AUTOMATED ALGORITHM FOR SWALLOWING SOUND DETECTION

Lisa J. Lazareck, Zahra K. Moussavi
Department of Electrical Engineering, University of Manitoba, Winnipeg, Canada, R3T 2N2
E-mail: lazareck@ee.umanitoba.ca, mousavi@ee.umanitoba.ca

ABSTRACT

This study examines the automated detection of swallowing sounds for normal subjects. A normal swallowing sound is characterized by three phases including oral, pharyngeal and esophageal where complications lead to swallowing disorder or dysphagia. Current gold standard testing for this abnormality is videofluorography, an x-ray based procedure with detrimental radiation side effects. New non-invasive techniques are necessarily explored to help assess the performance of the swallowing mechanism. Recent developed studies in acoustical airflow estimation indicate the need to detect and extract swallowing segments from sound signals. Extraction is currently a manual process, both subjective and time-consuming. Thus, an automated, objective and quick method is developed in the form of a "smart" algorithm with the ability to make decisions like trained technicians and physicians. Three sound signal features were explored to assist in the classification process (AR-coefficients, RMS values and average power). Utilizing the features, classification sequences were produced for six healthy subjects swallowing sounds. The results were compared with known values (acquired through visual and auditory means). RMS features in combination with the "smart" code yield the lowest error, on average 20.7 ± 4.6%. Future studies include testing variations of smart algorithm code in order to create a robust algorithm. Also, future work includes varying test subject ages, test media (bolus textures) and creating a program-user interface for decision-making assistance.

Keywords: *swallowing, dysphagia, spectrogram, RMS, automated detection.*

INTRODUCTION

The complex act of swallowing involves the coordination of various muscles with the simultaneous closing of the epiglottis and soft palate in order to prepare the body for consumption and to prevent the body from aspiration. [1]. Any lack of coordination or disorder in the swallowing mechanism is known as dysphagia [2]. Dysphagia is most common in patients who suffer from stroke or head injuries, polio, Guillian-Barre syndrome or spread vascular disease [3]. Dysphagic patients are at risk of choking, malnutrition, dehydration, cachexia and death. These health hazards reveal the essential need to understand both normal and abnormal swallowing mechanisms. In fact, many studies assess the maturity and competence of the swallowing mechanism by analyzing breath and swallow coordination [4]. Recent developments, involving acoustical airflow estimation [5, 6] reveal the need to detect swallowing segments and extract them from breath segments automatically. Currently, this is done in a time-consuming, manual and subjective manner. A quick, automatic and objective method is the subject of this paper where an algorithm is developed to perform such a task.

The sound signals used consist of breath and swallowing sections (Figure 1). A breath consists of inspiration and expiration; both respiratory sounds are relatively stationary signals. A swallowing signal consists of three segments, theoretically corresponding to the oral, pharyngeal and esophageal phases [7]. In research communities, there is no agreement on which exact physical mechanisms are involved per phase; however, this issue is beyond the scope of this study. The swallowing process begins with the insertion of the bolus (food, solid or liquid) into the mouth. The oral phase restricts the bolus in the mouth and out of the pharynx using the lips and tongue. The pharyngeal phase involuntary passes the bolus from the mouth into the esophagus once the bolus reaches the epiglottis [8]. This phase contributes to the first two swallowing segments called the "initial click" (for the opening of the cricoipharynx) and "non-click" (for bolus transition into the esophagus) [7]. Lastly, the esophageal phase involuntary pushes the bolus through the esophagus into the stomach via peristaltic waves and contributes to the last swallow segment called the "final click" (for the return of the epiglottis) [8]. Note that it is also possible to differentiate the various signals through auditory analysis.

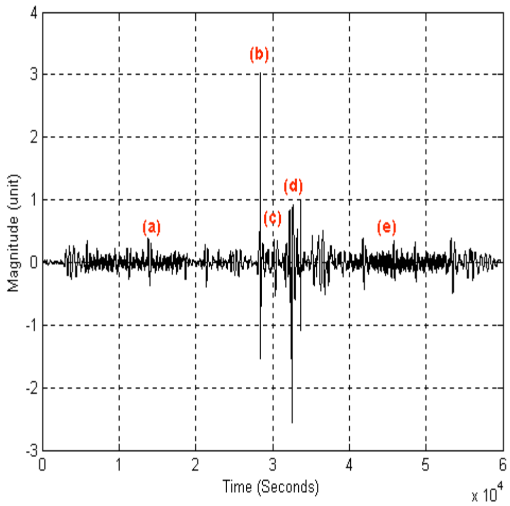

Figure 1. A typical breath and swallowing sound.
(a) inspiration, (b) initial click of swallowing sound,
(b) non-click segment of swallowing sound,
(d) final click of swallowing sound, (e) expiration.

The developed algorithm in this study utilizes feature extraction to best represent the major characteristics of the signals and signal analysis principles to non-invasively analyze the sound signal.

METHOD

Data

For the purpose of developing the automated swallowing detection algorithm, six respiratory and swallowing sound recordings were selected from a previous study data, in which healthy subjects (children of ages 5-15) participated in that study [6]. The breath and swallowing sounds were recorded by Siemens EMT 25 C accelerometers placed over the supersternal notch of the trachea. The breath and swallowing sounds were amplified, bandpass filtered (30-2500 Hz) and digitized at a 10240 Hz sampling rate.

Signal Processing

The three key features extracted, to represent the sound signal characteristics, include autoregressive (AR) coefficients, root-mean-square (RMS) values of the time signal and the average power of the signal (P_{ave}) within a frequency band. It is hypothesized that employing these features would allow the data to clearly separate into two distinct classes (or data clusters). More specifically, two trials were carried out that investigated the separation into breath and swallowing classes and "click" and "non-click" classes (where initial and final click are considered a single entity). The AR, RMS and P_{ave} features were selected because the time domain signal noticeably differs between sections; i.e., swallowing segments are much louder (of greater magnitude) than breath segments due to the "clicks," and also the breath segments are consistent in the frequency domain whereas swallowing segments are not.

The sound signals were sequestered into 100 ms segments with 50% overlap between successive segments. AR-coefficients, RMS and P_{ave} of each segment were calculated for classification. First, the coefficients of each segment were studied as the characteristic features. The sound signal in each segment was modeled by a Burg algorithm of order 7 [9]. Various combinations of the AR-coefficients were compared in a two-dimensional plane. For example, the 1st and 2nd order coefficients were compared to the 1st and 3rd set of coefficients and so on as the characteristic features for classification. Each segment was then classified as breath or swallowing segment using cluster analysis [10].

Second, the average power was examined. Spectrogram of the sound signal was calculated for every 100 ms of data with 50% overlap between adjacent segments applying a Hanning window (Figure 2). The three frequency ranges 100-450 Hz, 100-700 Hz, and 450-2500 Hz were studied for average power calculation. The first two ranges were selected to focus on the majority of the signal. The lower 100 Hz was selected to avoid background noise. The third range was selected to include a good majority of the signal and had an upper bound dictated by hardware limitations. Average power of the swallowing and breath segments were studied to see if they form two separate clusters. The classification error was calculated as the frequency in which the labeled sequences differed, per segment, from the actual known values.

Third, the RMS values were examined as characteristic features. First, the mean and standard deviation were calculated for a known breath section, creating a threshold of 95% interval of confidence ($\mu+2\sigma$). Each RMS segment was then compared to this threshold where values above yielded a swallowing label, below a breath label. The resultant labels were compared to known values in order to calculate classification error.

Fourth, the P_{ave} and RMS features were combined using an exclusive-or (giving priority to segments labeled as swallows). The frequency range utilized was the optimal range obtained from the previous spectrogram trial.

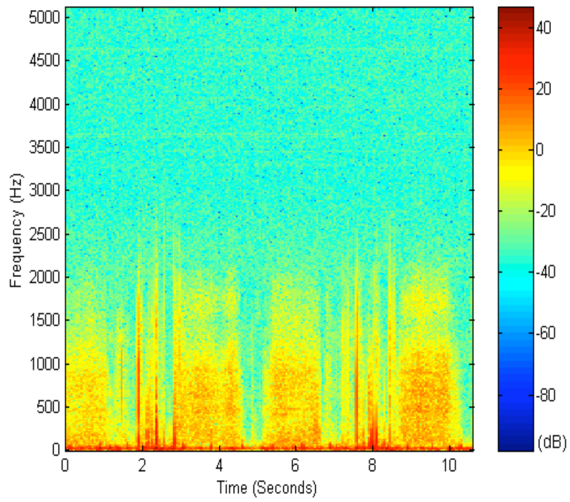

Figure 2. A typical signal spectrogram.

Finally, a "smart" algorithm was developed. The program had the ability to check previous segments and determine whether the next segment should be labeled as a breath or swallow, ultimately acting like a trained technician or physician tracking patterns within the sound signal. Three major constraints were used that stipulated what the "current" segment should be labeled as. First, a definite breath label was defined for a segment that preceded five (tentatively labeled) breath segments. (Five segments correspond to 250 ms). Second and third, a swallowing label was defined for a segment that preceded four swallowing segments or some combination of segments where the majority were (tentatively labeled) swallowing segments. Along with these constraints, the program used another smart code that replaced a misclassified swallowing segment should the majority of surrounding segments be labeled as swallowing segments. The smart code was applied to the RMS sequences and the combined (spectrogram and RMS) sequences. The fully labeled segments for both cases were compared to known values for error evaluation.

Following the smart code results, additional testing was performed that incorporated "quiet" segments into both the RMS and smart RMS code. "Quiet" segments referred to segments small in RMS magnitude (at the background noise level) that typically preceded swallowing segments, "initial click." Additionally, "quiet" segments were found within the "non-click" segments and between inspiration and expiration. Thus, the implementation was performed with results compared for error evaluation.

RESULTS & DISCUSSION

The AR-coefficients were the first features used to classify sound signal data distinctly. For both the click/non-click and breath/swallow classification, every possible combination of coefficients performed inadequately. This was proven by the obvious visual overlap between classes and misclassification error (Figure 3). This can be explained by the fact that AR-coefficients represent the entire signal in a compressed form. While swallowing and respiratory sections are different in their power over a certain frequency band, this difference would be blunted by AR-coefficients as they represent a smoother version of the original data for the entire frequency band.

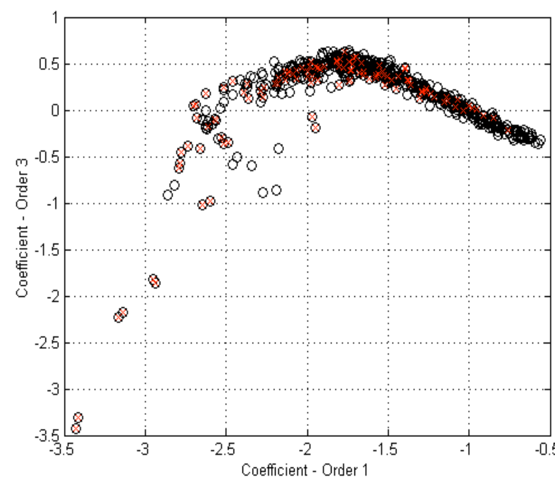

Figure 3. Typical AR-coefficient values for signal. ("x" = swallow, "o" = breath)

Next, the following illustrates the results obtained from utilizing the P_{ave} feature, P_{ave} and RMS features combined, and the smart algorithm (applied to the combined sequence). Each error was averaged for the six subject signals used. The optimal result was from the combined error, in the 100-450 Hz range.

Range (Hz)	*Error % (power)*	*Error % (combined)*	*Error % (smart & combined)*
100-450	24.9 ± 9.9	20.7 ± 4.6	23.7 ± 3.7
100-700	26.0 ± 9.7	22.0 ± 4.6	24.5 ± 3.5
450-2500	25.8± 10.8	21.3 ± 5.3	24.6 ± 3.2

Table 1. Spectrogram-based results (mean ± SE).

First, it was concluded that the combination of P_{ave} and RMS values consistently yielded smaller error than that using P_{ave} alone. Thus, two features were better than one. Second, the smart algorithm was applied to the combined features, as their error percentages were lowest. Opposite to initial hypotheses, the smart algorithm, in this case, actually increased the overall average error slightly. Referring back to the separate subject results, it was observed that if the combined error was quite low (~10%), the smart code failed. However, if the combined error was quite high (~30%), the smart code reduced the error considerably. Hence, on average (in this case) the smart code did not significantly perform well. Trial and error investigations indicated that an optimal frequency range existed per subject, where the smart algorithm reduced the error sufficiently. It is hypothesized that a relationship exists though not investigated further in this study (between optimal frequency range, combined error and smart (combined) error).

The final results from the study are depicted below, describing the averaged classification errors obtained using the RMS values, smart algorithm (applied to the RMS values) and the additional errors from incorporating the aforementioned "quiet" segments.

Error % (RMS)	*Error % (smart & RMS)*	*Error % (RMS & quiet)*
23.1 ± 4.0	21.6 ± 3.9	22.8 ± 3.8

Table 2. RMS-based results (mean ± SE).

RMS error was quite comparable to the combined error for two features (listed above as 20.7 ± 4.6%). However, the combined feature requires an appropriate frequency range, which differed significantly between the subjects. Thus, the smart RMS was chosen as the superior study feature. It was hypothesized that the reason for such a large (although optimal) RMS error was due to the loud expiratory sounds. Lastly, it was surprising that the "quiet" segments slightly enhanced the classification error (a result worthwhile of further investigation). Nevertheless, the errors were per segment, not per section (breath and swallowing sections). In another word, all the swallowing "click" sounds were classified correctly but some segments within the swallowing sections as well as some segments in forceful expiration sections were misclassified (especially for segments that were neither breath nor swallowing but included a noise due to tongue movement). Overall, the results of this study are encouraging for further investigative development of a "smart" program for automated swallowing sound detection.

FUTURE WORK

Given that our goal was 100% accuracy, the results of this study are only a beginning for research in this field. Future work includes retooling of the code using RMS values as the selected feature (as well as combining the "quiet" segments). With a robust program, future testing includes varying test subject ages and test media (bolus textures) as well as creating a program-user interface for assistance in decision-making. For example, for regions where the program is unsure of label, the user would be presented with a labeling choice.

ACKNOWLEDGMENT

This research has been appreciatively supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Reddy, N.P., Canilang, E.P., Grotz, R.C., Rane, M.B., Casterline, J., Costarella, B.R., "Biomechanical Quantification for Assessment and Diagnosis of Dysphagia," *IEEE Eng. Med. & Bio. Mag.*, Sept. 1988.

[2] Reddy, N.P., Costarella, B.R., Grotz, R.C., Canilang, E.P., "Biomechanical Measurements to Characterize the Oral Phase of Dysphagia," *IEEE Trans. on Biomed. Eng.*, vol. 37, no. 4, pp. 392-397, April 1990.

[3] Reddy, N.P., Costarella, B.R., Grotz, R.C., Canilang, E.P., "Biomechanical measurements to Characterize the Oral Phase of Dysphagia," *IEEE Trans. on Biomed. Eng.*, vol. 37, no. 4, pp. 392-397, April 1990.

[4] Rempel, G., Pasterkamp, H. and Moussavi, Z., "The Effect of Altering the Consistency of Food on the Breath and Swallow Pattern of Young People with Cerebral Palsy," *53rd Ann. Cong. of American Academy for Cerebral Palsy and Development Med.*, Sept. 1999.

[5] Moussavi, Z.K., Leopando, M.T., Rempel, G.R., "Automated Detection of Respiratory Phases by Acoustical Means," *Proc. of the 20th Annual Int. Conf.*, IEEE-EMBS., vol. 20, no. 1, pp. 21-24, 1998.

[6] Yap, Y., Moussavi, Z.K., "Acoustical Airflow Estimation From Tracheal Sound Power," *Proc. IEEE CCECE*, May 2002.

[7] Moussavi, Z., *Lecture notes of Biomedical Engineering Course*, Dept. of Elect. Eng., Univ. of Manitoba, 2001.

[8] Firmin, H., Reilly, S., Fourcin, A., "Non-Invasive Monitoring of Reflexive Swallowing," *Univ. College London, Dept. Phon. & Ling.;* http://www.phone.ucl.ac.uk/home/sh110/firmfour/swallow.htm (Nov. 9, 2001).

[9] Proakis, J.G., Manolaki, D.G., *Digital Signal Processing; Principles, Algorithms, and Applications*. New Jersey: Prentice-Hall Inc., 1996.

[10] Duda, R.O., Hart, P.E., Stork, D.G., *Pattern Classification*. NY: John-Wiley & Sons, Inc., 2001.