroke and patients with other causes of dysphagia, indicating that dysphagia due to stroke does not result in a single, well-defined functional change. These findings should greatly help the development of the cervical auscultation field and serve as a reference for future investigations.

7.0 UNSAFE SWALLOWING COMPARISON

The content of this chapter is currently under review with the Journal of Neuroengineering and Rehabilitation. Dudik, J. M.; Kurosu, A.; Coyle, J. L. & Sejdić, E.. A Statistical Analysis of Cervical Auscultation Signals from Adults with Unsafe Airway Protection. *Journal of Neuroengineering and Rehabilitation*, submitted July 2015.

7.1 MOTIVATION

In its classical, familiar form as in a public place in which meals are served, aspiration by healthy people is accompanied by choking, coughing and sometimes airway obstructions. In frail, immunologically or medically compromised people as well as those with diseases that directly cause dysphagia by damaging the sensorimotor substrates that enable swallowing, aspiration can be completely undetectable (silent) because airway protective reflexes are attenuated or disconnected. Silent aspiration of saliva, typically mixed with the normal and pathological bacteria residing in the oral cavity [159], is a known cause of aspiration pneumonia which constitutes up to 15.5% of all pneumonias [160]. Due to this significant health risk, early identification of aspiration, particularly silent aspiration, would be of great human importance and benefit. In this study, we investigate if cervical auscultation is capable of differentiating safe from unsafe swallows made by patients with known swallowing difficulties.

7.2 METHODS

7.2.1 Data Collection and Processing

Our data collection protocol, signal processing steps, and feature extraction techniques are all identical to our previous work with non-aspirating dysphagic subjects. These methods are described in sections 6.2.1 and 6.2.2. The only exception is that the analysed data came exclusively from patients with known swallowing difficulties with no data from healthy subjects. We included 17 patients with a primary diagnosis of stroke (10 men, 7 women, mean age 67) along with 51 with known swallowing difficulties but no indication of a stroke (34 men, 17 women, mean age 61). This eliminates the possibility of our previous findings (section 6) being repeated here as false positive results due to inconsistent subject variability.

7.2.2 Statistical Analysis

After calculating the relevant features (as detailed in section 6.2.2) we performed various statistical comparisons on our data set. First, we attempted to test for the normality of our data with the Shapiro-Wilk test as well as the equality of variances via the Levene's test in order to assess the viability of using parametric tests. However, after separating the data based on our chosen variables (PA score, participant's sex, presence of stroke, bolus viscosity) we found that approximately 60% of our feature distributions met these assumptions. At this point, we chose to incorporate non-parametric tests to analyse our data.

We used the Wilcoxon signed rank test to identify differences with regards to each feature of all three signals for safe (PA scores of 1-3) and unsafe (PA scores of 4-8) swallows and stratified by the consistency of the ingested bolus. A p-value of ≤ 0.05 was used to determine significance. This process was repeated to test for differences between dysphagic patients with and without stroke during 'unsafe' swallows. To mirror the results of our previous studies we performed another set of rank sum tests to examine sex-based differences in the signals recorded from the dysphagic population. Finally, the effects of bolus viscosity on our data was examined through the use of Wilcoxon signed-rank tests.

7.3 RESULTS

Tables 13 through 15 present the mean and standard deviation of each feature of our data set separated by bolus viscosity and whether it was a safe or unsafe swallow. Figure 17 displays the average wavelet decomposition of all three of our signals corresponding to unsafe swallows.

Table 13: Feature values corresponding to dysphagic anterior-posterior swallowing vibrations

	Thin		Viscous	
	Safe	Unsafe	Safe	Unsafe
Skewness	0.867 ± 3.743	0.642 ± 1.372	0.491 ± 3.815	0.759 ± 2.055
Kurtosis	87.04 ± 505.3	27.56 ± 52.63	96.20 ± 588.2	39.69 ± 89.38
Entropy Rate	0.987 ± 0.007	0.987 ± 0.007	0.989 ± 0.005	0.986 ± 0.008
L-Z Complexity	0.059 ± 0.022	0.065 ± 0.022	0.056 ± 0.025	0.064 ± 0.022
Peak Freq (Hz)	16.56 ± 51.67	7.162 ± 7.659	56.29 ± 559.1	15.22 ± 37.29
Center Freq (Hz)	189.7 ± 735.8	109.2 ± 245.8	204.7 ± 706.9	141.0 ± 232.7
Bandwidth (Hz)	221.1 ± 373.5	198.1 ± 431.5	273.7 ± 558.8	264.2 ± 393.9
Wavelet Entropy	1.034 ± 0.791	1.003 ± 0.691	0.928 ± 0.820	1.066 ± 0.895

We found no significant differences in any of our features for safe or unsafe thin liquid swallows. For viscous swallows, we found that the anterior-posterior vibrations had greater Lempel-Ziv complexities ($p = 0.039$) and lower entropy rates ($p = 0.022$) during unsafe swallows. We also found that the superior-inferior accelerometer bandwidth was greater for unsafe swallows ($p = 0.033$), while the microphone peak frequency was lower ($p = 0.048$) when compared to safe swallows.

Our contrasts with regards to bolus viscosity and the presence or absence of stroke showed no significant effects of either variable on unsafe swallows. However, we did note several differences with regards to patient sex. Specifically, we found that unsafe swallows made by male subjects showed greater anterior-posterior kurtosis ($p = 0.013$) and superior-inferior

Table 14: Feature values corresponding to dysphagic superior-inferior swallowing vibrations

	Thin		Viscous	
	Safe	Unsafe	Safe	Unsafe
Skewness	-0.557 ± 2.491	-0.435 ± 2.708	-0.129 ± 3.999	-0.441 ± 1.252
Kurtosis	28.91 ± 70.87	101.6 ± 379.0	66.26 ± 291.4	22.68 ± 33.04
Entropy Rate	0.988 ± 0.005	0.989 ± 0.004	0.989 ± 0.004	0.988 ± 0.005
L-Z Complexity	0.068 ± 0.024	0.067 ± 0.024	0.062 ± 0.026	0.069 ± 0.027
Peak Freq (Hz)	11.33 ± 23.41	19.52 ± 35.11	10.79 ± 19.51	30.60 ± 100.5
Center Freq (Hz)	67.53 ± 134.8	143.4 ± 444.2	105.7 ± 421.6	85.52 ± 105.7
Bandwidth (Hz)	114.7 ± 209.6	238.4 ± 587.7	145.3 ± 386.0	180.4 ± 252.7
Wavelet Entropy	1.160 ± 0.730	0.978 ± 0.787	1.138 ± 0.778	1.004 ± 0.803

Table 15: Feature values corresponding to dysphagic swallowing sounds

	Thin		Viscous	
	Safe	Unsafe	Safe	Unsafe
Skewness	-0.317 ± 6.056	-1.564 ± 4.309	-0.525 ± 5.492	-0.125 ± 4.314
Kurtosis	149.2 ± 321.2	187.2 ± 319.9	191.8 ± 413.5	157.3 ± 161.0
Entropy Rate	0.985 ± 0.008	0.986 ± 0.008	0.987 ± 0.006	0.987 ± 0.007
L-Z Complexity	0.055 ± 0.024	0.055 ± 0.031	0.050 ± 0.024	0.052 ± 0.030
Peak Freq (Hz)	94.10 ± 151.0	99.52 ± 187.2	99.46 ± 169.9	92.88 ± 121.6
Center Freq (Hz)	312.5 ± 318.0	348.5 ± 440.4	340.3 ± 578.6	382.2 ± 344.7
Bandwidth (Hz)	348.2 ± 435.8	393.6 ± 703.8	402.7 ± 587.9	399.0 ± 450.0
Wavelet Entropy	1.723 ± 0.724	1.641 ± 0.608	1.596 ± 0.787	1.697 ± 0.716

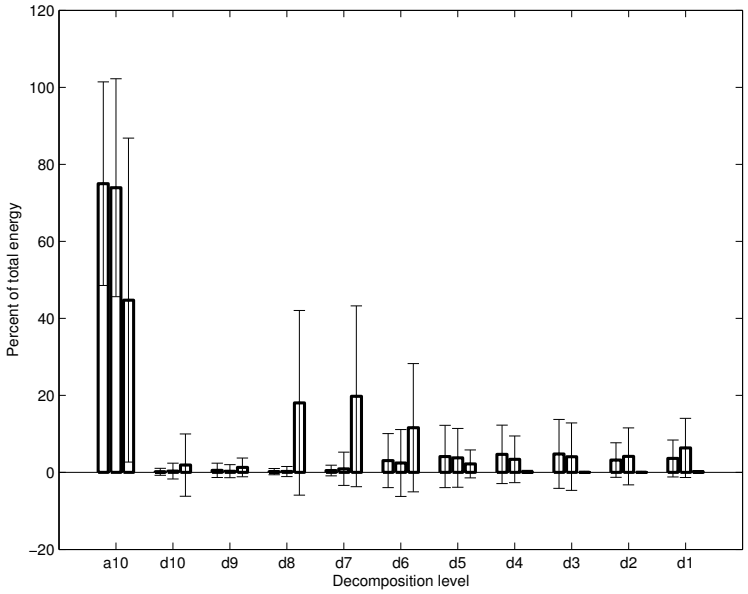


Figure 17: Wavelet energy composition of swallowing vibrations and sounds when the patient produced an unsafe swallow. From left to right, the bars for each decomposition level correspond to the signals recorded from the anterior-posterior accelerometer, the superior-inferior accelerometer, and the microphone.

Lempel-Ziv complexity ($p = 0.016$) corresponding to vibrations along with greater entropy rate ($p = 0.015$), centre frequency ($p = 0.045$), and bandwidth ($p = 0.047$) corresponding to swallowing sounds.

7.4 DISCUSSION

We found that cervical auscultation is able to detect several statistical differences between unsafe swallows of viscous fluid, in which clinically significant aspiration and laryngeal penetration occurred, and safe swallows that either exhibited no airway penetration or airway penetration that falls within the normal range for healthy people. This is of particular interest because aspiration of thicker liquids has been shown to produce higher rates of pneumonia

than aspiration of thin liquids, and longer hospitalization durations than those observed in aspirators drinking thinner liquids [161]. As cervical auscultation signals are not fully understood, we postulate the reasons for why only viscous swallows demonstrated significant differences in this situation. Past research has suggested that thickening agents used during videofluoroscopy exams exhibit non-Newtonian fluid properties, which lead to the reduced aspiration rate in dysphagic patients [162, 163]. It is possible that the penetration of this non-Newtonian fluid into the airway affects the recorded signals in ways that do not occur during thin or non-aspirating viscous swallows. For example, a sudden drop in the pressure exerted on the aspirated material as it enters the larynx could notably reduce the viscosity, and subsequently change the acoustic properties, of a viscous bolus while a thin bolus would be unaffected. Alternatively, viscous swallows are used in the clinical setting because, among other reasons, they provide greater feedback to the patient during a swallow [164,165]. Whether consciously or unconsciously, it is possible that the patient is better able to determine when swallowed material has entered the larynx and react accordingly when aspirating viscous material. This physiological change could alter the cervical auscultation signals as demonstrated in this study.

It is also interesting to note that, when compared to their values for safe swallows, the values of many of the features corresponding to unsafe swallows are closer to the values found in a previous study corresponding to safe swallows made by healthy subjects 6. It may be that our data indicates that deep laryngeal penetration or aspiration occurs when a subject with reduced airway protection performs a swallow as if they did not have a swallowing impairment. In this situation, the patient with dysphagia would behave identically to a healthy subject except for one small detail, such as delaying epiglottic inversion, that would allow material to enter the larynx. A patient that had dysphagia but swallowed safely may have developed a modified swallowing profile, possibly through neuroplastic developments, that compensates for their specific deficiency of airway protection. A similar but alternative explanation is that cervical auscultation is unable to detect the occurrence of aspiration itself, but instead is able to monitor the activity of related swallowing events. As an example, we can imagine a situation where we have a patient that aspirates due to delayed epiglottic inversion and our sensors can record the sounds and vibrations made by the bolus as it

travels through the pharynx, but not the larynx. If the patient does execute an unsafe swallow, then it may be because the bolus was travelling as it normally would in a person with adequate airway protection. On the other hand, if the patient executes a safe swallow it may be because of a longer than normal bolus transit time, which would allow for adequate airway protection in spite of the inversion delay. In this situation our sensors would be able to identify the abnormal swallowing pattern of the safe swallow, but the unsafe swallow would demonstrate little difference from a healthy subject. This distinction between aspiration and altered swallowing patterns could be a vital detail in future work, since aspiration is more common among, but not exclusive to, patients with dysphagia. However, many more statistical features and physiological events would need to be investigated in order to reach a proper consensus on any of these topics, which is beyond the scope of the current manuscript.

Lastly, our sex-based contrasts match our previous work, detailed in section 6, with males demonstrating higher frequency components and greater kurtosis than female counterparts. As described in those studies, we suggest that this is a result of the physical differences of the laryngeal prominence and that future studies should account for these differences during classification tasks. Fewer features showed statistical significance in this regard, however, which we believe to be a result of the added effects of dysphagia and poor airway protection as confounding variables.

Much past work has focused on classifying whether airway protection during swallowing was safe or unsafe, rather than directly characterizing unsafe swallows [40, 43, 46, 136, 166]. However in order to achieve the reported accuracies, these classification techniques simultaneously utilize multiple features that were selected either through principle component analysis [40, 46] or because the features were of particular interest to the researcher [43, 136, 166]. All of these studies found that using at least two features [166], if not more [40, 43, 46, 136], provided noticeable improvement of the data classification when compared to using the value of a single signal feature. Our findings demonstrate the reason for these findings. Though our feature value distributions are not identical between safe and unsafe swallows, we were able to find very few significant differences between individual features for the two states. Attempting to classify swallows using only a single, generalized statistical feature would

produce mediocre results at best. This is not to say that all of our chosen features would be useful for such a task, but that future research into classifying unsafe swallows would need to investigate the concurrent predictive value of their statistical features.

7.4.1 Limitations

These results come with three key limitations, however. First, it is possible that the effects of deep laryngeal penetration and aspiration on swallowing sounds and vibrations were masked or attenuated by other variables. Dysphagia is a highly varied condition that may take completely different forms between patients with the same diagnosis or even between individual swallows from the same patient. Our previous study as well as the work of others showed that safe swallows made by healthy subjects and dysphagic patients showed multiple statistical differences between, but relatively high variation of, individual feature values [29,30,118,167]. This study demonstrated that features corresponding to unsafe swallows are similarly variable. As mentioned previously, it is possible that the main source of cervical auscultation signals is not the deep laryngeal penetration and aspiration event itself, but other swallowing events that may be altered in these patients. Second, our lack of any notable statistical differences between unsafe swallows made by subjects with or without stroke matches our findings with respect to safe swallows [167]. It is possible that our findings indicate that there is not a single consistent physiological expression of dysphagia as a result of stroke, but may also demonstrate that cervical auscultation is unable to identify key existing features of dysphagia caused by a stroke. In either case, this demonstrates that additional investigations will need to be done to characterize the most common form of dysphagia before classification methods could be fully implemented. Finally, our results indicate that cervical auscultation can more easily identify unsafe viscous swallows than unsafe thin swallows. Since aspirating with thin boluses is more common and occurs more often outside of the clinical environment this may restrict the number of potential applications for cervical auscultation. However, we only utilized a small selection of very generalized statistical features in this study. A follow-up study that utilizes features more focused towards cervical auscultation signals or a full machine-learning study could provide a better estimate of the technique's usefulness.

7.5 CONCLUSION

In this study, we recorded swallowing sounds and vibrations from adult patients with dysphagia who exhibited either deep laryngeal penetration or aspirated on one or more swallows during a routine videofluoroscopy exam. We found only a very limited number of statistical differences between swallows during which deep laryngeal penetration or aspiration (unsafe swallows) and those during which only shallow or no laryngeal penetration occurred (safe swallows) based on our chosen features. This supports the findings of other studies and demonstrates the necessity of utilizing multiple statistical features to characterize aspiration. We suggest that the difference we did find is due to a complex interaction between the non-Newtonian nature of thickened liquids and the reduced airway protection in dysphagic patients. We also confirmed the findings of our earlier work with regards to the effects of stroke and sex on cervical auscultation signals. In summary, we conclude that no simple statistical feature can be used to characterize impaired airway protection in dysphagic patients, and that multiple features must be accounted for when aspiration is chosen as a variable in future work.

8.0 CLASSIFICATION OF HEALTHY SWALLOWS

The content of this chapter is currently under review with IEEE Transactions on Pattern Analysis and Machine Intelligence. Dudik, J. M.; Coyle, J. L.; El-Jaroudi, A.; Mao, Z-H.; Sun, M. & Sejdić, E. Deep Learning for Classification of Normal Swallows in Adults. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, submitted October 2015.

8.1 MOTIVATION

Cervical auscultation would be most useful as a screening method for dysphagia if it could function autonomously so that it might provide a more objective assessment than existing methods. In particular, such a method should be able to differentiate swallows made by healthy people from swallows made by people with dysphagia using only the recorded cervical auscultation signals. However, past studies into this aspect of the technique have been relatively limited. They have often used small sample sizes, pre-determined statistical features, linear classifiers, and additional transduction methods in order to classify these groups. By using a broader selection of data and a non-linear classification method we hope to provide a more accurate and more generalizable method of differentiating swallows made by healthy and dysphagic subjects.

8.2 METHODS

8.2.1 Data Set

The protocol for the study was approved by the Institutional Review Board at the University of Pittsburgh.

A total of 55 healthy participants (28 men, 27 women, mean age 39) were recruited from the neighbourhoods surrounding the University of Pittsburgh campus. Each confirmed that they had no history of swallowing disorders, head or neck trauma or major surgery, chronic smoking, or other conditions which may affect swallowing performance. The subjects were asked to complete a total of 20 independent swallows of several types of boluses (water, 'nectar' thick liquid, and 'honey' thick liquid) while their head was in a neutral position. This process was repeated with the subject's head in a 'chin-tuck' position. Five swallows of each bolus type were completed by each subject in both positions, resulting in a total of 1650 recorded swallows. The beginning and end points of each swallow were found using a custom algorithm that has been shown to provide results similar to those given by manual analysis [54].

The non-healthy participants consisted of a total of 53 patients (34 men, 19 women, mean age 63) with suspected dysphagia that were scheduled to undergo a videofluoroscopic swallowing evaluation at the University of Pittsburgh Medical Center Presbyterian Hospital (Pittsburgh, Pennsylvania). Any patient that was scheduled for this exam was confirmed by clinical examination to have evidence of probable dysphagia or a history of swallowing difficulties. Those patients that had a history of major head or neck surgery, were equipped with assistive devices that obstructed the anterior neck such as a tracheostomy tube, or were not sufficiently competent to give informed consent were not included in the study, but no other conditions were excluded. These patients did not undergo a standardized data collection procedure, as the videofluoroscopy examination is routinely modified by the examiner to suit the individual patient. Instead, presentation order, head position, and other environmental factors were unique for each patient. The materials swallowed during the examination were of comparable consistencies (thin liquid (< 5 cps), 'nectar' (≈ 300

cps),'honey' ($\approx$ 2000 cps)) to those provided to healthy subjects based on available product information and qualitative guidelines. Included swallows were limited to those completed in either a neutral or 'chin-tuck' head position. A total of 973 swallows were recorded from these subjects.

For swallows from non-healthy participants, the beginning and end points were defined as the time at which the leading edge of the swallowed bolus intersected with the shadow cast on the x-ray image by the posterior border of the ramus of the mandible and the time at which the hyoid bone completed motion associated with swallowing-related pharyngeal activity and returned to its resting or pre-swallow position, respectively. A trained speech language pathologist with established accuracy, inter-, and intra-rater reliability in analysis of kinematic videofluoroscopic swallowing data and detection of physiological swallowing events located these time points. The speech language pathologist also ensured that each included swallow was rated as a 3 or less on an accepted, 8-point ordinal clinical penetration-aspiration scale [168]. Such swallowing performance indicates no more than shallow layrngeal penetration of the bolus with minimal residue, which is common among elderly patients even without dysphagia [168].

The transducers and signal processing steps used in this study are identical to those used in our other experiments 6.2.1.

8.2.2 Deep Belief Network

8.2.2.1 Neural Network A neural network is a mathematical imitation of a biological neural system that is trained to perform a specific task [169]. One commonly used configuration, a feed-forward network, consists of three major components. The first is the visible layer, which can be anything ranging from individual samples of a signal to various statistical or categorical features. The second component is the hidden layer which consists of many individual 'neurons'. Each neuron outputs a sigmoidal function with an argument that is a weighted sum of all of the inputs in the visible layer. The final component is the output layer. Here, the weighted sum of all of the outputs of the hidden layer is used to indicate the group membership, and thereby classify, the input data. This system can also be thought

of as two restricted Boltzmann machines stacked on top of one another, since connections exclusively travel between adjacent layers and not between distant layers or between neurons in the same layer [169, 170].

Upon initial generation, the weights between the visible, hidden, and output layers are randomized and the network does not produce meaningful results [169]. However, supervised learning can be used to train the network to complete a specific task. Specifically, backpropagation can be used to adjust the weights between layers and ensure the proper output is produced for each input [169]. Using this technique, the weights are adjusted by using gradient descent:

$$w_{ij} = w_{ij} - \alpha \frac{\partial MSE}{\partial w_{ij}} \tag{8.1}$$

where w_{ij} is an array containing the weights between neurons i and j, MSE is the mean-squared error of the output, α is the user-defined learning rate, and $\frac{\partial}{\partial w_{ij}}$ indicates a partial derivative. It can be shown that the partial derivative term can be rewritten as the product of the inputs to the neuron and the error of its output. By implementing this algorithm alongside a labelled training set, a 3-layer, feed-forward neural network can be used to classify data into pre-defined categories [169].

8.2.2.2 Deep Learning Formulation Neural networks have become increasingly popular in recent years due to advances in computing speed and customizability of the algorithm. However, many researchers limit their models to containing a single hidden layer, which can potentially hinder the network's overall performance. Such a model can efficiently calculate only first and second order features and is unable to correlate those higher order features due to the limited arrangement of connections between neurons [171, 172]. One possible solution is to implement a multi-layer deep believe network. Structurally, this involves simply taking the most basic component of the neural network, the restricted Boltzmann machine, and connecting multiple copies together sequentially [170–172]. This produces a network that has multiple hidden layers and can calculate higher order features as well as correlations between higher order features for a given input [170–172]. This can potentially improve the network's classification accuracy since it can analyse data that was not available otherwise.

The final stage of a deep belief network is functionally identical to that of the basic neural network. It is trained through the same supervised learning and backpropagation algorithms and outputs the group membership of a given input. However, backpropagation cannot be used effectively to optimize the weights of the entire network. Instead, each restricted Boltzmann machine that we add to the network must individually undergo an unsupervised learning step [170–172]. To do so, we implement the contrastive divergence algorithm and attempt to minimize the negative log-likelihood of the training data:

$$-\log(P) = -\log(\frac{e^{-E}}{\sum e^{-E}}) \tag{8.2}$$

$$E = -b'_{visible}x - b'_{hidden}h - h'Wx \tag{8.3}$$

where E is the energy of one particular network configuration, x and h are the activations of the visible and hidden neurons, respectively, for a given training set, W is the array of weights between the visible and hidden layers, and b is the bias of the layer of neurons. Taking the gradient of Equation (8.2) provides the update law for the parameters and can be shown to be equal to

$$\frac{\partial}{\partial\theta} - \log(P) = \mu(\frac{\partial E}{\partial\theta}|x) - \mu(\frac{\partial E}{\partial\theta}) \tag{8.4}$$

where μ is the expected value operator, $\mu(*|x)$ is the conditional expected value of $*$ given a value for x, and θ is the vector of model parameters. In explicit form:

$$\frac{\partial}{\partial W} - \log(P) = \mu(-h'x|x) - \mu(-h'x) \tag{8.5}$$

$$\frac{\partial}{\partial b_{visible}} - \log(P) = \mu(-x|x) - \mu(-x) \tag{8.6}$$

$$\frac{\partial}{\partial b_{hidden}} - \log(P) = \mu(-h|x) - \mu(-h) \tag{8.7}$$

Since batch processing is not always available or the most efficient method of calculation, a stepwise contrastive divergence algorithm has been developed. It provides estimates of the negative term in Equations (8.5)-(8.7) and can be summarized as follows:

1. Randomize the initial weights and biases.
2. Determine the hidden layer neuron activations h based on a given input array x.
3. Generate an estimate of the input x' from the current hidden layer h.
4. Estimate a new hidden layer h' based on the estimate of the input x'.
5. Update the weights W based on the difference between the exact and estimated terms, multiplied by a learning rate: $W = W + \alpha(xh - x'h')$.
6. Update the biases in the same manner: $b_{visible} = b_{visible} + \alpha(x - x')$ $b_{hidden} = b_{hidden} + \alpha(h - h')$.

This process is followed for each restricted Boltzmann machine added to the network [170–172]. For the first layer, the input is the raw data. For the additional layers the input is simply the output of the previous layer for a given sample of data.

8.2.2.3 Deep Belief Network Details Our input signal was chosen to be the Fourier transform of each segmented swallowing vibration, normalized relative to the largest amplitude in the signal, with a resolution of 0.25 Hz. The spectrum was limited to the positive frequencies (0 to 5 kHz) to minimize redundancies. Past studies have shown that there is an insignificant amount of energy at or above this frequency, so no information should be lost [56]. We also eliminated the frequencies between 58 Hz and 62 Hz so that the differences in electrical noise for our two data sets would not affect the results of our classification.

Our training and testing sets consisted of a random selection of swallows taken evenly from both data categories. This amounted to 123 swallows from healthy subjects and another 123 swallows from subjects with dysphagia being used as a testing data set. Another 1700 swallows (850 from each category) were used as the training set. The surplus number of swallows from healthy subjects was ignored for the remainder of the study to ensure a balanced distribution. Data was presented to each network in mini batches of ten randomly selected swallows and the training was repeated for five epochs. Stratified ten-fold cross-validation was utilized to generalize the results of our training methods.

We first built 12 unique, independent networks. We constructed 3 'small' networks, which consisted of a 300 node, single layer network, a 2-layer network with 300 neurons in each layer, and a 3-layer network with 300 neurons in each layer. These networks were trained and tested

exclusively on anterior-posterior data. Another 3 'large' networks were constructed that contained 3000 neurons in each layer, but were otherwise used the same manner. Another 3 'small' and 3 'large' networks were built and subsequently trained and tested on superior-inferior data exclusively. The basic structure of these networks is illustrated in Figure 18.

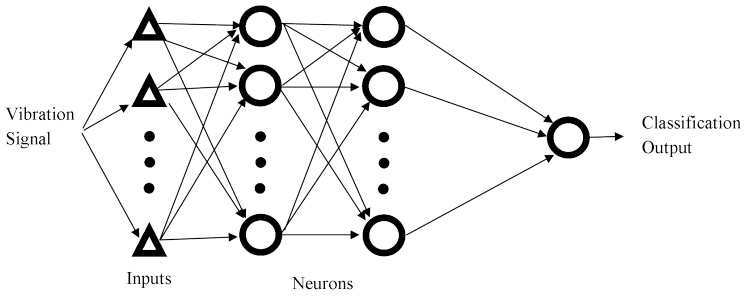


Figure 18: A 2-Layer deep belief network. For this study, the inputs (triangles) are the data points corresponding to the Fourier transform of the vibration signal.

We then constructed networks which combined anterior-posterior and superior-inferior data. Based on a previous study in deep learning [173] we divided our networks into two parts. The first stage consisted of two independent networks of equal size that each operated on one vibration signal. The labels are identical to those used previously, so a 'small 2-layer network' would indicate that both the anterior-posterior and superior-inferior vibrations would serve as input to two independent, 2-layer networks with 300 neurons in each layer. The second stage of these combined networks consisted of a single neural network of one, two, or three layers with 1000 neurons in each layer. The outputs of the first stage were used as the inputs to this second stage. The system is illustrated more clearly in figure 19. In total, 18 of these combined networks were built and tested for our study. Other researchers have demonstrated that combining multiple correlated signals, such as both sound and video data, into a single input vector to a deep belief network results in minimal interaction between the neurons corresponding to each input [173]. However, by allowing each input to be processed by their own networks before combining the in a third, the combination network is better able to identify interactions at higher-orders [173]. We employed this bimodal deep belief network since the higher order relations are of particular interest to us.

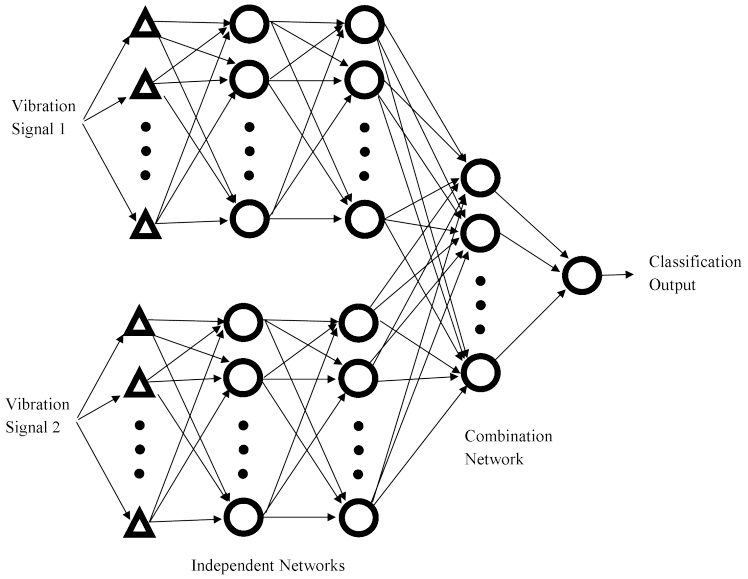


Figure 19: One of our combined networks, containing a pair of 2-Layer deep belief networks with an additional 1-Layer combination network. Note that the independent networks, which each analyse one of our two vibration signals, do not interact until reaching the combination network. For this study, the inputs (triangles) are the data points corresponding to the Fourier transform of the vibration signal.

For all of our networks we used a learning rate of 0.05. This was found, through trial and error, to provide a relatively steady and non-chaotic rate of weight adjustment for the size of our training set. It also demonstrated a minimal amount of over-tuning of the model when the networks were tested with the training data set. Similarly, we used logistic sigmoid activation functions for all of the neurons in all of our networks. This function is smooth, differentiable, and positive at all points which should minimize any potential difficulties with implementing our training algorithms.

8.3 RESULTS

Tables 16-18 provide the results of our tests. Tables 16 and 17 present the results for networks which utilized a single vibration axis' data as its input and contained either 300 or 3000 neurons in each layer, respectively. Table 18 presents the corresponding results when these two independent networks were merged by a third network with 1000 neurons in each layer. The number of correctly classified healthy and unhealthy swallows (of the 123 presented for each category) are given as the average of our ten-fold cross validation procedure. Sensitivity is defined as the percentage of swallows from patients with dysphagia that were correctly identified as such, while specificity is the percentage of correctly classified swallows from healthy subjects. Accuracy is the overall percentage of correctly classified swallows.

Table 16: Small network performance classifying healthy and safe swallows

	Classified Correctly		Statistical Metrics		
	Healthy	Dysphagia	Specificity	Sensitivity	Accuracy
AP 1-Layer	107.3	100.2	87.2	81.5	84.3
SI 1-Layer	120.3	84.1	97.8	68.4	83.1
AP 2-Layer	105.8	97.8	86.0	79.5	82.8
SI 2-Layer	121.0	77.8	98.4	63.3	80.8
AP 3-Layer	106.6	99.7	86.7	81.1	83.9
SI 3-Layer	120.2	74.5	97.7	60.6	79.1

We found that single-axis networks demonstrate overall accuracies between 76 and 86 percent. This was generally the result of very high specificity (95% or higher for many networks) paired with a much lower sensitivity (72% or less for many networks). Networks that had a greater number of neurons generally had lower sensitivities but greater specificities. This resulted in higher accuracy for anterior-posterior networks with more neurons, but lower overall accuracy for superior-inferior networks. We also found that deep belief networks

Table 17: Large network performance classifying healthy and safe swallows

	Classified Correctly		Statistical Metrics		
	Healthy	Dysphagia	Specificity	Sensitivity	Accuracy
AP 1-Layer	120.1	88.2	97.6	71.7	84.7
SI 1-Layer	114.9	76.5	93.4	62.2	77.8
AP 2-Layer	122.0	88.4	99.2	71.9	85.5
SI 2-Layer	117.7	69.4	95.7	56.4	76.1
AP 3-Layer	122.1	89.8	99.3	73.0	86.1
SI 3-Layer	122.5	67.4	99.6	54.8	77.2

provided little to no accuracy improvement beyond the simple, single layer neural network configuration, though the additional layers did tend to improve specificity at the cost of sensitivity.

The accuracies of our combined, deep belief networks was generally greater than those offered by the single-axis networks, ranging from 80 to 91 percent. While the specificity of these networks is somewhat lower than their single-axis counterparts, the sensitivity is noticeably increased. These networks demonstrated the same neuron number dependence as the single-axis networks, where networks with more neurons had higher specificity but lower sensitivity. In this case, however, overall accuracy remained nearly unchanged.

The networks exhibited marginally better accuracies when classifying the training data compared to the test data set. Single axis networks demonstrated overall accuracies between 80 and 89 percent while combined networks exhibited between 88 and 92 percent accuracies. No networks demonstrated more than a 3% improvement in classification accuracy when analysing the training data set.

Table 18: Combined network performance classifying healthy and safe swallows

Structure		Classified Correctly		Statistical Metrics		
Combination Network	Input Network	Healthy	Dysphagia	Specificity	Sensitivity	Accuracy
1-Layer	Small 1-Layer	111.5	103.0	90.7	83.7	87.2
	Small 2-Layer	114.8	100.4	93.3	81.6	87.5
	Small 3-Layer	111.8	101.1	90.9	82.2	86.5
	Large 1-Layer	114.4	100.5	93.0	81.7	87.4
	Large 2-Layer	119.2	105.4	96.9	85.7	91.3
	Large 3-Layer	121.2	92.6	98.5	75.3	86.9
2-Layer	Small 1-Layer	116.0	99.3	94.3	80.7	87.5
	Small 2-Layer	113.1	102.5	92.0	83.3	87.6
	Small 3-Layer	107.6	103.6	87.5	84.2	85.9
	Large 1-Layer	119.9	98	97.5	79.7	88.6
	Large 2-Layer	118.3	99.8	96.2	81.1	88.7
	Large 3-Layer	120.3	95.5	97.8	77.6	87.7
3-Layer	Small 1-Layer	112.1	103	91.1	83.7	87.4
	Small 2-Layer	113.2	102	92.0	82.9	87.5
	Small 3-Layer	102.4	95.2	83.3	77.4	80.3
	Large 1-Layer	119.9	96.3	97.5	78.3	87.9
	Large 2-Layer	117.3	96.0	95.4	78.0	86.7
	Large 3-Layer	121.2	93.3	98.5	75.9	87.2

8.4 DISCUSSION

Our study varied from past research on swallowing classification in a number of ways. Namely, we included a larger number of participants and swallow events as well as a much wider array of boluses and swallowing techniques. Despite this we see that our networks, particularly our single layer networks, provide similar swallow classification accuracy to that reported by several other studies [39, 42, 65, 134, 139]. In addition, our networks demonstrated only a minimal amount of over-tuning of the model parameters when classifying the training data set. This indicates that our network design methods are valid and that we can accurately compare our deep belief network results to other work in the field.

Unlike some previous studies, the results provided in this study are relative to the classification of individual swallows from a generic source rather than classification of individual patients. This is due to the limits of recording data during a modified barium swallow test, which does not guarantee that all bolus types are presented to all patients. As a result, we could not formulate a 'majority opinion rule' for classifying a patient's condition based on the classification of individual swallows, as seen in similar works. However, considering that the per-patient classification rate was higher than the per-swallow classification rate in studies that reported such information [39, 134, 139], the results of our study for individual swallows should provide a lower-limit for the classification of patients in typical clinical settings. Even so, our technique shows considerable potential improvement over existing, non-instrumental methods. These deep learning networks demonstrate similar or superior sensitivity and specificity compared to tests such as the Toronto bedside assessment or modified MASA [32, 174]. In summary, this study has demonstrated that deep learning combined with cervical auscultation has the potential to be a viable classification technique for dysphagia.

For single-axis networks, we found that using a multi-layer deep belief network configuration provided little to no benefit with respect to the overall classification accuracy. This indicates that higher-order features from a single vibration direction do not provide any additional classification information on their own. However, combining information from both axes in a multi-layer configuration did provide a significant increase in overall accuracy as

well as more than a 5% increase in maximum observed accuracy. In particular, the highest accuracies (more than 88%) were observed for the 'large, 2-layer' network with either a one or two layer combination and the 'large 1-layer' network with a two layer combination. This demonstrates that higher-order relationships between these two vibration signals do provide information important to classification of swallowing signals. Furthermore, due to structural limitations, these higher-order relationships cannot be obtained with a single layer network. As a result, we can conclude that multi-layer deep belief networks can provide significant improvements to swallowing classification when analysing multi-dimensional signals.

Our combined deep belief network configuration also demonstrated generally greater sensitivity than our single-axis networks. While all metrics used to assess a screening technique's performance are valuable, sensitivity is arguably of greater importance. Should this method of classification be used for a real-world application, it would likely be used in a clinical setting. Patients with swallowing difficulties are encountered much more often in this situation than in the general population and assessing their swallowing ability correctly is of greater importance. As a result, a classification's sensitivity (ability to identify swallows from a subject with dysphagia) is far more valuable than its specificity (ability to identify swallows from a healthy subject). This suggests that multi-layer deep belief network classification would be of greater benefit than single layer network classification specifically when utilized in a real-world setting, but further research would need to be done before any definitive conclusions can be made.

8.4.1 Limitations

Swallows included in this study were gathered in a typical clinical setting. This is not necessarily the optimal environment for data intended for mathematical analysis. It can introduce a number of confounding variables in the form of environmental factors and can result in a uneven distribution of variables, such as the number of swallows performed by each subject. However, if cervical auscultation is to be used as a screening method it would need to operate in such an environment. Testing our classification method on highly controlled and perfectly distributed data may result in much clearer results, but would not provide an

accurate assessment of its practical capabilities. In addition, this study also only included swallows that did not result in significant laryngeal penetration. This is because we focused chiefly on differentiating swallows from healthy subjects, who rarely if ever swallow with laryngeal penetration, and patients with dysphagia. Including penetrating swallows from only patients with dysphagia would greatly bias the training of the network and could easily result in a configuration with poor generalizability.

8.5 CONCLUSION

In this study, we sought to differentiate swallows made by healthy subjects from those made by patients with dysphagia using only cervical auscultation signals. To do this, we used the frequency spectrums of anterior-posterior and superior-inferior swallowing vibrations as inputs to a variety of single and multi-layer and deep belief networks. We found that single layer networks provided the greatest overall accuracy when analysing vibrations from a single axis. However, when incorporating information from both axes simultaneously, multi-layer deep belief networks offered a notable increase in overall accuracy and sensitivity. We conclude that higher-order features contain valuable information when analysing multi-dimensional swallowing signals and that multi-layer deep belief networks should be considered when classifying such data.

9.0 CLASSIFICATION OF UNSAFE SWALLOWS

9.1 MOTIVATION

Cervical auscultation would be most useful as a screening method for dysphagia if it could function autonomously so that it might provide a more objective assessment than existing methods. One subject that is difficult to monitor with existing screening and diagnostic methods is the occurrence of unsafe swallows during ordinary behaviour. As a small, non-invasive, and passive screening method, cervical auscultation could provide a wealth of valuable information by identifying when these swallows occur. However, as with differentiating healthy and dysphagic swallows, research into automated classification of these swallows has been limited with a few notable drawbacks. By using a broader selection of data and a non-linear classification method, this study attempts to provide a superior method of differentiating safe from unsafe swallows in patients with known swallowing difficulties.

9.2 METHODOLOGY

The methodology for this study, from data collection to design of our deep belief networks, is nearly identical to that used in our previous study which was described in section 8.2. The exceptions to our data set are as follows: We used swallowing vibration data that originated exclusively from subjects with known swallowing difficulties. Thirty-eight of these subjects (29 males, 9 females, mean age 68) had a primary diagnosis of a stroke while 124 (78 males, 46 females, mean age 62) had no history of stroke. Swallows with a PA-score of 4 or greater (199 in total) were declared 'unsafe' swallows while all wallows with a PA-score of 3 or less (1465

in total) were declared 'safe' swallows. A random selection of 360 safe and unsafe swallows (180 from each group) were designated as the training data set while 38 were designated as the test data set.

Our deep belief network formulation and structure remained unchanged with one exception. The learning rate of our networks was increased to a value of 1 in order to account for the reduced number of swallows included in this study.

We also chose to compare the results obtained with our networks to those obtain with an existing algorithm and utilizes conditional density estimates [44]. This algorithm demonstrated high overall accuracy in identifying safe and unsafe swallows in a previous study, but has not been fully tested on a less restricted data set. We used the average of the top five algorithm configurations detailed in the previous study [44] to provide a baseline for the performance of our networks.

9.3 RESULTS

Tables 19-22 provide the results of our tests. Tables 19 and 20 present the results for networks which utilized a single vibration axis' data as its input and contained either 300 or 3000 neurons in each layer, respectively. Table 21 presents the corresponding results when these two independent networks were merged by a third network with 1000 neurons in each layer. Table 22 summarizes the results given by the conditional density algorithm presented in a previous study. The entries for the categories of 'Wavelets', 'Features', and 'Levels' are given in the for of AP/SI, where the first entry corresponds to the parameters used with the anterior-posterior vibration signals, while the second corresponds to the same for superior-inferior signals. The 'log-energy' feature has been abbreviated to 'log-E' along with 'Coiflet' being written as 'Coif' and 'entropy' as 'ent'. The number of correctly classified safe and unsafe swallows (of the 19 presented for each category) are given as the average of our ten-fold cross validation procedure. Sensitivity is defined as the percentage of safe swallows that were correctly identified as such, while specificity is the percentage of correctly classified unsafe swallows. Accuracy is the overall percentage of correctly classified swallows.

Table 19: Small network performance classifying safe and unsafe swallows

	Classified Correctly		Statistical Metrics		
	Safe	Unsafe	Specificity	Sensitivity	Accuracy
AP 1-Layer	14.9	12.1	78.4	63.7	71.1
SI 1-Layer	14.7	10.2	77.4	53.7	65.5
AP 2-Layer	15.2	12.1	80.0	63.7	71.8
SI 2-Layer	14.6	11.4	76.8	60.0	68.4
AP 3-Layer	15.0	12.3	86.7	64.7	71.8
SI 3-Layer	14.6	10.8	76.8	56.8	66.8

Table 20: Large network performance classifying safe and unsafe swallows

	Classified Correctly		Statistical Metrics		
	Safe	Unsafe	Specificity	Sensitivity	Accuracy
AP 1-Layer	15.0	12.7	78.9	66.8	72.9
SI 1-Layer	14.9	10.9	78.4	53.4	67.9
AP 2-Layer	14.7	12.6	77.4	66.3	71.8
SI 2-Layer	14.8	11.3	77.9	59.5	68.7
AP 3-Layer	14.6	12.6	76.8	66.3	71.6
SI 3-Layer	14.9	11.2	78.4	58.9	68.7

We found that single-axis networks demonstrate overall accuracies between 65 and 73 percent. We were unable to find any notable trends in the single-axis data aside from specificity being consistently greater than sensitivity and AP-only networks performing better than SI-only networks.

Table 21: Combined network performance classifying safe and unsafe swallows

Structure		Classified Correctly		Statistical Metrics		
Combination Network	Input Network	Safe	Unsafe	Specificity	Sensitivity	Accuracy
1-Layer	Small 1-Layer	15.1	13.9	79.5	73.2	76.3
	Small 2-Layer	14.8	13.0	77.9	68.4	73.2
	Small 3-Layer	14.5	13.5	76.3	71.1	73.7
	Large 1-Layer	14.7	13.9	77.4	73.2	75.3
	Large 2-Layer	14.9	13.5	78.4	71.1	74.7
	Large 3-Layer	14.2	13.9	74.7	73.2	73.9
2-Layer	Small 1-Layer	14.6	13.5	76.8	71.1	73.9
	Small 2-Layer	14.5	13.7	76.3	72.1	74.2
	Small 3-Layer	14.1	13.7	74.2	72.1	73.2
	Large 1-Layer	15.2	14.1	80.0	74.2	77.1
	Large 2-Layer	14.9	13.8	78.4	72.6	75.5
	Large 3-Layer	14.7	13.8	77.4	72.6	75.0
3-Layer	Small 1-Layer	15.0	14.0	78.9	73.7	76.3
	Small 2-Layer	14.3	13.7	75.3	72.1	73.7
	Small 3-Layer	14.7	13.9	77.4	73.2	75.3
	Large 1-Layer	14.2	12.7	74.7	66.8	70.8
	Large 2-Layer	14.3	13.9	75.3	73.2	74.2
	Large 3-Layer	14.0	14.0	73.7	73.7	73.7

The accuracies of our combined, deep belief networks was generally greater than those offered by the single-axis networks, ranging from 70 to 77 percent. While the specificity of these networks is somewhat lower than their single-axis counterparts in some situations, the sensitivity is noticeably increased. We were unable to identify any other consistent trends.

Table 22: Conditional density classification performance

Wavelets	Features	Levels	Classified Correctly		Statistical Metrics		
			Safe	Unsafe	Specificity	Sensitivity	Accuracy
Coif5/Coif5	log-E/log-E	8/2	14.8	13.7	77.9	72.3	75.1
Meyer/Coif1	ent/log-E	2/8	14.1	14.3	74.2	75.0	74.6
Coif3/Coif1	ent/log-E	2/8	14.2	14.2	74.7	74.5	74.6
Coif5/Coif1	ent/log-E	2/8	14.4	14.0	75.8	73.8	74.8
Coif5/Coif3	log-E/log-E	8/2	14.9	13.8	78.4	72.4	75.4

The networks exhibited marginally better accuracies when classifying the training data compared to the test data set. Single axis networks demonstrated overall accuracies between 68 and 77 percent while combined networks exhibited between 75 and 80 percent accuracies. No networks demonstrated more than a 4% improvement in classification accuracy when analysing the training data set.

The conditional density algorithm demonstrated an average specificity of 76.2, sensitivity of 73.6, and accuracy of 74.9. Unlike in the previous study with this algorithm [44] which had a clear outlier, the results were fairly consistent regardless of the algorithm's configuration on this data set with no accuracy measure varying by more than one percent.

9.4 DISCUSSION

Our study varied from past research on swallowing classification in a number of ways. Namely, we included a larger number of participants and swallow events as well as a much wider array of boluses and swallowing techniques. Due to this, we note that our conditional density algorithm did not perform as well with this data set as it did on a more controlled

data set [44]. However, its performance with this new data set was comparable to the performance of our combined networks. In addition to the minimal amount of over-tuning of our model parameters, we feel that this demonstrates the validity of our results and methodology.

Regardless, the reduced performance of our classifiers on this data set is an important development. Past studies, most notably that which proposed the conditional density algorithm [44], often utilize a data set consisting of a limited selection of swallows. Our past work has demonstrated how variables such as viscosity [38] can affect cervical auscultation signals. It is possible, and perhaps likely, that the physiological effects of an unsafe swallow are overshadowed by concurrent effects due to these other factors and results in a reduction in classification performance. It is also possible that unsafe swallows made while in an alternate head position or with boluses of varying viscosities are more difficult to differentiate than the thin liquid, neutral head position swallows made in the previous study [44]. The preliminary work we have performed on this subject, detailed in section 7, suggested that any differences between safe and unsafe swallows would be rather small due to the highly variable nature of the two groups, so such concurrent factors may play a larger role than anticipated. Despite these differences, we note that both of the classification techniques utilized in this study produced comparable results to those reported by existing bedside swallowing screening procedures [32, 174]. This is significant because our data set, which was gathered during routine clinical exams, is similar to the data that would be analysed by this method in a real-world setting. This reinforces the validity of cervical auscultation as a method of swallowing assessment and the potential use of deep belief networks for swallowing signal classification.

As with our attempt at classifying healthy and dysphagic swallows in section 8, using a multi-layer deep belief network provided minimal benefit with regard to single axis classification performance. However, the combined networks likewise did offer a notable improvement in both sensitivity and overall classification accuracy. In particular, the highest classification accuracy was achieved with a 'large, 1-layer' network with a 2-layer combination. This again demonstrates the importance of both the use of information from both vibration axes as well as the use of higher order features and relationships between these two signals which cannot

be obtained with a single layer network. Furthermore, the improved sensitivity (ability to identify unsafe swallows) of our combined networks is of greater benefit with regards to the clinical situation, as identifying when such swallows occur is one of the key points of swallowing assessment. As an added benefit, we note that the best performing deep belief networks were able to out-perform the existing conditional density algorithm, though the perceived benefit is small and leaves much room for improvement. These facts again suggest that multi-layer deep belief network classification would be of greater benefit than single layer network classification specifically when utilized in a real-world setting, but further research would need to be done before any definitive conclusions can be made.

9.5 CONCLUSION

In this study, we sought to differentiate swallows that were performed safely from those that were unsafe and resulted in significant penetration of a swallowed bolus into the larynx. To do this, we used the frequency spectrums of anterior-posterior and superior-inferior swallowing vibrations as inputs to a variety of single and multi-layer and deep belief networks. We found that multi-layer deep belief networks provided the greatest overall classification performance and conclude that higher-order features contain valuable information about swallowing. However, we note the heterogeneous nature of our data set and the reduced performance of a comparable classification technique, suggesting the need to refine our general classification methodology.

10.0 RESEARCH SUMMARY

10.1 CONCLUSIONS

Swallowing disorders affect thousands of people each year and are one of the greatest issues faced by those with neurological impairments or trauma. Despite them being much easier to overlook, swallowing difficulties can develop into dangerous and harmful conditions if ignored to left unnoticed. As a result, identifying a method of swallowing assessment that has a high accuracy, but can also be deployed easily in the clinic has the potential to help many individuals. The research described in this manuscript attempted to justify and advance the use of cervical auscultation signals for such an automated assessment technique. While some advancements have been made on the subject, past research has been relatively shallow and disorganized and so a comprehensive overview of the technique could help to advance progress in the field. Specifically, we sought to characterize healthy, safe, and unsafe swallows made by adults with and without dysphagia as well as develop methods to automatically classify these groups. We also attempted to provide superior methods of denoising and segmenting in order to improve the processing methods used with cervical auscultation signals.

Ultimately, we were successful in our stated goals. We were able to characterize swallows made by healthy subjects as well as those with dysphagia using cervical auscultation signals and a variety of statistical features. We were also able to mathematically identify a number of different ways in which these signals varied, both with respect to each other and with respect to several external variables. Furthermore, we demonstrated that it is possible to differentiate these signals mathematically and achieve accuracies comparable to that of existing clinical screening methods while remaining completely blind to the results of any clinical examination. Finally, we offered some possible solutions and improvements to exist-

ing signal processing methodology which should allow for more accurate and clearer data in future applications. All of these items justify our key points: that cervical auscultation can be used to assess swallowing performance as an independent technique, and that this method has the potential to improve upon and offer new insights about swallowing physiology and existing swallowing assessment techniques. As a result, we advocate for the dedication of future work to refining this swallowing assessment method and the development of a formal, cervical auscultation based, swallowing assessment procedure.

Cervical auscultation as a swallowing assessment technique is still in the experimental phase. However, as this manuscript demonstrated, it does have the potential to be utilized in a greater context. Should future efforts to develop this technique be successful, the low hardware requirements of cervical auscultation ensures that it could easily be implemented alongside existing screening or diagnostic procedures in the clinical setting. This manuscript and other studies have demonstrated its potential to be deployed in a completely automated manner, and so cervical auscultation could provide an objective clinical assessment against which existing, subjective assessment techniques could be compared. Overall, this should allow swallowing disorders to be identified sooner and with greater certainty and ensure that this easily manageable condition does not cause undue harm on patients.

10.2 FUTURE WORK

There are a number of different areas that would be useful to investigate in future studies. First, would be the use of alternate signals. In all of the experiments detailed here we utilized anterior-posterior and superior-inferior swallowing vibrations, with only minor use of swallowing sounds for feature analysis. However, we have not put the same amount of effort into developing filters and classifiers for swallowing sound data. Past work [37] has shown that these signals provide unique information. As a result, we should be able to improve our analysis by refining algorithms to operate with swallowing sound data as well

as vibrations. It may also be worth investigating using medial-lateral vibrations, axes or rotation, or composite vectors as well, since these signals have not been studied as intensively as the vibrations chosen for our work.

In future studies, it would also be useful to refine upon the methods detailed in this work. The algorithms presented here were generally shown to offer a small improvement upon existing methodology or demonstrate the potential validity of the method. However, there is still a gap between what has been achieved here, be it in removal of noise from a signal, features selected for characterizing swallows, or otherwise, and what is possible or desirable for a clinical assessment technique. Further iterations could be made to improve upon the algorithms presented here, or alternative methods may be investigated and compared to these techniques.

Finally, there is still much work to be done in relating swallowing physiology to cervical auscultation signals. There have been a few studies that attempted to relate the clinically observed swallowing sounds as heard via a stethoscope to simple cervical auscultation signal features. However, there has not been a large, comprehensive effort to directly relate these two transduction methods despite the volumes of research done on either independent technique. Being able to provide a proper mathematical representation of swallows and swallowing events would provide great benefits to the study of both real and digital swallowing signals and could lead to dramatic improvements in swallowing assessment.

BIBLIOGRAPHY

[1] M. Spieker, "Evaluating dysphagia," *American Family Physician*, vol. 61, no. 12, pp. 3639–3648, June 2000.

[2] D. Castell and M. Donner, "Evaluation of dysphagia: A carcare history is crucial," *Dysphagia*, vol. 2, no. 1, pp. 65–71, January 1987.

[3] J. Logemann, P. Kahrilas, J. Cheng, B. Pauloski, P. Gibbons, A. Rademaker, and S. Lin, "Closure mechanisms of laryngeal vestibule during swallow," *American Journal of Physiology*, vol. 262, no. 2, pp. G338–G344, February 1992.

[4] D. Smithard, P. O'Neill, C. Park, J. Morris, R. Wyatt, R. England, and D. Martin, "Complications and outcome after acute stroke. does dysphagia matter?" *Stroke*, vol. 27, no. 7, pp. 1200–1204, July 1996.

[5] P. Leslie, M. Drinnan, I. Zammit-Maempel, J. Coyle, G. Ford, and J. Wilson, "Cervical auscultation synchronized with images from endoscopy swallow evaluations," *Dysphagia*, vol. 22, no. 4, pp. 290–298, October 2007.

[6] A. Castrogiovanni, "Communication facts: Special population: Dysphagia," American Speech-Language-Hearing Association, 2200 Research Boulevard, Rockville, MD 20850, Tech. Rep., January 2008.

[7] S. Orenstein, F. Izadnia, and S. Khan, "Gastroesophageal reflux disease in children," *Gastroenterology Clinics of North America*, vol. 28, no. 4, pp. 947–969, December 1999.

[8] D. Ramsey, D. Smithard, and L. Kalra, "Silent aspiration: What do we know?" *Dysphagia*, vol. 20, no. 1, pp. 218–225, January 2005.

[9] B. Garon, T. Sierzant, and C. Ormiston, "Silent aspiration: Results of 2,000 video fluoroscopic evaluations," *Journal of Neuroscience Nursing*, vol. 41, no. 4, pp. 178–185, August 2009.

[10] H. Gray, *Anatomy of the Human Body*, 20th ed., W. Harmon, Ed. Lea & Febiger, 1918.

[11] W. Dodds, E. Stewart, and J. Logemann, "Physiology and radiology of the normal oral and pharyngeal phases of swallowing," *American Journal of Roentgenology*, vol. 154, no. 5, pp. 953–963, May 1990.

[12] G. Malandraki, B. Sutton, A. Perlman, D. Karampinos, and C. Conway, "Neural activation of swallowing and swallowing-related tasks in healthy young adults: An attempt to separate the components of deglutition," *Human Brain Mapping*, vol. 30, no. 10, pp. 3209–3226, October 2009.

[13] J. Bosma, "Deglutition: Pharyngeal stage," *Physiological Reviews*, vol. 37, no. 3, pp. 275–300, July 1957.

[14] R. Goyal and H. Mashimo, "Physiology of oral, pharyngeal, and esophageal motility," *GI Motility Online*, vol. 1, no. 1, pp. 1–5, May 2006.

[15] R. Flaherty, S. Seltzer, T. Campbell, R. Weisskoff, and R. Gilbert, "Dynamic magnetic resonance imaging of vocal cord closure during deglutition," *Gastroenterology*, vol. 109, no. 3, pp. 843–849, September 1995.

[16] R. Martino, N. Foley, S. Bhogal, N. Diamant, M. Speechley, and R. Teasell, "Dysphagia after stroke: Incidence, diagnosis and pulmonary complications," *Stroke*, vol. 36, no. 12, pp. 2756–2763, November 2005.

[17] I. Cook and P. Kahrilas, "American gastroenterological association technical review on management of oropharyngeal dysphagia," American Gastroenterological Association, Tech. Rep. 116 (2), January 1999.

[18] J. Palmer, J. Drennan, and M. Baba, "Evaluation and treatment of swallowing impairments," *American Family Physician*, vol. 61, no. 8, pp. 2453–2462, April 2000.

[19] J. Logemann, "The evaluation and treatment of swallowing disorders," *Otolaryngology and Head and Neck Surgery*, vol. 6, no. 1, pp. 395–400, 1998.

[20] J. L. Coyle and J. Robbins, "Assessment and behavioral management of oropharyngeal dysphagia," *Otolaryngology and Head and Neck Surgery*, vol. 5, no. 1, pp. 147–152, 1997.

[21] J. Logemann and G. Shelley, "Should treatment for pharyngeal swallowing disorders begin before instrumental assesment is completed," *American Speech-Language-Hearing Association*, vol. 38, no. 4, p. 14, September 1996.

[22] M. Rugiu, "Role of videofluoroscopy in evaluation of neurologic dysphagia," *ACTA Otorhinolaryngologica Italica*, vol. 27, no. 6, pp. 306–316, December 2007.

[23] M. Rashid, "Case 1: Diagnosing difficult deglutition," *Paedicatrics and Child Health*, vol. 14, no. 7, pp. 453–454, September 2009.

[24] R. Bastian, "The videoendoscopic swallowing study: An alternative and partner to the videofluoroscopic swallowing study," *Dysphagia*, vol. 8, no. 4, pp. 359–367, September 1993.

[25] S. Langmore, K. Schatz, and N. Olsen, "Fiberoptic endoscopic examination of swallowing safety: A new procedure," *Dysphagia*, vol. 2, no. 4, pp. 216–219, 1988.

[26] A. Kelly, P. Leslie, T. Beale, C. Payten, and M. Drinnan, "Fiberoptic endoscopic evaluation of swallowing and videofluoroscopy: Does examination type influence perception of pharyngeal residue severity?" *Clinical Otolaryngology*, vol. 31, no. 5, p. 5, October 2006.

[27] C. Ertekin, I. Aydogdu, N. Yuceyar, S. Tarlaci, N. Kiylioglu, M. Pehlican, and G. Celebi, "Electrodiagnostic methods for neurogenic dysphagia," *Electroencephalography amd Clinical Neurophysiology/Electromyography and Motor Control*, vol. 109, no. 4, pp. 331–340, August 1998.

[28] R. Ding, C. Larson, J. Logemann, and A. Rademaker, "Surface electromyographic and electroglottographic studies in normal subjects under two swallow conditions: Normal and during the mendelsohn manuever," *Dysphagia*, vol. 17, no. 1, pp. 1–12, January 2002.

[29] P. Zenner, D. Losinski, and R. Mills, "Using cervical auscultation in the clinical dysphagia examination in long-term care," *Dysphagia*, vol. 10, no. 1, pp. 27–31, January 1995.

[30] P. Leslie, M. Drinnan, P. Finn, G. Ford, and J. Wilson, "Reliability and validity of cervical auscultation: A controlled comparison using videofluoroscopy," *Dysphagia*, vol. 19, no. 4, pp. 231–240, 2004.

[31] D. Suiter, S. Leder, and D. Karas, "The 3-ounce (90-cc) water swallow challenge: A screening test for children with suspected oropharyngeal dysphagia," *Otolaryngology and Head and Neck Surgery*, vol. 140, no. 2, pp. 187–190, February 2009.

[32] R. Martino, F. Silver, R. Teasell, M. Bayley, G. Nicholson, D. Streiner, and N. Diamant, "The toronto bedside swallowing screening test (tor-bsst): Development and validation of a dysphagia screening tool for patients with stroke," *Stroke*, vol. 40, no. 2, pp. 555–561, February 2009.

[33] J. Edmiaston, L. Connor, L. Loehr, and A. Nassief, "Validation of a dysphagia screening tool in acute stroke patients," *American Journal of Critical Care*, vol. 19, pp. 357–364, 2009.

[34] J. Cichero, S. Heaton, and L. Bassett, "Triaging dysphagia: Nurse screening for dysphagia in an acute hospital," *Journal of Clinical Nursing*, vol. 18, pp. 1649–1659, 2009.

[35] C. Pettigrew and C. O'Toole, "Dysphagia evaluation practices of speech and language therapists in ireland: Clinical assessment and instrumental examination decision-making," *Dysphagia*, vol. 22, pp. 235–244, 2007.

[36] B. Mathers-Schmidt and M. Kurlinski, "Dysphagia evaluation practices: Inconsistencies in clinical assessment and instrumental examination decision-making," *Dysphagia*, vol. 18, pp. 114–125, 2003.

[37] J. M. Dudik, I. Jestrović, B. Luan, J. L. Coyle, and E. Sejdić, "A comparative analysis of swallowing accelerometry and sounds during saliva swallows," *Biomedical Engineering Online*, vol. 14, no. 3, pp. 1–15, January 2015.

[38] I. Jestrović, J. Dudik, B. Luan, J. Coyle, and E. Sejdić, "The effects of increased fluid viscosity on swallowing sounds in healthy adults," *Biomedical Engineering Online*, vol. 12, no. 90, pp. 1–17, September 2013.

[39] L. Lazareck and Z. Moussavi, "Classification of normal and dysphagic swallow by acoustical means," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2103–2112, December 2004.

[40] M. Celeste, K. Azadeh, E. Sejdiċ, G. Berall, and T. Chau, "Quantitative classification of pediatric swallowing through accelerometry," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 34, pp. 1–8, June 2012.

[41] C. Steele, E. Sejdić, and T. Chau, "Noninvasive detection of thin-liquid aspiration using dual-axis swallowing accelerometry," *Dysphagia*, vol. 28, no. 1, pp. 105–112, March 2013.

[42] A. Spadotto, A. Gatto, R. Guido, A. Montagnoli, P. Cola, and J. P. A. Schelp, "Classification of normal swallowing and oropharyngeal dysphagia using wavelet," *Applied Mathematics and Computation*, vol. 207, no. 1, pp. 75–82, January 2009.

[43] M. Nikjoo, C. Steele, E. Sejdić, and T. Chau, "Automatic discrimination between safe and unsafe swallowing using a reputation-based classifier," *Biomedical Engineering Online*, vol. 10, no. 100, pp. 1–17, November 2011.

[44] E. Sejdić, C. Steele, and T. Chau, "Classification of penetration – aspiration versus healthy swallows using dual-axis swallowing accelerometry signals in dysphagic subjects," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1859–1866, July 2013.

[45] E. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, and M. Neuman, "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 626–633, March 2010.

[46] J. Lee, C. Steele, and T. Chau, "Classification of healthy and abnormal swallows based on accelerometry and nasal airflow signals," *Artificial Intelligence in Medicine*, vol. 52, no. 1, pp. 17–25, May 2011.

[47] S. Sarraf-Shirazi and Z. Moussavi, "Silent aspiration detection by breath and swallowing sound analysis," in *The 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, San Diego, CA, August 28 - September 1 2012, pp. 2599–2602.

[48] A. Das, N. Reddy, and J. Narayanan, "Hybrid fuzzy logic committee neural networks for recognition of swallow acceleration signals," *Computer Methods and Programs in Biomedicine*, vol. 64, no. 2, pp. 87–99, February 2001.

[49] O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, and M. Neuman, "Limited receptive area neural classifier for recognition of swallowing sounds using continuous wavelet transform," in *The 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, FR, August 22-26 2007, pp. 3128–3131.

[50] O. Makeyev, E. Sazonov, S. Schuckers, P. Lopez-Meyer, T. Baidyk, E. Melanson, and M. Neuman, *Recognition of Swallowing Sounds Using Time-Frequency Decomposition and Limited Receptive Area Neural Classifier*, T. Allen, R. Ellis, and M. Petridis, Eds. Springer London, 2009, vol. 15, no. 1.

[51] J. Fontana, P. Melo, and E. Sazonov, "Swallowing detection by sonic and subsonic frequencies: A comparison," in *The 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, Boston, MA, August 30 - September 3 2011, pp. 6890–6893.

[52] L. Lazareck and Z. Moussavi, "Swallowing sound characteristics in healthy and dysphagic individuals," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, San Francisco, CA, September 1-5 2004, pp. 3820–3823.

[53] M. Aboofazeli and Z. Moussavi, "Analysis of temporal pattern of swallowing mechanism," in *The 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, New York, NY, August 30 - September 3 2006, pp. 5591–5594.

[54] E. Sejdić, C. M. Steele, and T. Chau, "Segmentation of dual-axis swallowing accelerometry signals in healthy subjects with analysis of anthropometric effects on duration of swallowing activities," *IEEE Transactions of Biomedical Engineering*, vol. 56, no. 4, pp. 1090–1097, April 2009.

[55] J. M. Dudik, J. L. Coyle, and E. Sejdić, "Dysphagia screening: Contributions of cervical auscultation signals and modern signal processing techniques," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 465–477, August 2015, ©2015 IEEE. Reprinted, with permission.

[56] S. Hamlet, D. Penney, and J. Formolo, "Stethoscope acoustics and cervical auscultation of swallowing," *Dysphagia*, vol. 9, no. 1, pp. 63–68, January 1994.

[57] J. Logemann, S. Veis, and L. Colangelo, "A screening procedure for oropharyngeal dysphagia," *Dysphagia*, vol. 14, no. 1, pp. 44–51, January 1999.

[58] J. Hinchey, T. Shephard, K. Furie, D. Smith, D. Wang, S. Tonn, and S. P. I. N. Investigators, "Formal dysphagia screening protocols prevent pneumonia," *Stroke*, vol. 36, no. 9, pp. 1972–1976, September 2005.

[59] C. Hey, B. Lange, S. Eberle, Y. Zaretsky, R. Sader, T. Stöver, and J. Wagenblast, "Water swallow screening test for patients after surgery for head and neck cancer: Early identification of dysphagia, aspiration, and limitations of oral intake," *Anticancer Research*, vol. 33, no. 9, pp. 4017–4027, September 2013.

[60] C. Borr, M. Hielscher-Fastabend, and A. Lucking, "Reliability and validity of cervical auscultation," *Dysphagia*, vol. 22, no. 3, pp. 225–234, July 2007.

[61] J. Marrara, A. Duca, R. Dantas, L. Trawitzki, R. Cardozo, and J. Pereira, "Swallowing in children with neurologic disorders: Clinical and videofluoroscopic evaluations," *Pro-Fono Revista de Atualizacao Cientifica*, vol. 20, no. 4, pp. 231–237, December 2008.

[62] S. Youmans and J. Stierwalt, "An acoustic profile of normal swallowing," *Dysphagia*, vol. 20, no. 3, pp. 195–209, September 2005.

[63] K. Takahashi, M. Groher, and K. Michi, "Methodology for detecting swallowing sounds," *Dysphagia*, vol. 9, no. 1, pp. 54–62, January 1994, copyright Springer-Verlag New York Inc. 1994.

[64] M. Aboofazeli and Z. Moussavi, "Automated extraction of swallowing sounds using a wavelet-based filter," in *The 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY, August 30 - September 3 2006, pp. 5607–5610.

[65] A. Yadollahi and Z. Moussavi, "Feature selection for swallowing sounds classification," in *The 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, FR, August 22-26 2007, pp. 3172–3175.

[66] S. Hamlet, R. Nelson, and R. Patterson, "Interpreting the sounds of swallowing: Fluid flow through the cricopharyngeus," *Annals of Otology, Rhinology, and Laryngology*, vol. 99, no. 9, pp. 749–752, September 1990.

[67] J. Cichero and B. Murdoch, "Detection of swallowing sounds: Methodology revisited," *Dysphagia*, vol. 17, no. 1, pp. 40–49, January 2002.

[68] E. Reynolds, F. Vice, and I. Gewolb, "Variability of swallow-associated sounds in adults and infants," *Dysphagia*, vol. 24, no. 1, pp. 13–19, March 2009.

[69] W. Selley, F. Flack, R. Ellis, and W. Brooks, "The exeter dysphagia assessment technique," *Dysphagia*, vol. 4, no. 4, pp. 227–235, December 1990.

[70] W. Selley, R. Ellis, F. Flack, C. Bayliss, and V. Pearce, "The synchronization of respiration and swallow sounds with videofluoroscopy during swallowing," *Dysphagia*, vol. 9, no. 3, pp. 162–167, September 1994.

[71] M. Klahn and A. Perlman, "Temporal and durational patterns associating respiration and swallowing," *Dysphagia*, vol. 14, no. 3, pp. 131–138, September 1999.

[72] D. Smith, S. Hamlet, and L. Jones, "Acoustic technique for determining timing of velopharyngeal closure in swallowing," *Dysphagia*, vol. 5, no. 3, pp. 142–146, July 1990.

[73] L. Pinnington, K. Muhiddin, R. Ellis, and E. Playford, "Non-invasive assessment of swallowing and respiration in parkinson's disease," *Journal of Neurology*, vol. 247, no. 10, pp. 773–777, May 2000.

[74] B. Roubeau, S. Moriniere, S. Perie, A. Martineau, J. Falieres, and J. L. S. Guily, "Use of reaction time in the temporal analysis of normal swallowing," *Dysphagia*, vol. 23, no. 2, pp. 102–109, June 2008.

[75] S. Passler and W. Fischer, "Acoustical method for objective food intake monitoring using a wearable sensor system," in *The 5th International Conference on Pervasive Computing Technologies for Healthcare*, vol. 1, Dublin, IE, May 23-26 2011, pp. 266–269.

[76] S. Sarraf-Shirazi, J. Baril, and Z. Moussavi, "Characteristics of the swallowing sounds recorded in the ear, nose and on trachea," *Medical and Biological Engineering and Computing*, vol. 50, no. 8, pp. 885–890, August 2012.

[77] E. Sazonov, S. Schuckers, P. Lopex-Meyer, O. Makeyev, N. Sazonova, E. Melanson, and M. Neuman, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiological Measurement*, vol. 29, no. 5, pp. 525–541, May 2008.

[78] H. Firmin, S. Reilly, and A. Fourcin, "Non-invasive monitoring of reflexive swallowing," *Speech Hearing and Language*, vol. 10, no. 1, pp. 171–184, Janurary 1997.

[79] M. Aboofazeli and Z. Moussavi, "Comparison of recurrence plot features of swallowing and breath sounds," *Chaos, Solitons & Fractals*, vol. 37, no. 2, pp. 454–464, July 2008.

[80] S. Sarraf-Shirazi, C. Buchel, R. Daun, L. Lenton, and Z. Moussavi, "Detection of swallows with silent aspiration using swallowing and breath sound analysis," *Medical and Biological Engineering and Computing*, vol. 50, no. 12, pp. 1261–1268, December 2012, copyright International Federation for Medical and Biological Engineering 2012.

[81] K. Takahashi, M. Groher, and K. Michi, "Symmetry and reproducibility of swallowing sounds," *Dysphagia*, vol. 9, no. 3, pp. 168–173, September 1994.

[82] A. Santamato, F. Panza, V. Solfrizzi, A. Russo, V. Frisardi, M. Megna, M. Ranieri, and P. Fiore, "Acoustic analysis of swallowing sounds: A new technique for assessing dysphagia," *Journal of Rehabilitation Medicine*, vol. 41, no. 8, pp. 639–645, March 2009.

[83] D. Zoratto, T. Chau, and C. M. Steele, "Hyolaryngeal excursion as the physiological source of accelerometry signals during swallowing," *Physiological Measurement*, vol. 31, no. 6, pp. 843–855, May 2010.

[84] S. Morinière, M. Boiron, D. Alison, P. Makris, and P. Beutter, "Origin of the sound components during pharyngeal swallowing in normal subjects," *Dysphagia*, vol. 23, no. 3, pp. 267–273, September 2008, copyright Springer Science+Business Media, LLC 2007.

[85] E. Sejdić, V. Komisar, C. M. Steele, and T. Chau, "Baseline characteristics of dual-axis cervical accelerometry signals," *Annals of Biomedical Engineering*, vol. 38, no. 3, pp. 1048–1059, March 2010.

[86] N. Reddy, R. Thomas, E. Canilang, and J. Casterline, "Toward classification of dysphagic patients using biomechanical measurements," *Journal of Rehabilitation Research and Development*, vol. 31, no. 4, pp. 335–344, November 1994.

[87] N. Reddy, E. Canilang, R. Grotz, M. Rane, J. Casterline, and B. Costarella, "Biomechanical quantification for assessment and diagnosis of dysphagia," *Engineering in Medicine and Biology Magazine*, vol. 7, no. 3, pp. 16–20, September 1988.

[88] M. Pehlivan, N. Yuceyar, C. Ertekin, G. Celebi, M. Ertas, T. Kalayci, and I. Aydogdu, "An electronic device measuring the frequency of spontaneous swallowing: Digital phagometer," *Dysphagia*, vol. 11, no. 4, pp. 259–264, November 1996.

[89] T. Nakamura, Y. Yamamoto, and H. Tsugawa, "Measurement system for swallowing based on impedance pharyngography and swallowing sound," in *The 17th IEEE Instrumentation and Measurement Technology Conference*, vol. 1, Baltimore, MD, May 1-4 2000, pp. 191–194.

[90] K. Kohyama, H. Sawada, M. Nonaka, C. Kobori, F. Hayakawa, and T. Sasaki, "Textural evaluation of rice cake by chewing and swallowing measurements on human subjects," *Biosience, Biotechnology, and Biochemistry*, vol. 71, no. 2, pp. 358–365, February 2007.

[91] S. Eylgor, A. Perlman, and X. He, "Effects of age, gender, bolus volume and viscosity on acoustic signals of normal swallowing," *Journal of Physical and Medical Rehabilitation*, vol. 53, no. 1, pp. 94–99, January 2007.

[92] H. Preiksaitis and C. Mills, "Coordination of breathing and swallowing: Effects of bolus consistency and presentation in normal adults," *Journal of Applied Physiology*, vol. 81, no. 4, pp. 1707–1714, October 1996.

[93] J. Lee, C. Steele, and T. Chau, "Time and time-frequency characterization of dual-axis swallowing accelerometry signals," *Physiological Measurement*, vol. 29, no. 9, pp. 1105–1120, August 2008.

[94] C. Greco, L. Nunes, and P. Melo, "Instrumentation for bedside analysis of swallowing disorders," in *The 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, no. 1, Buenos Aires, AR, August 31 - September 4 2010, pp. 923–926.

[95] J. Lee, C. Steele, and T. Chau, "Swallow segmentation with artificial neural networks and multi-sensor fusion," *Medical Engineering & Physics*, vol. 31, no. 9, pp. 1049–1055, November 2009.

[96] N. Reddy, A. Katakam, V. Gupta, R. Unnikrishnan, J. Narayanan, and E. Canilang, "Measurements of acceleration during videofluorographic evaluation of dysphagic patients," *Medical Engineering & Physics*, vol. 22, no. 6, pp. 405–412, July 2000.

[97] D. Prabhu, N. Reddy, and E. Canilang, "Neural networks for recognition of acceleration patterns during swallowing and coughing," in *The 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, Baltimore, MD, November 3-6 1994, pp. 1105–1106.

[98] V. Gupta, D. Prabhu, N. Reddy, and E. Canilang, "Spectral analysis of acceleration signals during swallowing and coughing," in *The 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, Baltimore, MD, November 3-6 1994, pp. 1292–1293.

[99] A. Joshi and N. Reddy, "Fractal analysis of acceleration signals due to swallowing," in *The 21st Annual Conference of the Biomedical Engineering Society Engineering in Medicine and Biology*, vol. 2, Atlanta, GA, October 13-16 1999, p. 912.

[100] S. Damouras, E. Sejdić, C. M. Steele, and T. Chau, "An online swallow detection algorithm based on the quadratic variation of dual-axis accelerometry," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3352–3359, June 2010.

[101] Y. Guo, D. Cui, Q. Lu, S. Wang, and Q. Jiao, "Chaotic feature of water swallowing sound signals revealed by correlation dimension," in *2nd International Conference on Information Science and Engineering*, vol. 1, Hangzhou, CN, December 4-6 2010, pp. 171–173.

[102] M. Aboofazeli and Z. Moussavi, "Analysis of normal swallowing sounds using nonlinear dynamic metric tools," in *The 26th Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society, vol. 2, San Francisco, CA, September 1-5 2004, pp. 3812–3815.

[103] E. Sejdić, C. M. Steele, and T. Chau, "A method for removal of low frequency components associated with head movements from dual-axis swallowing accelerometry signals," *PLoS ONE*, vol. 7, no. 3, pp. 1–8, March 2012.

[104] ——, "A procedure for denoising of dual-axis swallowing accelerometry signals," *Physiological Measurements*, vol. 31, no. 1, pp. N1–N9, January 2010.

[105] S. Morinière, P. Beutter, and M. Boiron, "Sound component duration of healthy human pharyngoesophageal swallowing: A gender comparison study," *Dysphagia*, vol. 21, no. 3, pp. 175–182, July 2006.

[106] A. Spadotto, J. Papa, A. Gatto, P. Cola, J. Pereira, R. Guido, A. Schelp, C. Maciel, and A. Montagnoli, "Denoising swallowing sound to improve the evaluator's qualitative analysis," *Computers & Electrical Engineering*, vol. 34, no. 34, pp. 148–153, March 2008.

[107] S. Almeida, E. Ferlin, M. Parente, and H. Goldani, "Assessment of swallowing sounds by digital cervial auscultation in children," *Annals of Otology, Rhinology & Laryngology*, vol. 117, no. 4, p. 253, April 2008.

[108] N. Tanaka, K. Nohara, K. Okuno, Y. Kotani, H. Okazaki, M. Matsumura, and T. Sakai, "Development of a swallowing frequency meter using a laryngeal microphone," *Journal of Oral Rehabilitation*, vol. 39, no. 6, pp. 411–420, June 2012.

[109] S. Sarraf-Shirazi and Z. Moussavi, "Investigating the statistical properties of the swallowing sounds," in *The 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, Boston, MA, August 30 - September 3 2011, pp. 6021–6024.

[110] M. Taniwaki and K. Kohyama, "Fast fourier transform analysis of sounds made while swallowing various foods," *Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2478–2482, August 2012.

[111] O. Makeyev, P. Lopez-Meyer, S. Schuckers, W. Besio, and E. Sazonov, "Automatic food intake detection based on swallowing sounds," *Biomedical Signal Processing and Control*, vol. 7, no. 6, pp. 649–656, November 2012.

[112] M. Nagae and K. Suzuki, "Neck mounted interface for sensing the swallowing activity based on swallowing sound," in *The 33rd Annual Internation Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, Boston, MA, August 30 - September 3 2011, pp. 5224–5227.

[113] M. Taniwaki, Z. Gao, K. Nishinari, and K. Kohyama, "Acoustic analysis of the swallowing sounds of food with different physical properties using the cervical auscultation method," *Journal of Texture Studies*, vol. 44, no. 3, pp. 169–175, June 2013.

[114] N. Tanaka, K. Nohara, Y. Kotani, M. Matsumura, and T. Sakai, "Swallowing frequency in elderly people during daily life," *Journal of Oral Rehabilitation*, vol. 40, no. 10, pp. 744–750, October 2013.

[115] S. Moriniere, M. Boiron, L. Brunereau, P. Beutter, and F. Patat, "Pharyngeal swallowing sound profile assessed after partial and total laryngectomy," *Dysphagia*, vol. 26, no. 4, pp. 366–373, December 2011.

[116] A. Spadotto, A. Gatto, P. Cola, R. da Silva, A. Schelp, D. Domenis, and R. Dantas, "Components of the acoustic swallowing signal: Preliminary study," *Jornal da Sociedade Brasileira de Fonoaudiologia*, vol. 24, no. 3, pp. 218–222, January 2012.

[117] O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, and M. Neuman, "Limited receptive area neural classifier for recognition of swallowing sounds using short-time fourier transform," in *The 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, FR, August 22-26 2007, pp. 1417.1–1417.6.

[118] A. Stroud, B. Lawrie, and C. Wiles, "Inter- and intra-rater reliability of cervical auscultation to detect aspiration in patients with dysphagia," *Clinical Rehabilitation*, vol. 16, no. 6, pp. 640–645, 2002.

[119] S. Hamlet, R. Patterson, S. Fleming, and L. Jones, "Sounds of swallowing following total laryngectomy," *Dysphagia*, vol. 7, no. 3, pp. 160–165, September 1992.

[120] E. Reynolds, F. Vice, J. Bosma, and I. Gewolb, "Cervical accelerometry in preterm infants," *Developmental Medicine & Child Neurology*, vol. 44, no. 9, pp. 587–592, September 2002.

[121] E. Reynolds, F. Vice, and I. Gewolb, "Cervical accelerometry in preterm infants with and without bronchopulmonary dysplasia," *Developmental Medicine & Child Neurology*, vol. 45, no. 7, pp. 442–446, July 2003.

[122] M. Boiron, Z. Benchellal, and N. Huten, "Study of swallowing sound at the esophagogastric junction before and after fundoplication," *Journal of Gastrointestinal Surgery*, vol. 13, no. 9, pp. 1570–1576, September 2009.

[123] S. Afkari, "Measuring frequency of spontaneous swallowing," *Australasian College of Physical Scientists & Engineers in Medicine*, vol. 30, no. 4, pp. 313–317, December 2007.

[124] N. Reddy, E. Canilang, J. Casterline, M. Rane, A. Joshi, R. Thomas, and R. Candadai, "Noninvasive acceleration measurements to characterize the pharyngeal phase of swal-

lowing," *Journal of Biomedical Engineering*, vol. 13, no. 5, pp. 379–383, September 1991.

[125] M. Crary, L. Sura, and G. Carnaby, "Validation and demonstration of an isolated acoustic recording technique to estimate spontaneous swallow frequency," *Dysphagia*, vol. 28, no. 1, pp. 86–94, March 2013.

[126] S. Hamlet, D. Smith, and L. Jones, "Acoustical menifestations of velopharyngeal closure during swallowing," *Journal of the Acoustical Society of America*, vol. 85, no. S1, p. S147, January 1989.

[127] D. Donoho and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[128] E. Sejdić, C. Steele, and T. Chau, "Understanding the statistical persistence of dual-axis swallowing accelerometry signals," *Computers in Biology and Medicine*, vol. 40, no. 11, pp. 839–844, November 2010.

[129] D. Talkin, *Speech Coding and Synthesis*. Elsevier Science B.V., 1995, ch. A Robust Algorithm for Pitch Tracking (RAPT), pp. 495–518.

[130] E. Sejdić, T. Falk, C. Steele, and T. Chau, "Vocalization removal for improved automatic segmentation of dual-axis swallowing accelerometry signals," *Medical Engineering & Physics*, vol. 32, no. 6, pp. 668–672, July 2010.

[131] I. Orovic, S. Stankovic, T. Chau, C. Steele, and E. Sejdic, "Time-frequency analysis and hermite projection method applied to swallowing accelerometry signals," *Journal on Advances in Signal Processing*, vol. 2010, no. 1, pp. 1–7, March 2010.

[132] L. Lazareck and S. Ramanna, *Classification of Swallowing Sound Signals: A Rough Set Approach*, ser. Lecture Notes in Computer Science, S. Tsumoto, R. Slowinski, J. Komorowski, and J. Grzymala-Busse, Eds. Springer Berlin Heidelberg, 2004, vol. 3066.

[133] S. Youmans and J. Stierwalt, "Normal swallowing acoustics across age, gender, bolus viscosity, and bolus volume," *Dysphagia*, vol. 26, no. 4, pp. 374–384, December 2011.

[134] S. Suryanarayanan, N. Reddy, and E. Canilang, "A fuzzy logic diagnosis system for classification of pharyngeal dysphagia," *International Journal of Biomedical Computing*, vol. 38, no. 3, pp. 207–215, March 1995.

[135] J. Lee, E. Sejdić, C. M. Steele, and T. Chau, "Effects of stimuli on dual-axis swallowing accelerometry signals in a healthy population," *Biomedical Engineering Online*, vol. 9, no. 7, pp. 1–14, February 2010.

[136] S. Palreddy, N. Reddy, P. Green, and E. Canilang, "Neural networks in computer-aided diagnosis classification of dysphagic patients," in *The 14th Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society, vol. 4, Paris, FR, October 29 - November 1 1992, pp. 1517–1518.

[137] M. Aboofazeli and Z. Moussavi, "Automated classification of swallowing and breadth sounds," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 5, San Francisco, CA, September 1-5 2004, pp. 3816–3819.

[138] T. Chau, M. Casas, G. Berall, and D. Kenny, "Testing the stationarity and normality of paediatric aspiration signals," *Engineering in Medicine and Biology*, vol. 1, no. 1, pp. 186–187, November 2002.

[139] M. Aboofazeli and Z. Moussavi, "Analysis and classification of swallowing sounds using reconstructed phase space features," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, Philadelphia, PA, March 18-23 2005, pp. 421–424.

[140] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić, "A comparative analysis of DBSCAN, k-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals," *Computers in Biology and Medicine*, vol. 59, no. 1, pp. 10–18, April 2015, ©2015 Elsevier. Reprinted with permission.

[141] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, 1967, pp. 281–297.

[142] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, E. Simoudis, J. Han, and U. Fayyad, Eds., vol. 1. AAAI Press, 1996, pp. 226–231.

[143] R. Campello, D. Moulavi, and J. Sander, *Advances in Knowledge Discovery and Data Mining*, ser. Lecture Notes in Computer Science, J. Pei, V. Tseng, L. Cao, H. Motoda, and G. Xu, Eds. Springer-Verlag Berlin Heidelberg, April 2013, vol. 7819, no. 2.

[144] M. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine*, vol. 18, no. 3, pp. 145–156, September 1988.

[145] S. M. Molfenter and C. M. Steele, "Variation in temporal measures of swallowing: Sex and volume effects," *Dysphagia*, 2012, accepted.

[146] J. Cichero and B. Murdoch, "Acoustic signature of the normal swallow: Characterization by age, gender, and bolus volume," *Annals of Otology, Rhinology, and Laryngology*, vol. 11, no. 7, pp. 623–632, July 2002.

[147] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.

[148] J. O. Chapa and R. M. Rao, "Algorithms for designing wavelets to match a specified signal," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3395–3406, December 2000.

[149] J. Rosenbek, J. A. Robbins, E. Roecker, J. Coyle, and J. Wood, "A penetration-aspiration scale," *Dysphagia*, vol. 11, no. 2, pp. 93–98, 1996.

[150] Z. Wang and P. Willett, "Two algorithms to segment white gaussian data with piecewise constant variances," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 373–385, February 2003.

[151] M. Aboy, R. Hornero, D. Abasolo, and D. Alvarez, "Interpretation of the lempel-ziv complexity measure in the context of biomedical signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2282–2288, November 2006.

[152] T. Chau, D. Chau, M. Casas, G. Berall, and D. J. Kenny, "Investigating the stationarity of paediatric aspiration signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 99–105, Mar. 2005.

[153] E. Sejdić, I. Djurović, and J. Jiang, "Time-frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, Jan. 2009.

[154] S. Stanković, I. Orović, and E. Sejdić, *Multimedia Signals and Systems*. New York, NY: Springer US, 2012.

[155] A. Cohen and J. Kovačević, "Wavelets: The mathematical background," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 514–522, Apr. 1996.

[156] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[157] R. Baken and R. Orlikoff, *Clinical Measurement of Speech and Voice*, 2nd ed., ser. Speech Science. Cengage Learning, December 1999.

[158] N. George, F. de Mul, Q. Qiu, G. Rakhorst, and H. Schutte, "Depth-kymography: High-speed calibrated 3D imaging of human vocal fold vibration dynamics," *Physics in Medicine and Biology*, vol. 53, no. 10, pp. 2667–2675, April 2008.

[159] F. A. Scannapieco, "Role of oral bacteria in respiratory infection," *Journal of Periodontology*, vol. 70, no. 7, pp. 793–802, July 1999.

[160] W. Baine, W. Yu, and J. Summe, "Epidemiologic trends in the hospitalization of elderly Medicare patients for pneumonia," *American Journal of Public Health*, vol. 91, no. 7, pp. 1121–1123, July 2001.

[161] J. Robbins, G. Gensler, J. Hind, J. A. Logemann, A. S. Lindblad, D. Brant, H. Baum, D. Lilienfeld, S. Kosek, D. Lundy, K. Dikeman, M. Kazandjian, G. D. Gramigna,

S. McGarvey-Toler, and P. J. M. Gardner, "Comparison of 2 interventions for liquid aspiration on pneumonia incidence: A randomized trial," *Annals of Internal Medicine*, vol. 148, no. 7, pp. 509–518, April 2008.

[162] Y. Meng, M. Rao, and A. Datta, "Computer simulation of the pharyngeal bolus transport of Newtonian and non-Newtonian fluids," *Good and Bioproducts Processing*, vol. 83, no. C4, pp. 297–305, December 2005.

[163] M. O'Leary, B. Hanson, and C. Smith, "Viscosity and non-Newtonian features of thickened fluids used for dysphagia therapy," *Journal of Food Science*, vol. 75, no. 6, pp. 330–338, August 2010.

[164] P. Clave, M. D. Kraa, V. Arreola, M. Gircent, R. Farre, E. Palomera, and M. Serra-Prat, "The effect of bolus viscosity on swallowing function in neurogenic dysphagia," *Alimentary Pharmacology and Therapeutics*, vol. 24, no. 9, pp. 1385–1394, October 2006.

[165] A. Zargaraan, R. Rastmanesh, G. Fadavi, F. Zayeri, and M. A. Mohammadifar, "Rheological aspects of dysphagia-oriented food products: A mini review," *Food Science and Human Wellness*, vol. 2, no. 3, pp. 173–178, September 2013.

[166] J. Lee, S. Blain, M. Casas, D. Kenny, G. Berall, and T. Chau, "A radial basis classifier for the automatic detection of aspiration in children with dysphagia," *Journal of Neuroengineering and Rehabilitation*, vol. 3, no. 14, pp. 1–17, July 2006.

[167] J. M. Dudik, A. Kurosu, J. L. Coyle, and E. Sejdić, "The effects of dysphagia on swallowing sounds and vibrations in adults," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2014, under review.

[168] J. Robbins, J. Coyle, J. Rosenbek, E. Roecker, and J. Wood, "Differentiation of normal and abnormal airway protection during swallowing using the penetration-aspiration scale," *Dysphagia*, vol. 14, no. 4, pp. 228–232, 1999.

[169] D. Kriesel, *A Brief Introduction of Neural Networks*, 2007. [Online]. Available: availableathttp://www.dkriesel.com

[170] M. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.

[171] Y. Bengio, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, January 2009.

[172] R. B. Palm, "Prediction as a candidate for learning deep hierachical models of data," Master's thesis, Technical University of Denmark, 2012.

[173] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," in *The 28th International Conference on Machine Learning*, Bellevue, USA, June 28 - July 2 2011, pp. 689–696.

[174] N. Antonios, G. Carnaby-Mann, M. Crary, L. miller, H. Hubbard, K. Hood, R. Sambandam, A. Xavier, and S. Silliman, "Analysis of a physician tool for evaluating dysphagia on an inpatient stroke unit: The modified mann assessment of swallowing ability," *Journal of Stroke and Cerebrovascular Diseases*, vol. 19, no. 1, pp. 49–57, January 2010.