A Comparative Analysis of DBSCAN, K-Means, and Quadratic Variation Algorithms for Automatic Identification of Swallows from Swallowing Accelerometry Signals

Joshua M. Dudik[1], Atsuko Kurosu[2], James L Coyle[2], Ervin Sejdić [1,*]

Abstract

Background

Cervical auscultation with high resolution sensors is currently under consideration as a method of automatically screening for specific swallowing abnormalities. To be clinically useful without human involvement, any devices based on cervical auscultation should be able to detect specified swallowing events in an automatic manner.

Methods

In this paper, we comparatively analyze the density-based spatial clustering of applications with noise algorithm (DBSCAN), a k-means based algorithm, and an algorithm based on quadratic variation as methods of differentiating periods of swallowing activity from periods of time without

---

[*]Corresponding author

[1]Department of Electrical and Computer Engineering, Swanson School of Enginering, University of Pittsburgh, 3700 O'Hara Street, Pittsburgh, PA, USA, 15261. Phone: 412-624-0508. E-mails: jmd151@pitt.edu, esejdic@ieee.org

[2]Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, 4028 Forbes Tower, Pittsburgh, PA, USA, 15260. E-mail: jcoyle@pitt.edu

swallows. These algorithms utilized swallowing vibration data exclusively and compared the results to a gold standard measure of swallowing duration. Data was collected from 23 subjects that were actively suffering from swallowing difficulties.

Results

Comparing the performance of the DBSCAN algorithm with a proven segmentation algorithm that utilizes k-means clustering demonstrated that the DBSCAN algorithm had a higher sensitivity and correctly segmented more swallows. Comparing its performance with a threshold-based algorithm that utilized the quadratic variation of the signal showed that the DBSCAN algorithm offered no direct increase in performance. However, it offered several other benefits including a faster run time and more consistent performance between patients. All algorithms showed noticeable differentiation from the endpoints provided by a videofluoroscopy examination as well as reduced sensitivity.

Conclusions

In summary, we showed that the DBSCAN algorithm is a viable method for detecting the occurrence of a swallowing event using cervical auscultation signals, but significant work must be done to improve its performance before it can be implemented in an unsupervised manner.

1. Introduction

1.1. Dysphagia Background

Dysphagia is a general term that is used to refer to a number of swallowing disorders and impairments [1]. Often a co-morbid disorder alongside neurological impairments, stroke in particular, it is estimated that well over 500,000 Americans are affected by dysphagia every year [1, 2, 3]. There are several different methods used to identify the physiologic characteristics of swallowing disorders, but the most widely utilized gold standard diagnostic imaging examination is the videofluoroscopic swallow study (VFSS) [1, 4]. For this test, a patient is asked to swallow various foods and liquids that contain a radiopaque contrast agent while observed by a trained examiner [5, 6]. This examiner analyzes the kinematic x-ray data for biomechanical errors and subsequent misdirection of swallowed material [5, 6]. The examiner, typically a specially trained Speech Language Pathologist, can determine whether or not the patient exhibits pathologically abnormal swallowing, can identify the likely cause and nature of those problems, and can assess the efficacy of interventions designed to mitigate the physiologic abnormalities observed [6]. Their judgment is based on the pattern of bolus propulsion, timing of the opening and closure of various valves in the oropharyngeal mechanism, and the range and speed of motion of the anatomical structures of the throat [6, 7, 8]. These dynamics are all generators of the flow of the bolus through the oropharyngeal structures and into the digestive system [6, 7, 8]. While this test is widely used and is considered the gold standard for assessment of swallowing function, it is reliant on using a small amount of potentially harmful ionizing radiation, requires that the patient be able to actively participate during the exam, and assumes that

the patient can travel to the location of the x-ray instrumentation. These requirements all serve as potential obstacles to implementing proper healthcare in a timely manner, especially in settings in which videofluoroscopy is unavailable [6, 9, 10, 11, 12, 13, 14, 15]. However, screening to accurately predict which patients need a full diagnostic imaging evaluation remains a relatively crude and somewhat subjective process [16]. As a result, improved and more accurate methods of screening for dysphagia are widely sought-after [17].

While several non-instrumented screening procedures have been adopted in medical centers worldwide, efforts to develop improved dysphagia screening methods with both high sensitivity and specificity are currently in development. One such instrumented method, cervical auscultation augmented with accelerometry, aims to provide vibratory data reflecting specific kinematic events within the oropharyngeal mechanism [18, 19]. For traditional cervical auscultation for dysphagia screening, a device that can record sounds is placed on the patient's throat while they swallow [20, 21]. An examiner then listens to these sounds and makes subjective inferences about the patient's swallowing function [20, 21]. In the past this has typically involved the use of a stethoscope, with the examiner subjectively attempting to judge the flow of swallowed material and biomechanical performance of the throat by listening to the generated sounds. Unfortunately, the ability of a clinician to identify discrete impairments and make an accurate assessment from this raw information has been shown to have very poor inter-judge agreement and validity [20, 21, 22]. An alternative method that has been investigated involves using an accelerometer to record the data and then analyzing the data digitally. The ability to precisely filter the data and calculate specific signal features, including those that the human auditory system cannot de-

tect or analyze from the perceived acoustic signal alone, could potentially provide a more accurate and reliable assessment when compared to the bedside stethoscope method [23]. By removing this subjective component from cervical auscultation based screening it could be possible to produce a viable automated screening tool for dysphagia that will more accurately identify patients who are at elevated risk for adverse events and should undergo testing such as VFSS.

1.2. Cervical Auscultation Segmentation Methods

If an automatic swallowing screening tool is to be developed, one important issue that must be addressed is the ability to automatically identify the beginning and end of a swallow in the acoustic/vibratory data stream using only the cervical auscultation data itself. Swallowing vibrations are highly variable since they are produced by a physiological process that is subject to many different, uncontrollable variables. Something relatively minor such as the bolus consistency can change the cervical auscultation signal while a less than perfect connection between the device and the patient's skin can introduce fairly high amplitude artifacts [19]. As such, the resulting data set is not always "well behaved" and can be skewed in the feature space or contain a large number of outliers. This is especially true when looking at the data recorded from a single patient, as such signals will contain a relatively small number of data points from a statistical perspective. This can result in a less than ideal grouping of the data and a loss of both accuracy and precision when segmenting.

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 method that has received significant

attention is the neural network technique. The signal is windowed and then multiple time-varying features are calculated before being fed into the neural network [24, 25, 26, 27]. After sufficient training this network should be able to differentiate between periods of time where swallowing activity is present or absent based on the values of the inputs [24, 25, 26, 27]. However, it is very computationally intensive and researchers are still debating what features are valid inputs to use for the purposes of segmentation [24, 25, 26, 27]. More computationally simple techniques have also been utilized, such as thresholding the time domain signal, but their accuracy has been questionable [28, 29, 30, 24, 31]. With this technique, an amplitude threshold is declared and any part of the signal that lies above that value is considered to be a part of a swallow [28, 29, 30, 24]. This technique has also been modified slightly to threshold, not the signal itself, but various time-varying features with variable levels of accuracy [28, 29, 30, 24, 31, 32, 33, 34]. The quadratic variation algorithm is one notably successful example [35]. Based on the magnitude of the amplitude changes between successive points, the algorithm calculates the volatility and curvature of the time domain signal [35]. 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 [35]. While this method has high sensitivity, it over-identifies the presences of swallowing events because it is unable to differentiate periods of activity corresponding to swallows and periods of high activity corresponding to coughs or other signal artifacts [35].

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 [36, 37, 18, 19].

Like with other techniques, the signal is windowed and several time-varying features are calculated [38]. Unlike other techniques, however, these points are then mapped onto a feature space and grouped into two or more randomly chosen clusters [38]. Through iterative methods the k-means algorithm attempts to group together points with similar feature values by minimizing the distance between the location of a cluster's centroid and all of its members [38]. 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 [38, 37, 36]. However, there are some issues with this clustering technique that can cause problems when segmenting swallowing signals. Its sensitivity to outliers and non-spherically shaped clusters can be problematic when analyzing physiological data, as such a data set is unlikely to be well behaved. This could cause the swallowing signals to be segmented incorrectly or with reduced precision. Furthermore, the non-deterministic nature of the algorithm offers unique challenges if the algorithm is to be implemented without active human oversight.

For this project, our goal was to comparatively analyze existing k-means and quadratic variation algorithms along with a new DBSCAN-based algorithm in the context of swallowing vibration signals. Specifically, we sought to investigate the accuracy of each algorithm by evaluating their abilities to correctly detect swallowing segments that were obtained by human analysis of videofluoroscopic imaging data. By using a data set that consists of patients that have been diagnosed with swallowing disorders we hope to better understand how these methods perform in their intended role. We also hope to be able to characterize the benefits and drawbacks inherent to each method. We hypothesized that the DBSCAN algorithm would segment

7

our data more accurately and reliably than other segmentation methods. Its density-based sorting technique is better suited to physiological data than the k-means technique while being more applicable to an unknown data set than the quadratic variation thresholding methodology.

2. Materials

2.1. The DBSCAN Algorithm

The density-based spatial clustering of applications with noise algorithm, usually abbreviated as DBSCAN, is a recently developed alternative method for clustering data sets [39]. Unlike other clustering algorithms that require many parameters, 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 [39]. The algorithm operates by cycling through all points in the data set and calculating the number of neighbors each point has, which is defined as the number of other points that are within the minimum distance of the original point [39]. Any data point that has fewer neighbors than the minimum cluster size parameter is declared to be a noise point that is not associated with any cluster [39]. However, a point that has at least as many neighbors as the minimum cluster size is declared to be the start of a new cluster [39]. The neighbors of the starting point are added to this cluster as are the neighbors of those points provided that they meet the minimum cluster size requirements [39]. 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 [39].

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 [39, 40]. 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 is able to sort data points into a separate 'noise' cluster if a given point is too dissimilar to the rest of the data set [39, 40]. Rather than forcing every point to belong to a cluster to some degree like other clustering algorithms the DBSCAN algorithm can exclude points from being part of any cluster, which reduces the effects of outliers on its classification performance. Lastly, this algorithm is deterministic [39, 40]. 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.

2.2. Swallowing Vibration Recording

Data was collected by using a tri-axial accelerometer (ADXL 327, Analog Devices, Norwood, Massachusetts). The two main output signals from the accelerometer were bandpass filtered from 0.1 to 3000 Hz with ten times amplification by an analog filter (model P55, Grass Technologies, Warwick, Rhode Island) before being sent into a National Instruments 6210 DAQ and

9

recorded in a custom Labview application (National Instruments, Austin, Texas). A sampling rate of 20 kHz was used to ensure there would be no aliasing of the signal. Previous studies have shown that this setup is adequate to detect swallowing activity [18, 30, 19, 41]. This data set was also digitally processed by the same custom program in Matlab (Mathworks, Natick, MA) that was used in [18]. The only modification was the addition of the Robust Algorithm for Pitch Tracking (RAPT, [42]), which was used to identify and remove some patient vocalizations from the signal and reduce the number of false positives generated by the sorting algorithms [43]. Images output by the fluoroscopy instrumentation (Ultimax system, Toshiba, Tustin, CA) which was operated concurrently with the accelerometer were input to a video capture card (AccuStream Express HD, Foresight Imaging, Chelmsford, MA) and recorded with a custom LabView program (National Instruments, Austin, Texas).

3. Methods

3.1. 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 the gross characteristics of 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 of no more than 5 seconds, and frequencies up to 5 kHz added to a stream of Gaussian white noise. Our limits on duration and frequency are based on past swallowing studies and are intended to represent signal components of a duration and frequency that may be en-

countered during a true swallowing examination [18, 41, 44, 30]. 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 (Pittsburgh, Pennsylvania). Adult patients referred for diagnostic evaluation of swallowing function as part of their inpatient medical care underwent simultaneous videofluoroscopic examination and recording of vibratory data while swallowing radiopaque contrast in a neutral head-neck position. The accelerometer was placed over the cricoid cartilage, located through visual and tactile methods, to concurrently record vibratory data the occurred during the swallows produced by the patients. The main axes of the accelerometer were aligned approximately parallel to the cervical spine and perpendicular to the coronal plane, respectively. The third axis, which was aligned perpendicular to the sagittal plane, was not used for this study since the techniques being studied do not utilize the corresponding data. The cricoid cartilage was chosen as the mounting location as it was previously demonstrated to provide a high quality vibratory signal [45]. Patients with a history of head or neck cancer or major head or neck 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. Patients produced 191 discrete, single swallows with their heads in a neutral position and another 40 single, discrete swallows with their heads in a flexed posture, referred to as a chin-tuck position [7, 46, 47]. Bolus consistency and volume were not controlled for in order to assess the general viability of the segmentation technique, while postural techniques other than the chin tuck were excluded from our anal-

ysis due to their lower rate of use in the clinic. 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, was recruited for this study. They observed the frame-by-frame video recording only and determined the start and end points of each swallow 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 swallowed bolus intersected with the shadow cast on the x-ray image by the posterior border of the ramus of the mandible. The ending (offset) was the time that the hyoid bone completed its motion associated with swallowing-related pharyngeal activity and returned to its resting or pre-swallow position. The protocol for the study was approved by the Institutional Review Board at the University of Pittsburgh.

3.2. Data Analysis

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 rectangular 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 of the segmentation algorithm while still providing enough data points in each segment for properly representative feature calculations. Increasing or decreasing the size of the window by more than 50 ms was found to significantly decrease the accuracy and performance of the algorithm. We then calculated two features with respect to the data points within each window to serve as the basis of the DBSCAN's sorting. The first, standard deviation, is easily calculated

12

through the common formula

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

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

$$WD = \frac{\log L}{\log d} \tag{2}$$

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 [48]. Both of these features have been used in past research on swallowing segmentation [49, 32, 30]. 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 [49, 32]. By making use of both features in our analysis we can differentiate small noise perturbations that only affect one

13

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.

In our efforts, we generally chose time domain features to segment swallowing vibration signals. Time domain features, particularly those that we have chosen, are also relatively simple traits that are common among swallowing signals. Swallowing vibrations have not been thoroughly studied and the exact characteristics that form a swallow are not yet known. Rather than attempt to locate complex waveform shapes or attempt to filter our certain frequency bands that may not be present during all swallows, our chosen features allow us to simply divide a signal into active (swallowing) and non-active (resting) segments. This is not to say that frequency and time-frequency based analyses are not useful in this context. They are likely a closer analog to how cervical auscultation is implemented in the clinical setting, are more receptive to various filtering and noise-cancelling methods, and offer additional signal features that could be beneficial for a segmentation task after further investigation. However, these benefits do not outweigh the importance of time resolution when attempting to locate the start and ending times of an event, and so we have limited our analysis methods to time domain traits of our signal.

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 parallel to the cervical spine. Examining data from both axes concurrently ensures that information is not lost or attenuated when the signal is analyzed. 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. Prior research has shown that the duration of swallowing vibrations is one second or longer, so any event that lasts for less than half of that duration at most clearly cannot be a swallow event [18, 36, 50, 51]. Therefore, any segments in the second category that were less than 400 ms in duration (2 window lengths) were considered to be false positives, as no swallow could be completed in such a short time, 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.

To provide proper context for our results with the DBSCAN algorithm, we also analyzed the data with two alternative algorithms. The k-means algorithm detailed in [36] was used in order to compare the results of our new algorithm with an existing clustering-based technique. All parameters and procedures set forth by the authors remained unchanged for our experiment [36]. The quadratic variation algorithm detailed in [35] was used to provide an example of a non-clustering segmentation method. The process remained unchanged, but the parameter values were adjusted to better suit our data and filtering techniques. Using the process described in the previous work we chose to use a k value of 200 points, corresponding to a sub-sampling rate of 10 ms with our higher sampling rate [35]. We also decreased the volatility threshold to a value of 0.004 and the curvature threshold to a value of -0.00002. The attenuation from our filtering techniques resulted in these features having lower values than in the original work, and so we adjusted the threshold values to match the segments given by the original threshold values on an unfiltered signal [35]. 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 analyzing the imaging data. A 'missed' swallow was defined as a swallow that was not segmented as such by the algorithm, otherwise referred to as a false negative. 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 videofluoroscopic exposures containing

16

the imaging data were collected in short-duration bursts as each 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 video-recorded and therefore could not be compared to the signal data.

We utilized several measures to assess the performance of our segmentation algorithms. The sensitivity of the algorithm was calculated as the ratio of correctly identified swallow segments to the total number of swallow segments identified by the speech-language pathologist. The precision was calculated as the ratio of true positives to the total number of segments produced by the algorithm. The harmonic average (F_1 score) was calculated as twice the product of the sensitivity and precision divided by their sum. Specificity and negative predictive value were not considered for this experiment. While a swallow event can be identified objectively by a specialist, there is no clear definition of what constitutes a single non-swallowing event and so we cannot define the true negative rate of the algorithm. Furthermore, the algorithms do not sort data into multiple classes, but instead sort all data as non-swallowing events unless certain criteria are met. Attempting to characterize the specificity at this stage of the development of this screening method, then, would not provide any useful information that is not already detailed by the algorithm's precision and false positive rate. Two-proportion z-tests were used to search for any statistical differences in the sensitivities of our algorithms while pairwise Wilxocon rank sum tests compared the degree to which the algorithms overestimated the duration of swallows. Lastly, we tested the run time of each algorithm by averaging the time required to segment a random selection of 20 of our real swallowing

17

data signals. Each signal used for this purpose was trimmed to a length of 60 seconds to provide a consistent number of data points across trials.

4. 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 1 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 on average. The quadratic variation algorithm performed the best on this artificial data set with endpoints that were only 60 ms greater than the true endpoints on average.

We then compared the performance of the three algorithms when segmenting swallowing vibrations from patients with swallowing difficulties. Figure 2 provides an example of this data and the corresponding output from the DBSCAN algorithm for illustrative purposes. In this figure, the first swallow was segmented correctly whereas the second swallow was located correctly but the duration was over estimated due to a signal artifact.

18

For comparison, figure 3 provides a pair of swallows made by a subject with no swallowing difficulties. 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. Using a two-proportion z-test, we found that the sensitivities of the DBSCAN and quadratic variation algorithms were not significantly different ($p = 0.549$ and $p = 0.303$ for normal and chin tuck data, respectively). The sensitivity of the k-means algorithm was significantly different from both other algorithms for data gathered in a neutral head position ($p << 0.001$ for both), but was only significantly different from the sensitivity of the quadratic variance algorithm for chin tuck data ($p = 0.165$ and $p = 0.017$ for the DBSCAN and quadratic variance algorithms respectively).

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, on average, approximately 0.85 s before and after the true endpoints of the swallow (1.70 s total). This value increased to approximately 1.05 s for chin 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, on average, 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 an average of 0.65 s and 0.80 s for normal and chin-tuck swallows, respectively. All of these differences demonstrate statistical significance at the $p < 0.05$ level using pairwise Wilcoxon rank-sum tests with the exception of the overestimation of chin tuck swallows by the DBSCAN

19

and k-means algorithms ($p = 0.11$). The average runs times to segment a 60 second signal for the three algorithms were equal to 10.2 s, 8.9 s, and 177 s for the DBSCAN, k-means, and quadratic variation algorithms, respectively.

5. Discussion

5.1. Performance Summary

The performance of the quadratic variation algorithm, specifically Table 2, is comparable to the results presented in a previous study with non-healthy swallows [35]. As such, we can safely assume that our data set and testing procedure was similar to that used previously to test the k-means and quadratic variation algorithms' performances. The performance of the two previously tested algorithms, however, was noticeably worse than their performances with healthy data [36, 35]. This could be the result of multiple issues, but we feel that it is chiefly the result of a lower signal to noise ratio in this study. Dysphagic syndromes are notorious for producing a variety of perturbations in typical physiologic and kinematic oropharyngeal events that occur during swallows [52, 4, 53]. Weakened muscles that produce less forceful swallows and generate lower magnitude vibrations is one fairly common trait of dysphagia that could have played a substantial role in this study [54]. During this experiment, it was qualitatively noted that the vibrations corresponding to swallows were lower in magnitude than those signals recorded from healthy patients in a previous study [18]. This is apparent by comparing the amplitudes of the signals in figures 2 and 3. Even after implementing our signal processing strategies the signal to noise ratio of dysphagic signals did not achieve the same level that it did in the healthy patient experiment. All three algorithms operate on the assumption

20

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 with more controlled data sets [18, 36, 35], we can safely conclude that the lower performance observed 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. This resulted in somewhat more artifacts to be present in our data, as the examiner focused on gathering data via videofluoroscopy, than there would have been in an environment where the focus of the test was gathering unadulterated data with the accelerometer. This resulted in all algorithms producing more than the expected number of false positives, as certain motion artifacts can produce 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 to which those outliers belong will change its location and some points will change their membership as a result. In our case, since artifacts 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 by the sorting of outliers. 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 artifacts 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 comparably 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. However, the DBSCAN algorithm has two key benefits to offset this greater over-estimation of the swallow duration. 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 [35]. 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 [35]. 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. Our results demonstrate this disparity for signal lengths of one minute, which are practical to obtain during a clinical swallowing examination. We show that in this situation, the DBSCAN algorithm offers a run time that is less than one tenth of that offered by the quadratic variation algorithm.

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 noticeably 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 by the algorithm, but it would produce multiple false positive segments for every true positive for the second patient's data set and be 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. It uses the normalized standard deviation of each data point 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 those produced by 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 to these feature differences, the clustering technique used by the DBSCAN algorithm has less 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.

24

5.2. DBSCAN Advantages and Disadvantages

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 neighboring points, that must be adjusted in any significant capacity while the others can be simply chosen to suit the task [39]. Even the window size has a minimal affect on the overall performance of the algorithm provided it does not significantly alter the number of data points, and therefore the density of the feature space, obtained from a given signal. A 50 ms increase or decrease in the window size affected the classification of no more than 5 total swallows in this study. 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 for a given signal in order to segment the data set [35]. 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 [35]. 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 to some extent without additional classification methods or reduced performance 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 corresponding to background noise being crowded into the low standard deviation and low waveform fractal dimension quadrant of the feature space. Only a few other points that form the segments that contain swallowing activity are spread around the remainder of the feature space. There are two ways to solve this issue. Our chosen method was to simply turn the multi-cluster sorting into a novelty detection algorithm, where any point that was 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 non-swallow events we do not currently have the knowledge necessary and so little can be done at the moment to correct the issue. Future research, however, could potentially address this shortcoming. 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 would require an independent study dedicated to this method.

6. 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, our initial hypothesis, that the DBSCAN algorithm was a superior method of segmenting swallowing accelerometry data, was proven incorrect. We found that the k-means algorithm was objectively inferior in all respects, but the quadratic variation algorithm had similar results to those produced by the DBSCAN algorithm for our chosen performance metrics. We still feel that, given all considerations, the DBSCAN algorithm is the superior option because it offers several usability and consistency improvements while providing similar performance to the quadratic variation algorithm. In spite of these advan-

tages, there is still room for improvement when it comes to automatically segmenting swallowing data in an unsupervised manner.

7. Conflict of Interest

References

[1] M. Spieker, Evaluating dysphagia, American Family Physician 61 (12) (2000) 3639–3648.

[2] D. Smithard, P. O'Neill, C. Park, J. Morris, R. Wyatt, R. England, D. Martin, Complications and outcome after acute stroke. does dysphagia matter?, Stroke 27 (7) (1996) 1200–1204.

[3] A. Castrogiovanni, Communication facts: Special population: Dysphagia, Tech. rep., American Speech-Language-Hearing Association, 2200 Research Boulevard, Rockville, MD 20850 (January 2008).

[4] J. Coyle, L. Davis, C. Easterling, D. Graner, S. Langmore, S. Leder, M. Lefton-Greif, P. Leslie, J. Logemann, L. Mackay, B. Martin-Harris, J. Murray, B. Sonies, C. M. Steele, Oropharyngeal dysphagia assessment and treatment efficacy: Setting the record straight (response to campbell-taylor), Journal of the American Medical Directors Association 10 (1) (2009) 62–66.

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

[6] M. Rugiu, Role of videofluoroscopy in evaluation of neurologic dysphagia, ACTA Otorhinolaryngologica Italica 27 (6) (2007) 306–316.

[7] J. Logemann, Evaluation and Treatment of Swallowing Disorders, 2nd Edition, Pro ed, 1998.

[8] J. Logemann, Manual for the Videofluorographic Study of Swallowing, 2nd Edition, Pro ed, 1993.

[9] H. Bonilha, K. Humphries, J. Blair, E. Hill, K. McGrattan, B. Carnes, Radiation exposure time during mbss: Influence of swallowing impairment severity, medical diagnosis, clinician experience, and standardized protocol use, Dysphagia 28 (1) (2013) 77–85.

[10] I. Zammit-Maempel, C. Chapple, P. Leslie, Radiation dose in videofluoroscopic swallow studies, Dysphagia 22 (1) (2006) 13–15.

[11] A. da Silva, J. L. Neto, P. Santoro, Comparison between videofluoroscopy and endoscopic evaluation of swallowing for the diagnosis of dysphagia in children, Otolaryngology - Head and Neck Surgery 143 (2) (2010) 204–209.

[12] S. Langmore, K. Schatz, N. Olsen, Fiberoptic endoscopic examination of swallowing safety: A new procedure, Dysphagia 2 (4) (1988) 216–219.

[13] S. Langmore, K. Schatz, N. Olson, Endoscopic and videofluoroscopic evaluations of swallowing and aspiration, The Annals of Otology, Rhinology, and Laryngology 100 (8) (1991) 678–681.

[14] B. Martin-Harris, J. Logemann, S. McMahon, M. Schleicher, J. Sandidge, Clinical utility of the modified barium swallow, Dysphagia 15 (3) (2000) 136–141.

[15] A. Nacci, F. Ursino, R. L. Vela, F. Matteucci, V. Mallardi, B. Fattori, Fiberoptic endoscopic evaluation of swallowing (fees): Proposal for informed consent, Acta Otorhinolaryngologica Italica: Organo Ufficiale Della Societa Italiana Di Otorinolaringologia E Chirurgia Cervico-Facciale 28 (3) (2008) 206–211.

[16] G. McCullough, R. Wertz, J. Rosenbek, Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke, Journal of Communication Disorders 34 (1-2) (2001) 55–72.

[17] D. Doggett, K. Tappe, M. Mitchell, R. Chapell, V. Coates, C. Turkelson, Prevention of pneumonia in elderly stroke patients by systematic diagnosis and treatment of dysphagia: An evidence-based comprehensive analysis of the literature, Dysphagia 16 (4) (2001) 279–295.

[18] J. M. Dudik, I. Jestrović, B. Luan, J. L. Coyle, E. Sejdić, A comparative analysis of swallowing accelerometry and sounds during saliva swallows, Medical and Biological Engineering and ComputingUnder review.

[19] I. Jestrović, J. Dudik, B. Luan, J. Coyle, E. Sejdić, The effects of increased fluid viscosity on swallowing sounds in healthy adults, Biomedical Engineering Online 12 (90) (2013) 1–17.

[20] P. Leslie, M. Drinnan, P. Finn, G. Ford, J. Wilson, Reliability and

validity of cervical auscultation: A controlled comparison using videofluoroscopy, Dysphagia 19 (4) (2004) 231–240.

[21] P. Zenner, D. Losinski, R. Mills, Using cervical auscultation in the clinical dysphagia examination in long-term care, Dysphagia 10 (1) (1995) 27–31.

[22] K. L. Shem, D. Castillo, S. L. Wong, J. Chang, M.-C. Kao, S. Kolakowsky-Hayner, Diagnostic accuracy of bedside swallow evaluation versus videofluoroscopy to assess dysphagia in inidividuals with tetraplegia, PM&R 4 (4) (2012) 283–289.

[23] Full revision of the international standards for equal-loudness level contours (iso 226) (October 2003).

[24] J. Lee, C. Steele, T. Chau, Swallow segmentation with artificial neural networks and multi-sensor fusion, Medical Engineering & Physics 31 (9) (2009) 1049–1055.

[25] A. Das, N. Reddy, J. Narayanan, Hybrid fuzzy logic committee neural networks for recognition of swallow acceleration signals, Computer Methods and Programs in Biomedicine 64 (2) (2001) 87–99.

[26] O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, 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, 2007, pp. 3128–3131.

[27] O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, 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, 2007, pp. 1417.1–1417.6.

[28] O. Makeyev, E. Sazonov, S. Schuckers, P. Lopez-Meyer, T. Baidyk, E. Melanson, M. Neuman, Recognition of Swallowing Sounds Using Time-Frequency Decomposition and Limited Receptive Area Neural Classifier, Vol. 15, Springer London, 2009.

[29] J. Fontana, P. Melo, 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, 2011, pp. 6890–6893.

[30] J. Lee, C. Steele, T. Chau, Time and time-frequency characterization of dual-axis swallowing accelerometry signals, Physiological Measurement 29 (9) (2008) 1105–1120.

[31] I. Orovic, S. Stankovic, T. Chau, C. Steele, E. Sejdic, Time-frequency analysis and hermite projection method applied to swallowing accelerometry signals, Journal on Advances in Signal Processing 2010 (1) (2010) 1–7.

[32] L. Lazareck, S. Ramanna, Classification of Swallowing Sound Signals: A Rough Set Approach, Vol. 3066 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2004.

[33] L. Lazareck, Z. Moussavi, Swallowing sound characteristics in healthy and dysphagic individuals, in: The 26th Annual International Confer-

ence of the IEEE Engineering in Medicine and Biology Society, Vol. 2, San Francisco, CA, 2004, pp. 3820–3823.

[34] L. Lazareck, Z. Moussavi, Classification of normal and dysphagic swallow by acoustical means, IEEE Transactions on Biomedical Engineering 51 (12) (2004) 2103–2112.

[35] S. Damouras, E. Sejdić, C. M. Steele, T. Chau, An online swallow detection algorithm based on the quadratic variation of dual-axis accelerometry, IEEE Transactions on Signal Processing 58 (6) (2010) 3352–3359.

[36] E. Sejdić, C. M. Steele, 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 56 (4) (2009) 1090–1097.

[37] M. Nikjoo, C. Steele, E. Sejdić, T. Chau, Automatic discrimination between safe and unsafe swallowing using a reputation-based classifier, Biomedical Engineering Online 10 (100) (2011) 1–17.

[38] 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.

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

[40] R. Campello, D. Moulavi, J. Sander, Advances in Knowledge Discovery and Data Mining, Vol. 7819 of Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg, 2013.

[41] E. Sejdić, V. Komisar, C. M. Steele, T. Chau, Baseline characteristics of dual-axis cervical accelerometry signals, Annals of Biomedical Engineering 38 (3) (2010) 1048–1059.

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

[43] E. Sejdić, T. Falk, C. Steele, T. Chau, Vocalization removal for improved automatic segmentation of dual-axis swallowing accelerometry signals, Medical Engineering & Physics 32 (6) (2010) 668–672.

[44] E. Sejdić, C. M. Steele, T. Chau, A method for removal of low frequency components associated with head movements from dual-axis swallowing accelerometry signals, PLoS ONE 7 (3) (2012) 1–8.

[45] S. Youmans, J. Stierwalt, An acoustic profile of normal swallowing, Dysphagia 20 (3) (2005) 195–209.

[46] T. Shanahan, J. Logemann, A. Rademaker, B. Pauloski, P. Kahrilas, Chin-down posture effect on aspiration in dysphagic patients, Archives of Physical Medicine and Rehabilitation 74 (7) (1993) 736–739.

[47] M. Welch, J. Logemann, A. Rademaker, P. Kahrlas, Changes in pharyngeal dimensions effected by chin tuck, Archives of Physical Medicine and Rehabilitation 74 (2) (1993) 178–181.

[48] M. Katz, Fractals and the analysis of waveforms, Computers in Biology and Medicine 18 (3) (1988) 145–156.

[49] M. Aboofazeli, Z. Moussavi, Comparison of recurrence plot features of swallowing and breathe sounds, Chaos, Solitons & Fractals 37 (2) (2008) 454–464.

[50] S. M. Molfenter, C. M. Steele, Variation in temporal measures of swallowing: Sex and volume effects, DysphagiaAccepted.

[51] J. Cichero, B. Murdoch, Acoustic signature of the normal swallow: Characterization by age, gender, and bolus volume, Annals of Otology, Rhinology, and Laryngology 11 (7) (2002) 623–632.

[52] J. Logemann, A. Veis, L. Colangelo, A screening procedure for oropharyngeal dysphagia, Dysphagia 14 (1) (1999) 44–51.

[53] G. Bours, R. Speyer, J. Lemmens, M. Limburg, R. de Wit, Bedside screening tests vs. videofluoroscopy or fiberoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: Systematic review, Journal of Advanced Nursing 65 (3) (2009) 477–493.

[54] L. Burkhead, C. Sapienza, J. Rosenbek, Strength-training exercise in dysphagia rehabilitation: Principles, procedures, and directions for future research, Dysphagia 22 (2007) 251–265.

Summary

Cervical auscultation is currently under consideration as a method of automatically screening for specific swallowing abnormalities. To be clinically useful without human involvement, any devices based on cervical auscultation should be able to detect specified swallowing events in an automatic manner. In this paper, we comparatively analyze the density-based spatial clustering of applications with noise algorithm (DBSCAN), a k-means based algorithm, and an algorithm based on quadratic variation in order to differentiate periods of swallowing from periods of time without swallows using only swallowing vibration data. The later two algorithms have been previously designed for use with cervical auscultation, but their performance on data with non-healthy participants has been mixed. A new algorithm based on the DBSCAN method was implemented because it provides several benefits to the existing algorithms. Its clustering-based sorting is not as strict as the thresholding of the quadratic variation algorithm, but also is better as sorting arbitrarily shaped and noisy data when compared to the k-means algorithm. Both attributes make it an attractive choice for segmenting physiological data.

Data was collected from 23 subjects that were currently suffering from swallowing difficulties and were scheduled to undergo a videofluoroscopic evaluation at the University of Pittsburgh Medical Center. Comparing the performance of the DBSCAN algorithm with a proven segmentation algorithm that utilizes k-means clustering showed that the DBSCAN algorithm has a higher sensitivity and correctly segments more swallows. Comparing its performance with a threshold-based algorithm that utilized the quadratic variation of the signal showed that the DBSCAN algorithm offered no di-

36

rect increase in performance, but offered several other benefits including a faster run time and more consistent performance between patients. All algorithms showed noticeable differentiation from the endpoints provided by a videofluoroscopy examination, where a trained speech-language pathologist observed only the videofluoroscopic output, as well as reduced sensitivity. In summary we showed that the DBSCAN algorithm is a viable method for detecting the occurrence of a swallowing event using cervical auscultation signals. However, it provides only small advantages over the best alternative algorithm and significant work must be done to improve its performance before it can be implemented in an unsupervised manner.
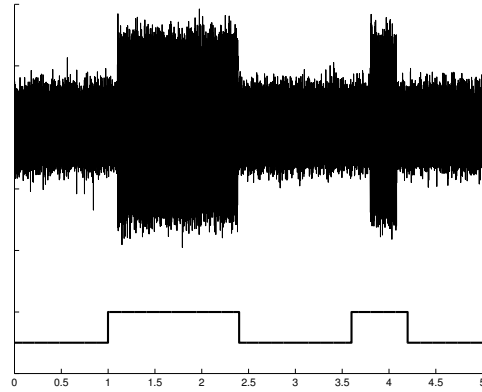
Figures

Figure 1: DBSCAN segmentation of artificial signal. The upper signal is our artificial test signal used to test our 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.
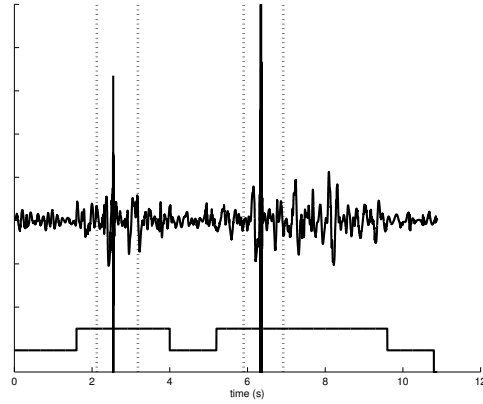
Figure 2: DBSCAN segmentation of real dysphagic signal. The upper signal shows two swallows made by a subject suffering from swallowing difficulties as recorded by the accelerometer channel aligned in the anterior-posterior direction. The lower signal consists of the output of the DBSCAN segmentation 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.
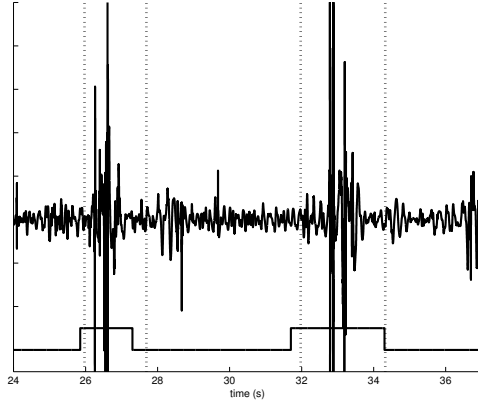
Figure 3: DBSCAN segmentation of real healthy signal. The upper signal shows two swallows made by a patient with no swallowing difficulties as recorded by the accelerometer channel aligned in the anterior-posterior direction. The lower signal consists of the output of the DBSCAN segmentation algorithm, 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 k-means algorithm, which has been proven to correspond to the endpoints provided by a speech-language pathologist for healthy subjects [36]. In comparison to figure 2, this signal is noticeably stronger than that created by the subject with dysphagia.

40

Tables

Table 1: Raw 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

Table 2: Statistical Summary of Algorithm Performance

	DBSCAN		K-Means		Quadratic Variation	
	Neutral	Chin Tuck	Neutral	Chin Tuck	Neutral	Chin Tuck
Sensitivity	0.749	0.700	0.542	0.550	0.775	0.800
Precision	0.611	0.549	0.505	0.449	0.554	0.478
Harmonic Average (F_1 score)	0.673	0.615	0.523	0.494	0.646	0.598