Neurocomputing 285 (2018) 1–9

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Brief papers

Deep learning for classification of normal swallows in adults

Joshua M. Dudik[a], James L. Coyle[b], Amro El-Jaroudi[a], Zhi-Hong Mao[a], Mingui Sun[c], Ervin Sejdić[a,*]

[a] Department of Electrical and Computer Engineering, Swanson School of Enginering, University of Pittsburgh, Pittsburgh, PA, USA
[b] Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA
[c] Department of Neurological Surgery, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 15 September 2016
Revised 19 September 2017
Accepted 26 December 2017
Available online 31 January 2018

Communicated by Shaoting Zhang.

Keywords:
Dysphagia
Cervical auscultation
Deep learning
Classification

ABSTRACT

Cervical auscultation is a method for assessing swallowing performance. However, its ability to serve as a classification tool for a practical clinical assessment method is not fully understood. In this study, we utilized neural network classification methods in the form of Deep Belief networks in order to classify swallows. We specifically utilized swallows that did not result in clinically significant aspiration and classified them on whether they originated from healthy subjects or unhealthy patients. Dual-axis swallowing vibrations from 1946 discrete swallows were recorded from 55 healthy and 53 unhealthy subjects. The Fourier transforms of both signals were used as inputs to the networks of various sizes. We found that single and multi-layer Deep Belief networks perform nearly identically when analyzing only a single vibration signal. However, multi-layered Deep Belief networks demonstrated approximately a 5–10% greater accuracy and sensitivity when both signals were analyzed concurrently, indicating that higher-order relationships between these vibrations are important for classification and assessment.

1. Introduction

Dysphagia is a term used to describe swallowing impairment [1]. It is seen as a symptom of many conditions, but most commonly occurs as a result of neurological conditions such as physical trauma or stroke [1,2]. Though typically not an immediate threat to a patient's well-being, dysphagia can quickly lead to more serious health complications including pneumonia, malnutrition, dehydration, and even death [2,3]. The first attempt at identifying this condition in the clinic before these serious complications occur is a bedside assessment of the patient's actions and behavior while swallowing. Should this prove inconclusive or is deemed insufficient by the administering clinician, more complex instrumental examinations are utilized. Nasopharyngeal flexible endoscopic evaluations involve visualization of the pharynx and upper airway during oral intake, while videofluoroscopic assessment collects dynamic radiographic images of the oral cavity, pharynx, upper airway and proximal esophagus, throughout the entire swallow event [1,4]. The goal of these assessments is to determine the nature of swallowing pathophysiology, and determine appropriate methods of treatment more accurately than the current bedside assessments allow. However, both of these instrumental examinations require skilled expertise, specialized equipment, and a patient that is able to travel to the site of testing. Previous studies agree that an accurate, simple, non-invasive method of evaluating swallowing function would be a desirable addition to the available tools for assessment.

Multiple different swallowing screening tests have been investigated and implemented in the past. Non-instrumental methods, such as the 3 ounce water challenge [5], the Toronto bedside test [6], or the modified MASA [7] among others, have been widely implemented in the clinical setting. Though they generally have a high sensitivity for detecting aspiration, they have poor specificity and can lead to unnecessary interventions [5–8]. Instrumental-based screening methods have also produced mixed results, but efforts have been made to improve these methods and allow for their use alongside existing screening techniques. Cervical auscultation, in particular, has been studied in significant detail in recent years [9]. Traditionally, this technique has utilized stethoscopes at the bedside to allow a clinician to listen to a patient swallow a bolus of liquid or food in real time. This non-instrumental screening method has not demonstrated adequate predictive value for swallowing disorders [3] but has given rise to a similar instrumental method in the form of digital microphones and accelerometers. In this digitized form any number of signal processing algorithms, such as those meant to filter noise or quantify statistical features, can be used to process the data [9]. The result is a signal that

* Corresponding author.
E-mail addresses: jmd151@pitt.edu (J.M. Dudik), jcoyle@pitt.edu (J.L. Coyle), amro@pitt.edu (A. El-Jaroudi), maozh@engr.pitt.edu (Z.-H. Mao), drsun@pitt.edu (M. Sun), esejdic@ieee.org (E. Sejdić).

https://doi.org/10.1016/j.neucom.2017.12.059
0925-2312/

is much cleaner and easier to analyze accurately and consistently than the human-interpreted signals obtained through non-digital techniques [9].

While methods to classify the results of existing, qualitative swallowing tests are well established, the classification of digital swallowing signals has not been studied in as much detail. This has resulted in multiple studies that demonstrate promising preliminary results, but which still have key experimental deficiencies that call the generalizability of the methods into question. As one example, a recent study by Sarraf-Shirazi and Moussavi [10] sought to differentiate swallowing vibrations that originated from a swallow with no aspiration from swallowing vibrations that originated from swallows that did not result in aspiration. They gathered data from 10 individuals with dysphagia, identified the average spectral power of each swallowing signal over 3 key frequency bands, and used a fuzzy k-means classifier to classify each swallow [10]. Their results demonstrated slightly greater than 80% classification accuracy [10]. A study with similar goals by Nikjoo et al. [11] was published about the same time. This study also sought to differentiate between vibrations from swallows that did or did not result in aspiration, but they instead utilized a support-vector machine classifier with a selection of 8 statistical features as inputs, gathered from 30 participants with dysphagia [11]. They, too, achieved an overall classification accuracy slightly greater than 80% [11]. However, past studies were not limited to these specific aspirating/non-aspirating classes. Das et al. [12] sought to differentiate swallowing vibrations produced by healthy subjects from various artifact signals as well as differentiate dysphagic swallowing vibrations from similar artifact signals. They were able to achieve an overall accuracy of 97% for this task by using hybrid fuzzy logic committee neural networks with a limited selection of statistical features as inputs [12]. Conversely, Suryanarayanan et al. [13] attempted to use swallowing vibrations and pressure measurements in order to classify the severity of aspirated swallow on a 4-point scale. By using simple fuzzy logic, they were able to achieve an overall accuracy of slightly more than 80% on their 22 person data set. These four studies [10–13] are not the sum of all research into classifying swallowing vibrations, but they are representative of the larger body of work and, by extension, demonstrate certain key flaws of past swallowing research. The first point, which some of these studies have addressed themselves, is that the number of subjects used to collect data was limited. In addition to being more susceptible to biases from individual subjects, this has also led to researchers not appropriately differentiating training and testing data sets. The similarly limited choice of input features is another significant drawback of previous research. By manually preselecting mathematical or physiological features to use as inputs to their classifier, researchers may have unintentionally biased their results or reduced the maximum potential accuracy of their classification method. Both the small sample sizes and manual feature selection limit the generalizability of the researchers' findings and accentuate the need for greater refinement of swallowing classification methods.

Further investigating the literature related to swallowing classification reveals that studies that utilize neural network-based classifiers [12,14–17] tend to report higher overall classification rates. While the details of their methodologies vary, one trait these studies have in common is that nearly all of them apply user-selected input features of a mathematically complex nature. Lee et al. [15] explored this topic and finds that high-order features such as normality and dispersion ratio are only quadratically separable rather than linearly. Aboofazeli and Moussavi [17] further support the necessity of such high-level investigation of swallowing vibrations and demonstrate the benefits of both nonlinear analysis techniques and neural networks with multiple hidden layers. While the higher-order analysis of swallowing signals demonstrates clear benefits, these studies acknowledge that they are investigating a limited selection of mathematical signal descriptions and that alternate choices may offer benefits to classification. Such trends were also acknowledged by Makeyev et al. [16], who advocates the use of unsupervised learning combined with a highly redundant signal representation in order to avoid the biases of preselecting mathematical features. From these previous attempts at classifying swallowing vibrations, it is clear that the field would benefit from a technique that was both able to analyze higher-order signal features and could self-select features to analyze through use of unsupervised learning methods. One relatively new classification technique, deep learning, has not yet been used in swallowing research. However it does possess these desirable traits and could easily be implemented as a method of classifying swallows, thereby combining much of the past research on the topic into a single method.

In this study, we propose a method that allows for the differentiation of swallows made by a healthy subject and swallows that did not result in a significant amount of laryngeal penetration that were made by a dysphagic subject. This will be performed using only cervical auscultation signals that were recorded in a clinical environment during typical swallowing examination procedures. A previous study supports this possibility, as it asserts that these two events do produce significantly variable cervical auscultation signals [18]. We also propose that our chosen classification technique, a Deep Belief network, will provide more reliable classification than previously implemented techniques. Its ability to classify data in a non-linear manner based on higher order relationships than a simple, feed-forward neural network should allow for the best possible swallowing classification.

2. Methods

2.1. Participants

The protocol for the study was approved by the Institutional Review Board at the University of Pittsburgh. The data collected in this study has been utilized by other published studies [18–20] and the methods used to collect it have been published previously, but will be summarized here for convenience.

A total of 55 healthy participants (28 men, 27 women, mean age 39) were recruited from the neighborhoods 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 30 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 eligible healthy 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 fluoroscopic analysis [21]. We note that a portion of this data was not used in our experiment in order to balance the number of data points with the amount of data available from our second class, described in the following paragraph.

The non-healthy (dysphagic) 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 (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