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

Towards Automated Ingestion Detection: Swallow Sounds

William P. Walker, *Student Member, IEEE*, Dinesh Bhatia, *Member, IEEE*

***Abstract*— Obesity is a worldwide epidemic and is a cause of many major chronic diseases. In most cases, obesity is a result of an imbalance between food intake and calories burned. Steps toward automated ingestion detection are being made. In order to automate the process of capturing ingestion, a method for detecting, analyzing, and recording sounds related to ingestion is being developed. In this paper, preliminary swallow sound analysis is presented and compared with various other noises captured from a throat mounted microphone. Initial frequency analysis indicates a stronger presence at high frequency intervals for swallow sounds in relation to other captured sounds such as voice. Comparisons show that a single high-pass filter can offer similar results as wavelet decomposition. Two simple methods for event detection are given.**

***Index Terms*— Health Monitoring Systems, Swallow Detection, Wireless Sensors**

I. Introduction

OBESITY is an epidemic spreading throughout the world [1]. With an increased Body Mass Index (BMI) comes an increased risk of a myriad of chronic diseases such as type 2 diabetes, hypertension, coronary heart disease, and sleep apnea [2] [3]. Along with these and other co-morbidities, obesity and overweight results in a reduced quality of life [4] and imposes a financial burden on the health care system [5].

Progress is being made toward the implementation of an obesity/chronic disease monitoring system which will aid both clinician and patient by automating the processes of capture, transmission, storage, processing, and display of information related to chronic disease and obesity. Presently, the ability to capture, transmit, store, and display information obtained from a pulse oximeter, blood pressure sensor, and digital weight scale has been presented [6] [7] [8] [9]. The platform for sensor interfacing and base processing are devices termed "Biote", which are designed in-house. The Biote consists of an MSP430F2619 microcontroller [10] and TI Chipcon CC2420 RF transceiver [11].

Principle to obesity management is the concept of energy balance, consisting of Energy Expenditure (EE) and Energy Intake (EI) [12]. Methods for the automated detection of EI are not as prevalent as those for EE detection. Amft et al. [13] [14] [15] [16] have performed extensive work in the three areas of eating gesture, chewing, and swallowing detection, for the purpose of automated EI detection. For eating gesture detection, accelerometers were placed on arms and sternum. For chewing detection, a microphone was placed at the ear canal to pick up chewing sounds via bone conduction. Classification of various foods into the types of dry-crisp, wet-crisp, and soft was made using these sounds emitted from chewing. Swallowing detection was performed using a Surface Electromyography (SEMG) sensor for detecting muscle activation in the throat and a sound sensor to detect presence of a swallowed bolus. A detection accuracy of around 73-75% was reported.

Sazonov et al. [17] [18] [19] used both a piezoelectric strain sensor and a microphone to detect the event of ingestion and an estimate of the type of ingested bolus. The strain sensor, located just below the ear, detected instances of chewing and the microphone, located just over the laryngopharynx detected instances of swallowing. For detecting instances of food ingestion, a higher level of accuracy was reported when using the results from both the strain sensor and the microphone than when either was used alone. The Instantaneous Swallow Frequency (ISF) was used to determine the type of ingestion (none, food, or liquid). ISF is calculated using the time between successive swallows. It was observed that the average ISF was highest for liquid ingestion and lowest for no ingestion, thus, thresholds could be set at differing ISF levels to determine what type, if any, of ingestion is occurring.

The work presented here is largely inspired by the works of Aboofazeli and Moussavi [20] and Yadollahi and Moussavi [21]. In both works, wavelet decomposition of the swallow sound recording is presented. In [20] swallow sounds are separated from breath sounds in healthy and dysphagic subjects with a high accuracy using a process involving wavelet decomposition and reconstruction and thresholding. In [21] a model for swallow sound generation is proposed, where the swallow process is given as the convolution of two systems. The first system is the interaction with the bolus as well as bone and muscle movement. The output of this system is a series of impulses and is the input to the second system. The second system

Manuscript received March 26, 2011.
W. P. Walker is with the Embedded and Adaptive Computing Group at the University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: wpw021000@utdallas.edu).
D. Bhatia is with the Embedded and Adaptive Computing Group at the University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: dinesh@utdallas.edu).

represents the skin, pharyngeal wall, and other tissue between the microphone and the bolus traveling down the esophagus. Wavelet analysis is shown to have the ability to extract the output of the first system. It is proposed that the number of impulses detected during a swallow may offer insight into the type of bolus being swallowed. It was observed that swallowing juice offered a higher number of impulses than swallowing a yogurt bolus.

The overall goal for the work presented here is to implement automated swallow detection in an existing Health Monitoring System (HMS) and provide feedback to the individual in order to promote awareness of consumption as well as serve as a tool for prompting the user for more information concerning the ingestion period. The remainder of this paper is organized as follows. The following section offers a discussion of swallow sounds and two methods used for comparison, wavelet decomposition and digital filtering. This paper is concluded with a discussion of two detection methods.

II. METHODS

A. Tracheal Recordings

The iASUS NT3 Throat Mic System (TMS) [22] was used to record all tracheal sounds. The TMS consists of an elastic band containing two surface microphones. The TMS is placed around the throat over the laryngeal prominence and is held in place by a magnetic clasp. All sound signals presented were obtained by connecting the TMS to the microphone input of an integrated sound card on a Personal Computer (PC) running Windows 7. Microphone levels were set to a volume of '5' and a boost of +20 dB. Recording was performed through the MATLAB™ object 'audiorecorder' at a sampling rate of 44.1 kHz at 24 bits per sample. A '.wav' file was created for each recording using the MATLAB™ function 'wavwrite'. Recordings were taken while the subject was in a sitting position. Unless otherwise noted, all data is assumed to have a sampling rate of 44.1 kHz. A throat microphone was chosen due to its relative insensitivity to outside sources of noise [19]. The recordings presented here were taken from two male subjects, labeled S1 and S2.

Recordings of various sounds emanating from the throat were made from S1 using the TMS. These sounds included the following: swallows (of a liquid, solid, and no substance), vocal cord activation (a hum, whisper, and speaking), clearing of the throat, and coughing. Fourier analysis in the form of the Short Time Fourier Transform (STFT) indicates that swallow sounds have a strong presence in the upper frequencies when compared with the other sounds.

B. Wavelet Decomposition

To obtain a higher temporal resolution than that offered by the STFT at high frequency intervals [23], wavelet analysis is implemented in the form of the Discrete Wavelet Transform (DWT). The Daubechies 20 wavelet [24] was chosen for decomposition/reconstruction. Decomposition is performed using the DWT available in the wavelet toolbox in MATLAB™. Wavelet decomposition through the DWT extracts characteristics of a signal within frequency intervals that decrease along a dyadic scale as the level of decomposition increases. The input signal at each level i is filtered into two signals, one containing the lower half of the frequency band of the input signal, called the "coefficients of approximation" (a_i) and one containing the upper half of the frequency band of the input signal, called the "coefficients of detail" (d_i). Both signals are then downsampled by 2 (every other sample is dropped). The coefficients of approximation are sent to the next level of decomposition. Note that this act of downsampling reduces the frequency interval of a_i at each level by 2. This process is continued until a desired level of decomposition is achieved or there are no samples left to pass to the next level.

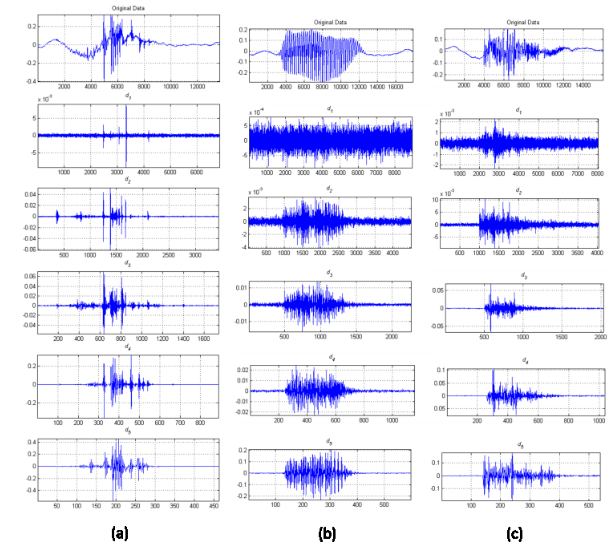

Fig. 1. First five levels of decomposition of a recording from S1 of (a) swallow of water, (b) vocal cord activation, and (c) a clearing of the throat. The DFT based on the Daubachies 20 scaling function is used.

Fig. 1 displays the result of 5 levels of wavelet decomposition (d_1 to d_5) of a swallow of water, vocal cord activation, and clearing of the throat, from left to right respectively. The recordings were taken from S1. The top row depicts the original signal.

The largest differences in amplitude of the d_i between sounds exist in the first and second levels of decomposition. In the given example (Fig. 1) and based on previous observations, swallow sounds resulting from the swallowing of a bolus have a greater presence in the d_i obtained from the first two levels of decomposition (d_1 and d_2) than most other sounds, such as voice and swallows with no bolus. The frequency interval covered by d_1 and d_2 is $\left(-\pi, -\frac{\pi}{4}\right) \cup \left(\frac{\pi}{4}, \pi\right)$, which is termed the Frequency Interval of Interest (FII) for this work. Note that the values given for this frequency interval are in radians, where π is half of the sampling frequency.

C. Digital Filtering

In the above, the DWT is used as a sequence of filters. Since the FII is the union of contiguous intervals (the high frequencies returned by the first two levels of decomposition), a single digital filter based on the ideal high-pass half-band filter [25] is implemented and comparisons with the DWT at two levels of decomposition are performed.

$$h_{high}(n) = \begin{cases} \frac{\sin \pi n - \sin\frac{\pi}{2}n}{\pi n}; n \neq 0 \\ \frac{\pi - \frac{\pi}{2}}{\pi} \quad\quad ; n = 0 \end{cases}$$

(1)

Equation (1) gives the impulse response for the ideal half-band high-pass filter used, which is infinite and non-causal. Thus, the impulse response given in (1) is time adjusted and windowed to create a causal Finite Impulse Response (FIR) filter. The impulse response is further adjusted to cover the FII.

D. Event Detection

Computational performance becomes more of an issue when using low-cost low-power microcontrollers such as the MSP430, which is the intended processing platform for the work described here. Keeping computational performance in mind, two simple event detection methods are devised, both aimed at detecting a high level of activity or presence in the FII.

The first event detection method involves creating a *Windowed Energy* (W_E) of the signal. Let the windowed energy of a signal $x(n)$ be the following.

$$W_E(i) = \sum_{n=i-W+1}^{i} |x(n)|^2$$

(2)

The windowed energy is the aggregate of the squared signal over a window of length W. Use of the rectangular window (1 over the desired interval, 0 everywhere else) was chosen for ease of computation. High values of W_E indicate a high amount of activity over the window. A simple thresholding or hill-climbing algorithm can be used to detect an event.

The second method for event detection scans the absolute value of the signal $x(n)$ over a window of length W for the maximum value, termed the *Windowed Maximum* (W_M). At each detected maximum, the value is saved if it is also the maximum over the next window of length W starting from the next sample.

III. Results

Fig. 2 depicts the results of two levels of decomposition and reconstruction using the DWT (top) and the signal filtered using the single high-pass filter based on (1) (bottom) with a filter order of 41, implemented on a five second recording containing a single swallow of an apple bolus of indeterminate size. The recording was made by S1. Evident in Fig. 2, the results of the single high-pass filter and the DWT at two levels of decomposition offer similar results at a similar amplitude.

Implementation of the single high-pass filter requires a single convolution between data and filter impulse response. Implementation of the DWT requires two convolutions at each level. Given filter impulse responses of relatively equal size (40 for each filter based on the Daubachies 20 wavelet, and 41 for the single high-pass filter) and the FII given above, implementation of the single high-pass filter will computationally outperform the DWT. It is a simple matter to show this and has been observed in simulations. It should be noted that this assumes no parallel processing, which is in line with a low-power low-cost microcontroller. Further, this assumes the FII proposed above. As the cutoff frequency for the FII approaches zero, the order of the high-pass filter must be increased to offer a proper frequency response. An increased deviation between results of the wavelet decomposition and the high-pass filter has been observed for levels of decomposition beyond 3 or 4 when using a high-pass filter order of 41. The authors have found that increasing the high-pass filter order by a factor of 2 for each increased level of decomposition offers results more in line with the wavelet decomposition. Increasing the filter order will, however, result in a corresponding increase in execution time.

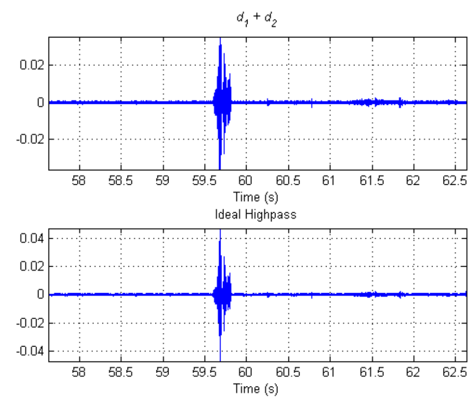

Fig. 2. Five second sample of a recording of S1 eating an apple. A single swallow is captured in this window, occurring just after the 59.5 second mark. The top plot is the result of the combination of d_1 and d_2. The bottom plot is the result of using the single high-pass filter covering the FII as described in the text.

Fig. 3 depicts the results from implementing W_E and W_M on a five second recording of S2 swallowing water. A swallow at the six second mark, not easily seen in the original time series recording, is detected as well. All swallows were noted manually during the recording process, and the notations are used as a means for verification. It should be noted that W_E and W_M do not offer swallow detection, but event detection, as some of the other possible noises such as clearing the throat and coughing can have a presence in the FII. A method for classification [26] will prove helpful in further separating swallow sounds from these various other sounds.

IV. Conclusion

This paper gives preliminary analysis of tracheal sound recordings with emphasis on separating swallow sounds from non-swallow sounds. The intended processing application is a low-cost low-power microcontroller; as such, light weight computational algorithms for ingestion detection are desired. The suggested approach is to filter the sound recordings using a high-pass filter. Two simple event detection methods, to be used on the filtered signal, are suggested. Further classification is likely necessary. Currently, Kohonen's Self Organizing Maps (SOM) [27] are being explored for classification. Future work includes refining the detection algorithms as well as examining recordings taken from a larger sample size of subjects.

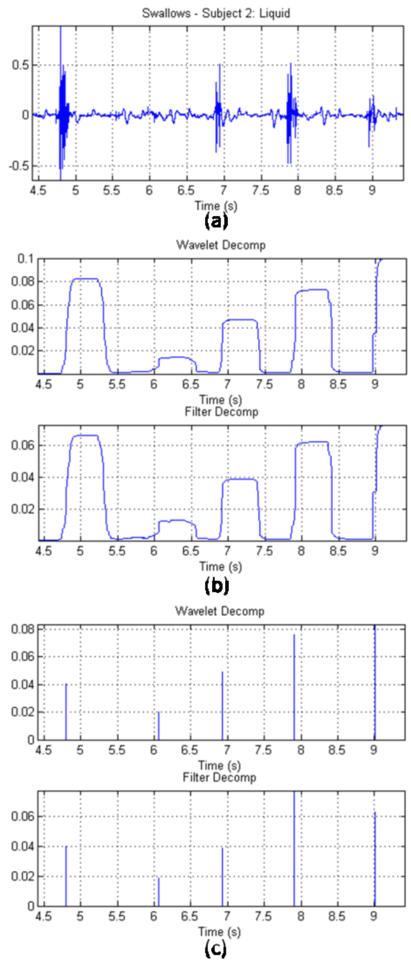

Fig. 3. Five second recording of (a) S2 swallowing water, (b) windowed energy of the recording with window size 0.5 seconds, and (c) windowed maximum with window size of 0.5 seconds.

REFERENCES

[1] S. Rossner, "Obesity: the disease of the twenty-first century," *International Journal of Obesity* , vol. 26, no. 4, 2002, pp. S2-S4.

[2] F. X. Pi-Sunyer, "The obesity epidemic: pathophysiology and consequences of obesity," *Obesity Research* , vol. 10, no. 2, 2002, pp. 97S-104S.

[3] A. E. Field, E. H. Coakley, A. Must, J. L. Spadano, N. Laird, W. H. Dietz, et al., "Impact of overweight on the risk of developing common chronic diseases during a 10-Year Period," *Archives of Internal Medicine* , vol. 161, 2001, pp. 1581-1586.

[4] A. Anandacoomarasamy, I. D. Caterson, S. Leibman, G. S. Smith, P. N. Sambrook, M. Fransen, et al., "Influence of BMI on health-related quality of life: comparison between an obese adult cohort and age-matched population norms," *Obesity* , vol. 17, 2009, pp. 2114-2118.

[5] E. A. Finkelstein, J. G. Trogdon, J. W. Cohen, and W. Dietz, "Annual medical spending attributable to obesity: payer and service specific estimates," *Health Affairs* , vol. 28, no. 5, 2009, pp. w822-w831.

[6] A. L. Aroul, D. Bhatia, and L. Estevez, "Energy-efficient ambulatory activity monitoring for disease management," *Proceedings of the 5th International Workshop on Wearable and Implantable Body Sensor Networks*, 2008, pp. 201-204.

[7] A. L. Aroul, A. Manohar, D. Bhatia, and L. Estevez, "Power efficient multi-band contextual activity monitoring for assistive environments," *Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments*, vol. 282, 2008, pp. 19.

[8] W. Walker, T. Polk, A. Hande, and D. Bhatia, "Remote blood pressure monitoring using a wireless sensor network," *IEEE Sixth Annual Emerging Information Technology Conference*, 2006.

[9] W. Walker, A. L. Aroul, and D. Bhatia, "Mobile health monitoring systems," *Engineering in Medicine and Biology Society, EMBC 2009*, 2009, pp. 5199-5202.

[10] *16-bit Ultra Low Power Mixed Signal Microcontroller*. (n.d.). (Texas Instruments Inc.) Available: http://www.ti.com/msp430

[11] *2.4 GHz IEEE 802.15.4 RF Transceiver.* (n.d.). Available: http://www.chipcon.com/files/CC2420_Data_Sheet_1_4.pdf

[12] A. M. Sharma and R. Padwal, "Obesity is a sign – over-eating is a symptom: an aetiological framework for the assessment and management of obesity," *Obesity Reviews* , vol. 11, no. 5, 2009, pp. 362-370.

[13] Amft, O., & Troster, G. (2009, April/June). "On-Body Sensing Solutions for Automatic Dietary Monitoring". *Pervasive Computing* , 62-70.

[14] Amft, O., & Troster, G. (2008). "Recognition of Dietary Activity Events Using On-Body Sensors". *Artificial Intelligence in Medicine , 42*, 121-136.

[15] O. Amft, H. Junker, and G. Troster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," *IEEE International Symposium on Wearable Computers*, 2005, pp. 160-163.

[16] O. Amft, M. Stager, P. Lukowicz, and G. Troster, "Analysis of chewing sounds for dietary monitoring," *International Conference on Ubiquitous Computing*, 2005, pp. 56-72.

[17] E. S. Sazonov, and S. Schuckers, "The energetics of obesity: a review," *IEEE Engineering in Medicine and Biology Magazine* , January/February 2010, pp. 31-35.

[18] E. Sazonov, S. Schuckers, P. L. Meyer, O. Makeyev, E. L. Melanson, M. Neuman, et al., "Toward objective monitoring of ingestive behavior in free-living populations," *Obesity* , vol. 17, no. 10, 2009, pp. 1971-1975.

[19] E. Sazonov, S. Schuckers, P. L. Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, et al., "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiological Measurement* , vol. 29, 2008, pp. 525-541.

[20] M. Aboofazeli, and Z. Moussavi, "Automated extraction of swallowing sounds using a wavelet-based filter," *28th IEEE EMBS Annual International Conference*, 2006, pp. 5607-5610.

[21] A. Yadollahi, and Z. Moussavi, "A model for normal swallowing sounds generation based on wavelet analysis," *CCECE 2008*, 2008, pp. 827-830.

[22] *NT3 Throat Mic For Two-way Radios, Wargamers, Paintball, and Airsoft Enthusiasts*. (n.d.). (iAsus) Retrieved January 2011, from http://www.iasus-concepts.com/nt/nt-3.htm

[23] M. Vetterli, and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Transactions on Signal Processing* , vol. 40, no. 9, 1992, pp. 2207-2232.

[24] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics* , vol. 41, 1988, pp. 909-996.

[25] J. G. Proakis, and D. G. Manolakis, *Digital Signal Processing - 4th ed.*, Upper Saddle River, New Jersey: Pearson Education, Inc., 2007.

[26] R. Polikar, "Pattern recognition in bioengineering,". In M. Akay (Ed.), *Wiley Encycloedia of Biomedical Engineering*, vol. 4, New York: Wiley Interscience, 2006, pp. 2695-2716.

[27] T. Kohonen, "The self-organizing map," *Neurocomputing* , vol. 21, 1998, pp. 1-6.