33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Swallowing Detection by Sonic and Subsonic Frequencies: a Comparison

Juan M. Fontana, Pedro L. Melo and Edward S. Sazonov, *Member, IEEE*

Abstract—**The detection of swallowing events by acoustic means represents an important tool to assess and diagnose swallowing disorders as well as to objectively monitor ingestive behavior of individuals. Acoustic sensors used to register swallowing sounds may also capture sound artifacts arising from intrinsic speech and external noise affecting the detection. In this paper we tested if subsonic frequencies are less prone to artifacts from speech, chewing and other intrinsic sounds than sonic frequencies. A simple method using a throat and an ambient microphone was employed to compare the swallowing detection accuracy by acoustic signals acquired in the sonic (20-2500 Hz) and subsonic (≤ 5 Hz) ranges. Averaged recall values were higher than 85% for both ranges. However, averaged precision values of 50% for subsonic frequencies and of 42% for sonic frequencies were caused by a high number of false positives. These results indicated no significant difference between averaged precision values which may suggest that subsonic frequencies were not less prone to intrinsic sound artifacts than frequencies in the sonic range. Further examination with the addition of a signal classification layer is proposed as a future step to confirm this statement.**

I. Introduction

SWALLOWING is a complex neuromuscular activity composed of three phases and controlled by different neurological mechanisms: oral phase, pharyngeal phase and esophageal phase. During the pharyngeal phase, a bolus of food created in the oral phase passes through the pharynx causing swallowing sounds. Detection and analysis of these sounds by non-invasive methods can be of special importance to assess and diagnose damages in certain areas of the brain and associated nerves [1], [2]. As an example, an automatic swallowing detection system can be used to diagnose dysphagia, which is a swallowing dysfunction that may lead to aspiration, choking and even death. On the other hand, monitoring of swallowing by acoustic means can be used to objectively measure the ingestive behavior of individuals by detecting periods of food intake. A direct application of this method is to gain understanding of etiology of obesity and overweight [3], [4].

In the field of swallowing analysis, the gold standard method is video fluoroscopy, which allows the visualization of recorded swallows in slow motion for a more accurate diagnosis. However, it is an invasive, time-consuming procedure that involves expensive equipment and radiation exposure [5], [6]. Non-invasive methods have been proposed for swallowing analysis by evaluating signals recorded by means of accelerometers [7], [8] or microphones. The resulting signals are used to train models in complex pattern recognition systems that would help to identify swallowing pattern. Use of accelerometers may not allow a reliable detection of swallows due to sensitivity to orientation in gravity field and to body motion. Acoustic sensors, on the other hand, are less sensitive to body motion leading to more effective systems. However, these sensors may capture sound artifacts arising from intrinsic speech and external noise affecting the accuracy of detection. Previous studies [9-11] assessed swallowing sounds signals using pattern recognition techniques in applications to dysphagia. The reported recognition rates were up to 93% [11] although speech and ambient noise was not included in the sound recordings.

In [12] and [13], swallowing sounds were collected during periods of food intake by using a throat microphone. The acoustic signals were contaminated by head motion artifacts, external noise and speech. The methodology implemented in those studies combined features from the sonic range of frequencies with supervised classification methods. The models obtained were robust to sound artifacts although the high computational burden of the methods may limit their practical application to large datasets.

An accurate recognition of swallowing sounds with a high rate of sound artifacts rejection arises as one of the main needs in the field. The swallowing process consists of several nonstationary short bursts of acoustic energy observed within a large frequency range (0-8000 Hz) [14]. On the other hand, the spectrograms of speech and other sound artifacts present more stationary patterns than swallowing sound in the sonic frequency range [15]. Most of the previous studies assessed swallowing sounds in the sonic range due to most of the signal power is contained in this range. However, the analysis of the acoustic signals in the subsonic range of frequencies may be addressed as a possible way to avoid interference with speech and other sound artifacts as previous studies showed promising results when swallowing sounds in this range were used to describe

Manuscript received March, 25 2011. This work was partially supported by a grant 1R21DK085462 "Objective Monitoring of Energy Intake and Ingestive Behavior in a Free Living Population" from the National Institutes of Health. This work was also supported by the Brazilian Council for Scientific and Technological Development (CNPq) and Rio de Janeiro State Research Supporting Foundation (FAPERJ).

Juan M. Fontana is with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487 USA (e-mail: jmfontana@bama.ua.edu).

Pedro L. Melo is with the Biomedical Instrumentation Laboratory, Rio de Janeiro State University, Rio de Janeiro, Brazil (email: plopeslib@gmail.com).

Edward S. Sazonov is with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487 USA (phone: 205-348-1981; e-mail: esazonov@bama.ua.edu)

physiological events during water swallowing [16], [17].

The goal of this study was to verify if subsonic frequencies are less prone to artifacts from speech, chewing and other intrinsic sounds than sonic frequencies. A simple method for swallowing detection was used to test the incidence of sound artifacts at both frequencies ranges. The proposed approach used a throat and an ambient microphone to register the sounds generated during periods of talking and food ingestion. Performances of sonic and subsonic signals were evaluated in terms of swallowing detection accuracy. The main question addressed by the study was if registering the sound of swallowing in subsonic range can lead to better rejection of intrinsic sound artifacts and thus to higher accuracy of swallowing detection.

II. METHODS

A. Data collection

Data was collected from 7 healthy volunteers who did not present any medical condition that would interfere with normal swallowing. Each subject participated in a single visit, which consisted of three parts: 1) 5 minutes of resting where the subject remained seated quietly; 2) 5 minutes of reading aloud; and 3) a meal period consisting of unlimited time to eat four food items. The consumption of each food item was further divided into ten repetitions of the following sequence: a) subject talking for a 20 seconds period, and b) a single bite-chewing-swallow phase. In this way, we ensured the presence of speech during food intake and that at least 10 swallowing instances were recorded for each food item. These food items consisted on: an apple, 40 g of crackers, can of low-fat yogurt, and 250 ml of water. These items were selected to represent different physical properties of everyday foods, which ensured that the proposed approach was tested on a representative sample.

Volunteers were monitored by a multi-modal sensor system that included acoustic swallowing sensor, ambient microphone and self-report signal (Fig. 1). Swallowing sounds in the subsonic range were detected by a condenser microphone whereas swallowing sounds in the sonic range were detected by a piezoelectric microphone. The rationale behind using two different microphones was based on that results of previous studies indicated that they are the best performers in each category (sonic and subsonic) [4], [8].

The two microphones were implemented in two different data collection setups. In the first setup, an IASUS NT (IASUS Concepts Ltd) microphone was placed over the laryngopharynx to acquire swallowing sounds in the sonic frequency range of 20-2500 Hz. Swallowing sounds beyond 2500 Hz were not relevant for this study because they are not frequent and only appear as sudden peaks of duration substantially shorter than a normal swallow [14]. This microphone provided a dynamic range of 46 ± 3 dB. A DC coupled amplifier was designed for conditioning the acoustic signal. The gain of the amplifier was set experimentally to reliably capture swallowing sounds while avoiding saturation of amplifiers during normal speech. The top-right graph in Fig. 2 presents an example of a signal recorded with the sonic microphone during a period of talking and yogurt ingestion.

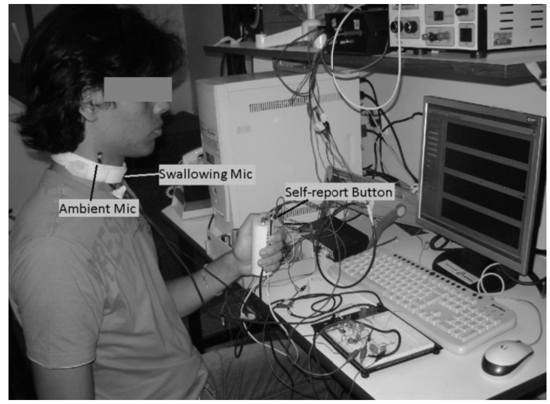

Fig. 1. Data collection setup dedicated to subsonic measurements

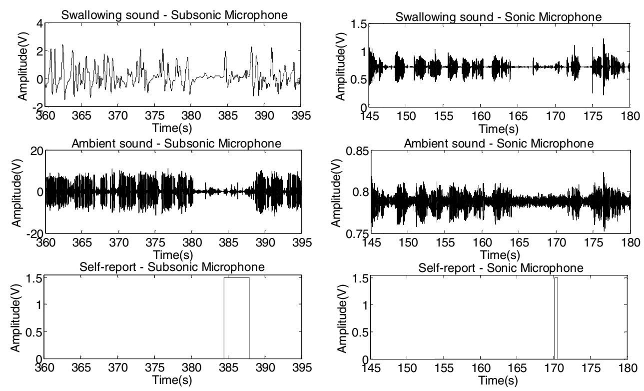

Fig. 2. Throat microphone (top), ambient (middle) and self-report (bottom) signals for subsonic (left) and sonic (right) data collection setups recorded during talking and ingestion of yogurt.

In the second setup, a CZN-15E microphone was placed on the throat at the level of the thyroid cartilage. This cartilage surrounds part of the larynx, thus the location may slightly differ from the sonic microphone location but without affecting the performance. It presented a signal-to-noise ratio of 60 dB and a sensitivity of -44 dB (30-18000 Hz bandwidth). The obtained acoustic signal was amplified using a DC amplifier and then filtered with an analog 4^{th} order Butterworth low-pass filter. The filter's cutoff frequency of 5 Hz only allowed the presence of subsonic frequency components in the acquired signal. The top-left graph in Fig. 2 illustrates an example of subsonic microphone signal recorded during talking and yogurt ingestion.

Each data collection setup was provided with a condenser microphone placed at one side of the subject's neck to record ambient sounds (Fig. 2). Such placement allowed to reliably capture sounds of intrinsic speech while not capturing the swallowing sound. Examples of these ambient signals are shown in the middle graphs of Fig. 2.

During all three parts of the visit, subjects were asked to press a handheld push-button to mark every swallow instance as a 1.5 V pulse (bottom graphs of Fig. 2). These marks were considered the true swallows instances in the evaluation phase. Amplified signals were collected through a data acquisition board at a sampling rate of 5000 Hz. This sampling rate was chosen to allow collection of signals in the subsonic and sonic range of frequencies while rejecting high-frequency noise.

B. Signal processing

The detection of swallowing events was performed by a simple method that used signals from the throat and the ambient microphone. The suggested methodology was based on the fact that the ambient microphone would register intrinsic speech but not swallowing sounds. Therefore, the intervals of speech could be identified in the ambient sound signal and then removed from the signal of the throat microphone to detect swallows. While this method may not be very practical in everyday noisy environments, it enabled very simple comparison of the microphones in controlled laboratory conditions. This method was applied to both sonic and subsonic microphone setups.

A moving average filter was used to extract the envelope of the ambient microphone signal. Two different windows widths, 1.0 s and 1.5 s were tested in this stage to determine their influence in the swallowing detection process. The periods where the subject was talking were detected by placing a threshold on the envelope. The optimal threshold value was selected based on the levels of subject's speech observed in the reading signal. Talking intervals were removed from the swallowing sound signal by zeroing the values in the same time interval. The resulting signal was smoothed using a moving average filter and anew threshold level was used to detect swallows. A swallowing instance was counted when the amplitude of the smoothed signal was higher than the threshold level for a period longer than 0.6 seconds, which represents the minimum time needed to cover a complete swallow [4].

The performances of each microphone setup was evaluated using the number of true positives (T_+), false negatives (F_-) and false positives (F_+). A true positive was counted when the algorithm and the self-report button indicated the presence of swallowing; a false positive was counted when the algorithm marked a swallow that was not present in the self report signal. Finally, a false negative was counted when the self-report signal indicated a swallow that was not marked by the proposed algorithm. Precision, recall and accuracy were calculated to compare microphones performances:

$$\text{Precision} = \frac{T_+}{T_+ + F_+}\ ;\quad \text{Recall} = \frac{T_+}{T_+ + F_-} \qquad (1)$$

$$\text{Accuracy} = (\text{Precision} + \text{Recall})/2 \qquad (2)$$

III. RESULTS

For each type of food ingested, precision, recall and accuracy values were calculated. When only subsonic frequencies were present in the analysis of the swallowing signal, the proposed method was able to achieve an accuracy of 68.2% (50.1% precision and 86.1% recall) averaged across all subjects and food items. On the other hand, when frequency components in the sonic range were included in the analysis, the accuracy of swallowing detection slightly decreased to 67% (41.8% precision and 92.2% recall) averaged across all subject and food items. The average accuracy values decreased (67.9% for subsonic and 64.7% for sonic frequencies) when the window width in the speech detection stage was reduced from 1.5 s to 1.0 s.

When each food item was analyzed separately (Fig. 3), there was no major differences in the detection of swallowing instances when using sonic and subsonic frequencies in any case. Water was the food item that presented the highest accuracy with 73.9% for subsonic frequencies and 72.6% for sonic frequencies. The lower detection accuracy was observed for cracker ingestion (64% for subsonic frequencies and 58.7% for sonic frequencies).

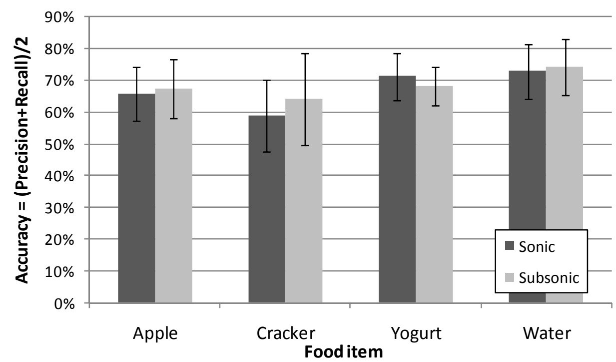

Fig. 3. Comparison of the best performances of swallowing detection for each food type using subsonic and sonic frequencies

IV. DISCUSSION

This study evaluated if the swallowing sounds registered in the subsonic range were less prone to speech, chewing and other intrinsic sounds than swallowing sounds registered in the sonic range. The top plots in Fig. 2 clearly illustrates the existence of speech in the throat microphone signal in both ranges. The first 20 seconds and the last 7 seconds in all plots corresponded to subject talking aloud. During those periods, sounds caused by the subject's voice were observed in both ambient (middle plots) and throat (top plots) microphone signals. This high influence of the voice in the throat microphone signal suggested that it was extremely important to implement a speech removal phase in the methodology to detect instances of swallow.

The lower accuracy values observed for both sonic and subsonic ranges were mainly due to the high number of false positives observed (low precision values). A precision value of 50.1% for the subsonic range indicated that only half of the swallows detected by the algorithm were true swallows. A reason for the high number of false positives may be the presence of chewing and other intrinsic sounds (i.e. head motion, respiration, throat cleaning, etc) that were captured by the microphone. These artifacts may have not been registered by the ambient microphone; consequently they were not captured by the ambient envelope and were not removed from the throat microphone signal. Another reason may be that those sounds showed up as short time events in the ambient signal and were discarded by the ambient envelope. In fact, the ambient signals on Fig. 2 illustrate that during the food ingestion periods (380-388 s for the left-hand side plots and 165-172 s for the right-hand side plots) the ambient microphone registered sounds other than talking.

The higher accuracy value for water swallowing with respect to apple and cracker was expected due to chewing is not present when drinking liquids. However, even though water presented higher precision, this value was still low suggesting that intrinsic sounds other than chewing were also registered during food ingestion.

Based on the assumptions mentioned in the paragraphs above, the higher precision value of the subsonic microphone (50.1%) may be an indication of a better rejection of chewing and other intrinsic throat sounds in the subsonic frequency range compared to that of the sonic range (precision value 41.8%). However, the lower recall value of the subsonic microphone (86.4% vs. 92.2%) did not allowed to conclusively establish this advantage.

The proposed methodology could be improved by adding a pattern recognition stage after speech has been removed. Support Vector Machines models have been proven to detect swallowing instances relying on high frequency components of swallowing sounds [4]. These models showed high efficiency for separating swallowing sounds from artifacts originated from respiration, head motion, etc. For the presented approach, new models would need to be trained and evaluated using frequency-domain features covering the subsonic range. A classification layer would help to discriminate swallowing sounds from intrinsic sounds by extracting unique characteristics of swallows that are not present in other sounds.

V. Conclusion

Rejection of intrinsic sound artifacts in sonic and subsonic frequency ranges was tested by processing throat microphone signals from several volunteers. A simple methodology for swallowing detection was used to evaluate the performances of each frequency range. The proposed approach detected most swallowing events but with high number of false positives caused by chewing and other intrinsic sounds captured by the throat microphones. Precision value of 50% averaged across all subjects may suggest that subsonic frequencies were sensitive to sound artifacts. The implementation of a pattern recognition algorithm may potentially help to confirm these findings by discriminating swallow sounds from other intrinsic sounds.

References

[1] T. Nakamura, Y. Yamamoto, and H. Tsugawa, "Measurement system for swallowing based on impedance pharyngography and swallowing sound," in *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE*, 2000, vol. 1, pp. 191-194 vol.1.

[2] M. Aboofazeli and Z. Moussavi, "Automated classification of swallowing and breadth sounds," in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, 2004, vol. 2, pp. 3816-3819.

[3] E. S. Sazonov et al., "Toward Objective Monitoring of Ingestive Behavior in Free-living Population," *Obesity (Silver Spring, Md.)*, vol. 17, no. 10, pp. 1971-1975, Oct. 2009.

[4] E. S. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. Neuman, "Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behavior," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 3, pp. 626-633, 2010.

[5] C. H. Smith, J. A. Logemann, L. A. Colangelo, A. W. Rademaker, and B. R. Pauloski, "Incidence and patient characteristics associated with silent aspiration in the acute care setting," *Dysphagia*, vol. 14, no. 1, pp. 1-7, 1999.

[6] T. A. Hughes, P. Liu, H. Griffiths, B. W. Lawrie, and C. M. Wiles, "Simultaneous electrical impedance tomography and videofluoroscopy in the assessment of swallowing," *Physiological Measurement*, vol. 17, no. 2, pp. 109-119, May. 1996.

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

[8] C. S. S. Greco, L. G. M. Q. Nunes, and P. L. Melo, "Instrumentation for bedside analysis of swallowing disorders," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 923-926.

[9] M. Aboofazeli and Z. Moussavi, "Analysis of swallowing sounds using hidden Markov models.," *Medical & Biological Engineering & Computing*, vol. 46, no. 4, pp. 307-314, Apr. 2008.

[10] M. Aboofazeli and Z. Moussavi, "Analysis of Temporal Pattern of Swallowing Mechanism," in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 2006, pp. 5591-5594.

[11] M. Aboofazeli and Z. Moussavi, "Automated Extraction of Swallowing Sounds Using a Wavelet-Based Filter," in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 2006, pp. 5607-5610.

[12] O. Makeyev et al., "Recognition of Swallowing Sounds Using Time-Frequency Decomposition and Limited Receptive Area Neural Classifier," in *Proceedings of AI-2008, The Twenty-eighth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Cambridge, 2008, pp. 33-46.

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

[14] J. A. Y. Cichero and B. E. Murdoch, "The Physiologic Cause of Swallowing Sounds: Answers from Heart Sounds and Vocal Tract Acoustics," *Dysphagia*, vol. 13, no. 1, pp. 39-52, Jan. 1998.

[15] S. R. Youmans and J. A. G. Stierwalt, "An Acoustic Profile of Normal Swallowing," *Dysphagia*, vol. 20, no. 3, pp. 195-209, Dec. 2005.

[16] G. P. Esteves, E. P. Silva, L. G. M. Q. Nunes, C. S. S. Greco, and P. L. Melo, "Configurable portable/ambulatory instrument for the analysis of the coordination between respiration and swallowing," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, pp. 90-93.

[17] C. S. Souza, J. A. Mesquita Júnior, and P. L. Melo, "A novel system using the Forced Oscillations Technique for the biomechanical analysis of swallowing.," *Technology & Health Care*, vol. 16, no. 5, pp. 331-341, 2008.