Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behavior

Edward S. Sazonov*, *Member, IEEE*, Oleksandr Makeyev, *Member, IEEE*, Stephanie Schuckers, *Senior Member, IEEE*, Paulo Lopez-Meyer, Edward L. Melanson, and Michael R. Neuman, *Life Senior Member, IEEE*

***Abstract*—Our understanding of etiology of obesity and overweight is incomplete due to lack of objective and accurate methods for monitoring of ingestive behavior (MIB) in the free-living population. Our research has shown that frequency of swallowing may serve as a predictor for detecting food intake, differentiating liquids and solids, and estimating ingested mass. This paper proposes and compares two methods of acoustical swallowing detection from sounds contaminated by motion artifacts, speech, and external noise. Methods based on mel-scale Fourier spectrum, wavelet packets, and support vector machines are studied considering the effects of epoch size, level of decomposition, and lagging on classification accuracy. The methodology was tested on a large dataset (64.5 h with a total of 9966 swallows) collected from 20 human subjects with various degrees of adiposity. Average weighted epoch-recognition accuracy for intravisit individual models was 96.8%, which resulted in 84.7% average weighted accuracy in detection of swallowing events. These results suggest high efficiency of the proposed methodology in separation of swallowing sounds from artifacts that originate from respiration, intrinsic speech, head movements, food ingestion, and ambient noise. The recognition accuracy was not related to body mass index, suggesting that the methodology is suitable for obese individuals.**

***Index Terms*—Biomedical signal processing, obesity, pattern recognition, swallowing, wearable devices.**

I. Introduction

THE WORLD is still losing in the battle with the obesity epidemic. In 2005, according to World Health Organization (WHO), there were approximately 1.6 billion overweight and at least 400 million obese adults worldwide [1]. Current trends are unsettling: 2015 projections predict 2.3 billion overweight and 700 million obese adults worldwide. Obesity is one of the risk factors for development of chronic diseases and presents a serious health problem. A recent study [2] suggested that effects of obesity on global health may be comparable to those of cancer. Though the etiology of obesity is a topic of ongoing scientific debate, regulation of food intake may be an important factor for maintaining a healthy weight [3] in the environment that provides abundance of inexpensive, highly palatable, and energy-dense foods, while requiring only minimal levels of physical activity [4].

While various methods have been developed for accurate and objective characterization of physical activity [5], at the present time, there is no accurate, inexpensive, nonintrusive way for objective monitoring of ingestive behavior (MIB) in free-living conditions. The most precise method of measuring energy intake is the doubly labeled water (DLW) technique that provides accurate estimates of caloric energy intake over long periods of time (10–14 days), if subjects remain weight-stable. However, the DLW technique cannot identify daily intake patterns. Dietary self-report methods like food-frequency questionnaires [6], self-reported diet diaries [7], and multimedia diaries [8] have been shown to be inaccurate and under-report daily intake.

Our recent research [9] has shown that frequency of swallowing can serve as a predictor for accurate detection of food intake, differentiation between liquid and solid foods, and estimation of ingested mass, with high frequency of swallowing being indicative of ingestion. Thus, an affordable wearable MIB device can be created for objective characterization of food intake. Such a device would use the proposed acoustical method to detect both spontaneous and food intake swallows, as they happen throughout the day without any conscience input from the user. A second-stage algorithm [9] would use the recognized swallows to detect and characterize food intake based on the frequency of swallowing. Potentially, such a device can reduce intake under-reporting as: 1) monitoring is objective and does not rely on self-report; and 2) continuous capture of spontaneous swallows indicates whether the sensor system is being worn or not, thus preventing or detecting intentional misreport.

While the weight gain is ultimately determined by energy balance (energy intake minus energy expenditure) and the proposed MIB device by itself cannot capture the energy content of a meal, such a device can provide valuable information about ingestion that is not available at this time. The device can also help diagnose and treat dangerous behaviors, leading to weight

Manuscript received April 27, 2009; revised July 27, 2009. First published September 29, 2009; current version published February 17, 2010. This work was supported in part by the National Institute of Heart, Lung, and Blood under Grant R21HL083052-02. *Asterisk indicates corresponding author.*

*E. S. Sazonov is with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699 USA (e-mail: esazonov@ieee.org).

O. Makeyev, S. Schuckers, and P. Lopez-Meyer are with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY 13699 USA (e-mail: mckehev@cias.clarkson.edu; sschucke@clarkson.edu; lopezmp@clarkson.edu).

E. L. Melanson is with the Division of Endocrinology, Metabolism, and Diabetes, University of Colorado Denver, Aurora, CO 80045 USA (e-mail: ed.melanson@ucdenver.edu).

M. R. Neuman is with the Department of Biomedical Engineering, Michigan Technological University, Houghton, MI 49931 USA (e-mail: mneuman@mtu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2009.2033037

gain, such as unconscious snacking [10], night eating [11], and evening [12] or weekend overeating [13]. The device may also find applications in diagnostics and treatment of disorders not directly related to obesity, such as inadvertent weight loss (cachexia), anorexia, and bulimia, as well as dysphagia.

The proposed method for acoustical detection of swallowing events is the first and fundamental step in implementation of the wearable MIB device. The swallowing detection does not need to differentiate between spontaneous and food intake swallows as the food-intake detection relies only on frequency of swallowing events. Algorithms presented in [9] can be applied as the second step of processing to detect and characterize food intake from the sequence of swallows.

This paper demonstrates high accuracy of swallowing event detection by acoustical means on the largest dataset to date by the methodologies based on mel-scale Fourier spectrum (msFS) and wavelet packet decomposition (WPD) for time-frequency representation, and support vector machines (SVM) for automatic recognition of characteristic sound of swallowing. It also contains assessment of the size of a near-optimal time decomposition window and effects of the decomposition level and epoch lagging on accuracy of swallowing detection suggesting that epoch duration used in earlier publications may not be optimal. Furthermore, assessment of recognition accuracy as a function of subject's body mass index (BMI) shows that the proposed acoustical method is suitable for obese individuals. Finally, it is demonstrated that proposed methods have substantial tolerance to the sound artifacts resulting from food intake, intrinsic speech, and background noise, and therefore may be suitable for free-living conditions.

The description is organized as follows: Section II presents the background on assessment of swallowing sound signals and currently used automatic swallowing detection methods. Section III provides a brief description of the data collection process. Section IV presents a detailed description of the proposed methodology. Experimental results are presented in Section V followed by the discussion and conclusions.

II. Acoustical Detection of Swallowing Events

Currently, videofluoroscopy and electromyography (EMG) are considered the gold standard in studies of deglutition. Videofluoroscopy depends on bulky and potentially unsafe equipment, while EMG is too invasive due to frequently used subcutaneous placement of electrodes in the masseter, suprahyoid, and infrahyoid muscles [14] to avoid interference from the muscles of the neck. Other reported sensors include a variety of strain devices [14]–[16]. However, most of the reported results indicate that detection of swallowing by a laryngeal strain sensor is not appropriate for obese subjects because under chin adipose deposits inhibit reliable detection of swallows. Use of accelerometer placed over the suprasternal notch of trachea, as suggested by [17]–[19], may also be not appropriate for obese individuals for the same reasons. Detection of the characteristic swallowing sound created by the specific motion of laryngopharynx can be performed by a microphone that is significantly less invasive and more effective for obese individuals than the methods listed above.

Several methods have been proposed for assessment of swallowing sounds using signal processing and pattern-recognition techniques. Methodologies for automatic decomposition of the tracheal sound signal into swallowing and respiratory segments in applications to dysphagia are presented in [17]–[20]. Signal decomposition techniques utilized such features as autoregressive coefficients, root-mean-square values of the signal in time domain, average power of the signal within several frequency bands, waveform fractal dimension, and discrete wavelet transform on time windows (epochs) ranging in duration from 12.5 to 100 ms. Reported recognition rates were in the range from 78.54% [17] to 93% [19], although the sound recordings did not include any speech or noise.

Rejection of artifacts arising from ingestion, intrinsic speech, and external noise is an issue that needs serious consideration. In the MIB applications, artifacts such as breathing, talking, throat cleaning, head movements, etc., may be confused with swallowing, thus decreasing the accuracy of the recognition [21]. The feasibility of sound artifact rejection was tested in [22], where swallowing sound recognition was performed using the limited receptive area neural classifier in combination with short-time Fourier transform and continuous wavelet transform. The methods described in [22] achieved 100% accuracy in classification of swallowing sounds on a limited dataset containing swallowing sounds, motion artifacts, talking, and music, although practical applications to large datasets were limited by high computational burden of the method.

A recently reported method of automated swallowing detection that was tested in the presence of artifacts originating in talking, head movements, food ingestion, and respiration was presented in [23]. The data were collected from six healthy subjects using a sensor collar containing surface EMG electrodes and a stethoscope electret microphone. A total of 7.93 h of data with 1265 swallows were acquired. Feature similarity search combined with an agreement of the detectors fusion method was used for classification. Fourfold cross validation was used with threefolds used for training and onefold for validation. The average recognition rate of 70% was obtained for labeling epochs of 250 ms, as belonging to swallows or nonswallows.

In summary, acoustical detection of swallowing events, as presented herein, may present a noninvasive and convenient method suitable for use by obese individuals. However, the field of swallowing sound detection is relatively unexplored with a significant need to focus on realistic conditions with presence of various sound artifacts. Another key consideration is the choice of the epoch duration and lagging for signal analysis. Epoch sizes used in [18], [20], and [23] are substantially shorter (12.5–250 ms) than the average duration of a swallow (0.86 s) [24], and thus, may represent only a partial segment of a swallowing sound or require a large number of time lags. The goal of the methodology proposed, here, is to consider acoustical swallow recognition as a method that may be appropriate for obese individuals; compare two popular signal time-frequency decompositions; investigate selection of key parameters of time-frequency transforms, such as epoch duration and level of decomposition; and to test the proposed methods on a challenging dataset that resembles free-living conditions and includes artifacts of various origins.

III. Data Collection

The data used in this paper were collected from a human study reported in [25], where the details of the protocol, hardware, sensors, and reliability of the manual scoring procedure are reported, but with no attempt to automatically recognize swallowing events. The following is a summary of the human study. The subject population included 20 volunteers, of which seven had BMI greater than 30 (obese). Each subject participated in four visits, each of which consisted of a 20-min resting period, followed by a meal, followed by another 20-min resting period. Out of 80 collected visits, ten were discarded due to data collection errors [25]. Selection and sequence of foods were fixed for each meal and represented different physical properties of the food such as crispiness, softness/hardness, and tackiness, all of which may impact both the artifacts arising from chewing sounds and the swallowing sound itself. To evaluate the impact of a meal-time conversation on the accuracy of swallowing detection, the subjects were involved in a dialogue with a member of the research team during the second and fourth visits and ate in silence during the first and third visits. Additionally, background noise (city noise, restaurant noise, and music) was played during the second and fourth visits to simulate realistic environments where people may be eating. The subjects were monitored by a multimodal sensor system that included an IASUS NT (IASUS Concepts, Ltd.) throat microphone located over laryngopharynx. The microphone provided a dynamic range of 46 ± 3 dB with a frequency range of 20–8000 Hz. Amplified signals were recorded through a line-in input of a standard sound card at a sampling rate of 44 100 Hz. The recordings were manually scored to mark the boundaries of each swallow. The evaluation of interrater reliability reported in [25] showed high reliability of manual scores (MSs) (0.98 average intraclass correlation) for manual scoring of swallows.

IV. Methods

The proposed methods are based on popular time-frequency decompositions: msFS and WPD with classification performed by SVM. Time-frequency decomposition and feature extraction based on WPD and msFS is widely used for processing of physiological signals, such as heart sounds [26] and lung sounds [27]. SVM is a supervised learning method that has a sound theoretical basis, robust to overfitting (loss of generalization on noisy or incomplete data [28]), and capable of producing very complex decision boundaries.

A. Feature Extraction by Wavelet Packet Decomposition

First, the sound stream was split into a series of overlapping epochs with fixed duration D and step S. A Hanning window was applied to each epoch. Second, a time-frequency decomposition of each epoch was obtained using WPD creating 2^N wavelet packets (where N is the level of decomposition) [29]. A packet on the previous level is decomposed into two packets on the next level as $w_{2n}(t) = \sqrt{2}\sum_k h_k w_n(2t-k)$, $w_{2n+1}(t) = \sqrt{2}\sum_k g_k w_n(2t-k)$ where h_k is the low-pass finite-impulse response (FIR) filter and g_k is the high-pass FIR filter such as $g_k = (-1)^k h_{1-k}$. The WPD was computed using Coiflet C4 wavelet. Advantages of the Coiflet wavelet include near-linear phase, good amplitude response, and fast computation [30]. WaveLab [31] package for MATLAB was used to perform WPD. Third, each wavelet packet was converted into a scalar feature forming a feature vector f_i of length 2^N for each epoch. The chosen feature was the unbiased estimate of entropy [32]. Fourth, to account for the time-varying structure of a swallow, a time-lagged feature vector was produced by merging feature vectors of the K adjacent epochs: $f'_i = \{f_{i-K}, f_i, f_{i+K}\}$.

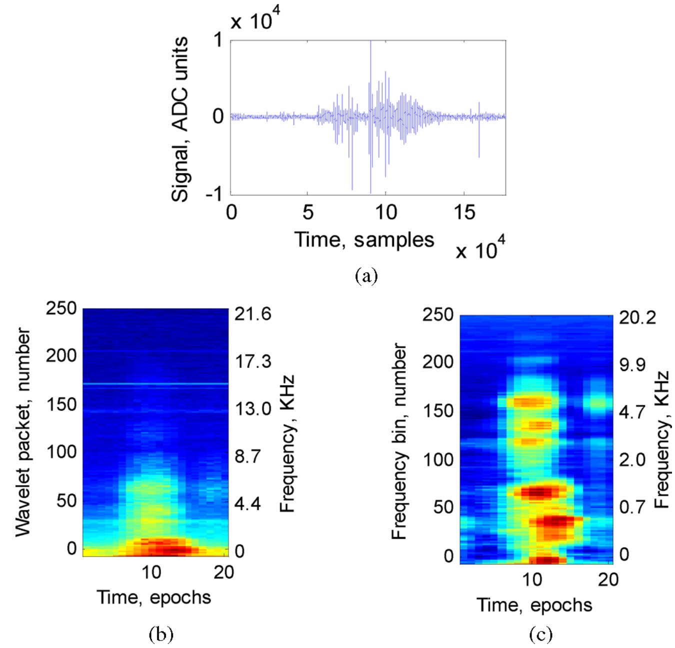


Fig. 1. (a) 4.0-s fragment of a sound recording including a swallow. (b) Features extracted by WPD processing. (c) Features extracted by msFS processing. Frequencies are shown for the center of each packet or bin.

B. Feature Extraction by msFT

First, segmentation of the sound signal into overlapping epochs was performed identically to the one used for WPD. Second, the Fourier amplitude spectrum $F(k)$ of length L was computed for every epoch. Third, a mel-scale triangle filter bank $M_i(k)$ [33] was used to computer 2^N point feature vector f_i (where N is an equivalent to WPD's level of decomposition) defined as $f_i = \log(\sum_{k=0}^{L/2} F(k)M_i(k))$, $i = 0, \ldots, N-1$. Finally, the time-lagged vector f'_i was obtained in the same way as for WPD. Fig. 1 shows a segment of the sound recording containing a swallow and its respective representation obtained by WPD and msFS processing with decomposition level $N = 8$, epoch duration $D = 1.5$ s, and step $S = 0.2$ s.

C. Support Vector Machines

The time-lagged feature vectors f'_i obtained either through WPD or msFS processing were used as inputs for training and validation of an SVM classifier [28]. The choice of the SVM as a classifier was defined by sound theoretical foundation and robust performance of SVM classifiers. A comparison of SVM performance with the performance of 16 classification and nine regression methods on 21 datasets for classification and 12 datasets for regression [34] ranked SVM as one the best techniques on most datasets, especially for classification. LibSVM package for MATLAB [35] was used for training the SVM classifier using the Gaussian radial-basis kernel function. Optimal parameters of the SVM classifier were found by a grid search procedure.

D. Optimal Epoch Duration and Decomposition Level

Optimal epoch duration D, epoch step size S, decomposition level N, and number of lags K were determined in a grid search procedure. The epoch duration and step size (D/S) were taken from a set $\{3.0/0.4, 1.5/0.2, 0.75/0.1, 0.375/0.05\}$ seconds that represents progressively finer time resolutions. Decomposition level both for WPD and msFS was taken as $N \in \{5, 6, 7, 8, 9\}$, thus producing features from 32 to 512 for each epoch. The number of lags K was either 0 or 1 since a higher number of lags produced long feature vectors that substantially slowed the classifier. Since a grid search procedure is time-consuming, it was performed on randomly selected two visits that included noise and talking during the meal, and thus, presented a harder classification case. The grid search procedure repeatedly trained classifiers defined by various combinations of D/S, N, and K. The validation accuracy was used to evaluate the goodness of parameters. Training and validation were performed with 34% of the data (one fold) used for training and 66% (two folds) used for validation. The accuracy of swallowing detection was estimated as described further.

E. Training and Validation

The pairs of feature vectors and class labels to be used as inputs for the SVM classifier were obtained in the following way: if any part of the epoch belonged to a swallow marked in the MS, the epoch label was set as “1” (swallow epoch), otherwise it was set as “-1” (nonswallow epoch). Individual intravisit models were built for 70 visits of 20 subjects. The training and validation sets were formed by taking into account the highly nonhomogeneous structure of each visit. For example, a period of quiet resting with no talking and no food intake will not have enough variability in the data to train a classifier that would work reliably if talking or food intake is introduced. Since talking, food intake, and external noise are introduced at various times in each visit, a longitudinal segmentation was used. Each visit was divided into 55 segments of equal duration, each segment of 1 min in duration on average. Threefold cross validation was performed with two folds used for training and one fold used for validation.

F. Accuracy of Detecting Swallowing Instances

Predicted class labels represent accuracy of the classifier on epoch level and do not correspond well to the accuracy of detection of swallowing events. Transition from the epochs to

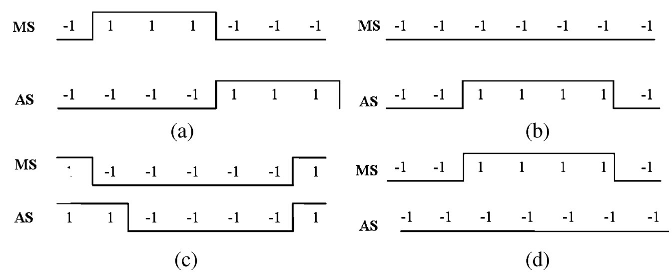


Fig. 2. Examples of (a) true positive, (b) false positive, (c) true negative, and (d) false negative. Each number represents a class label for an epoch (“-1”—nonswallow epoch, and “1”—swallow epoch).

swallowing events was done by identifying all situations where either MS or Automatic Score (AS) indicated presence of a swallow and calculating the numbers of true positives, false positives, and false negatives in terms of swallowing events. A true positive (T_+) was counted if both MS and AS contained continuous sequences of epochs marked as swallows intersecting at one or more epochs or on the sequence boundary [see Fig. 2(a)]. A false positive (F_+) was identified if the AS marked a swallow that was not present in the MS [see Fig. 2(b)]. A true negative (T_-) was counted if both MS and AS contained continuous sequences of epochs marked as nonswallows intersecting at one or more epochs [see Fig. 2(c)]. A false negative (F_-) was counted if the MS marked a swallow that was not present in the AS [see Fig. 2(d)]. The accuracy of swallowing events detection was then estimated using weighted accuracy $= T_+ + T_-/T_+ + T_- + F_+ + F_-$, sensitivity $= T_+/T_+ + F_-$, and specificity $= T_-/T_- + F_+$.

V. Results

The graphs obtained by the grid search of optimal epoch duration, decomposition level, and number of lags on a subset from two visits are shown in Fig. 3 that suggests the best parameters for WPD processing, i.e., ninth level of decomposition on 1.5-s epoch. For msFS processing, the best parameters are at the seventh level of decomposition on 1.5-s epoch. These parameters with and without lagging were used to process throat microphone signal collected in 70 visits. SVM training was performed with misclassification penalty $C = 10$ and Gaussian kernel width parameter $\gamma = 0.05$. Results obtained in per-epoch recognition and detection of swallowing events are presented in Table I.

The best average weighted accuracy in terms of epochs and swallows was produced by msFS-7 with three lags and found to be 96.8 $\pm$ 1.4% for epochs and 84.7 $\pm$ 6.9% for swallows. The distribution of average weighted accuracy in classification of epochs and swallowing events versus the subject’s BMI and corresponding linear fit of the data are presented in Fig. 4. To assess the impact of sound artifacts on accuracy of identifying swallowing events, the average weighted swallowing accuracy was also computed individually for the four nonoverlapping parts of the validation set corresponding to the following categories: periods of no food intake and no talking (88.0%), periods of no food intake with talking (86.4%), periods of food intake

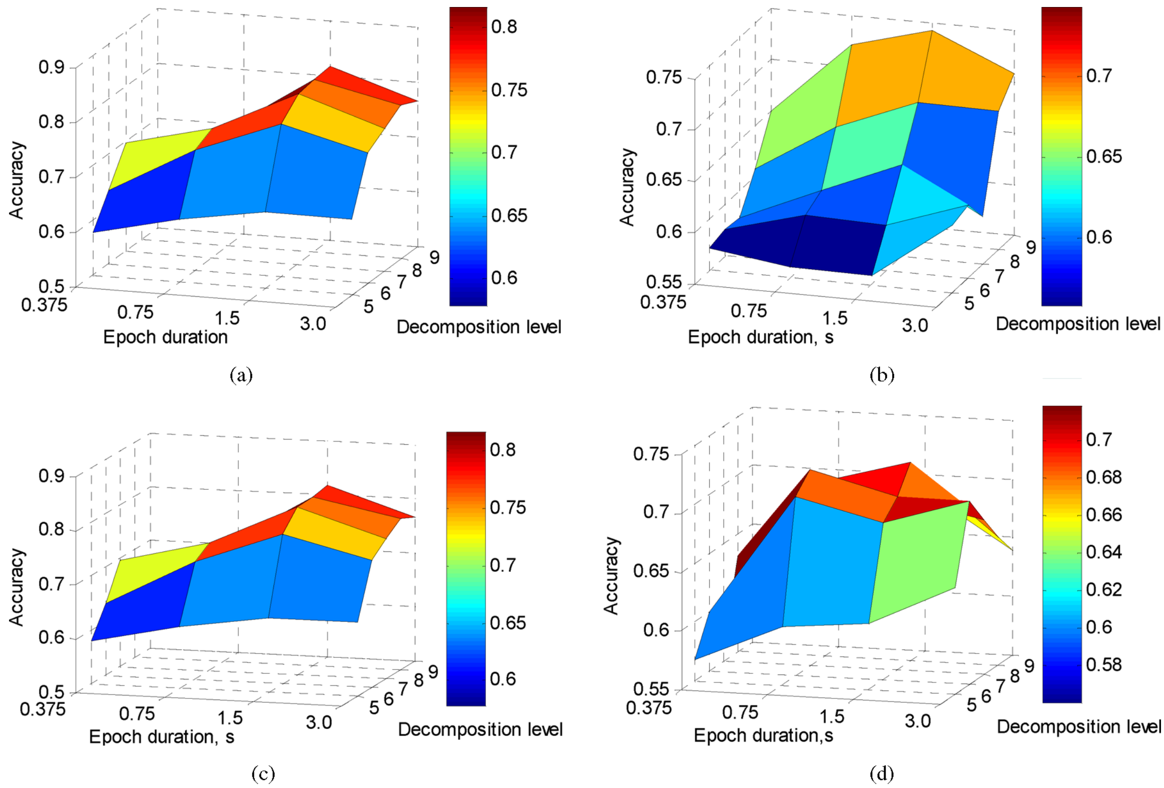


Fig. 3. Accuracy of swallowing sound recognition as a function of epoch duration and decomposition level (a) msFS with no lags. (b) WPD with no lags. (c) msFS with $K = 1$ (three lags). (d) WPD with three lags.

TABLE I
ACCURACY OBTAINED IN DETECTION OF SWALLOWING EVENTS FOR THREEFOLD CROSS VALIDATION

Feature	WPD 9	WPD 9	msFS 7	msFS 7
Number of lags	1	3	1	3
Average per-epoch accuracy (%)	95.9	96.4	96	96.8
Average per-swallow accuracy (%)	79.4	79.4	79	84.7

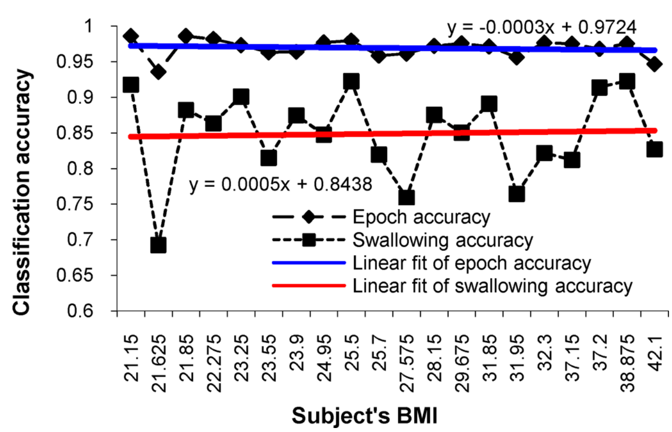


Fig. 4. Distribution of average weighted accuracy in classification of epochs and swallowing events versus subject's BMI and corresponding linear fit of the data.

and no talking and background noise (86.2%), and periods of food intake with talking and background noise (82.9%).

VI. DISCUSSION

One of the goals of this paper was to determine the optimal duration of an epoch, since durations reported in existing literature [18], [19], and [23] varied over a wide range of 12.5–250 ms. As Fig. 3 shows that the epoch duration of 1.5 s clearly demonstrates the highest recognition accuracy both for msFS and WPD with or without lagging. This corresponds well with the mean duration of swallow, which in our study was found to be 1.15 s with a standard deviation of 0.29 s (based on analysis of 10 686 swallows), comparable to previously reported duration of 0.86 s [24]. Thus, the epoch duration of 1.5 s is sufficient to completely include an average swallow. We believe that such a choice of the epoch duration is one of the reasons that our recognition rate is substantially higher than per epoch accuracy of 70% reported in [23] where the authors used a 0.25-s epoch that cannot cover a complete swallow. Although our study excluded dysphagic subjects, we may anticipate that recognition of longer than normal dysphagic swallows [36] may benefit from a longer epoch.

Fig. 3 also demonstrates accuracy growth with increase in the level of decomposition. The most pronounced increase for msFS is observed until the seventh level of decomposition. As Table I shows a lagged version produces higher overall accuracy due to better preservation of accuracy during transitioning from epochs to swallows on some of the visits. Lagging takes the feature evolution over time into account, and thus, produces more accurate results. The nonlagged version of WPD processing clearly peaks at the ninth level of decomposition and trends toward further growth. Unfortunately, higher levels of decomposition result in unacceptably long processing times both for feature extraction and classification. The lagged version of WPD behaves somewhat erratically (which may be attributed to a limited dataset used in the grid search procedure), but clearly peaks at the eighth level of decomposition, confirming that the eighth–ninth levels are probably near the optimum for WPD. Fig. 3 and Table I also demonstrate that msFS time-frequency decomposition clearly outperforms WPD resulting in higher recognition accuracy. A possible explanation is nonlinear scaling of the frequencies by msFS that allows for a better representation of the lower frequencies that contain most of the energy of a swallowing sound.

One of major advantages of the present study is that it was designed to be close to real-life conditions and include sound artifacts originating from chewing of food of different textures, talking, head movements, occasional intrinsic sounds (for example, coughing), and background noise of various origins. Thus, the classifier had to deal with a significantly more complex problem than previous studies while achieving a comparable (versus 93% in [19]) or better performance (versus 79% in [18]). The closest study that allows direct comparison is [23], which achieved the epoch-accurate average recognition rate of 70%. For comparison, the methodology proposed here yielded the average weighted epoch accuracy of 96.8% that relates to 84.7% average weighted accuracy in detection of swallowing events. Furthermore, our study utilized a wider variety of solid foods (cheese pizza, an apple, and a peanut butter sandwich) with varying physical properties that directly impact the sounds of mastication [37] and, subsequently, influence swallowing recognition. Another advantage is unrestricted consumption of liquids, which were limited in [23] to 5 and 15mL of volume at a time. Liquid consumption is characterized by a very high swallowing frequency [9] in which identification of individual swallows is difficult due to the fact that consecutive swallows may be recognized as one. The results show that artifact sounds negatively impact the recognition accuracy, but not to a degree that would render the method unusable. As expected, the highest recognition accuracy is observed for quiet periods of no food intake (88.0%), and the lowest recognition accuracy is observed for periods of food intake combined with talking and background noise (82.9%). Thus, application of noise cancellation techniques may further improve on the classification accuracy.

As Fig. 4 suggests, the accuracy of detecting swallowing events is very likely not to be dependent on the subject's BMI. While more data are needed to appropriately test the effect of obesity, this may be an important advantage of the acoustical approach of detecting swallowing events. The highest BMI of a volunteer in the study was 42.1, which is considered as severe (morbid) obesity. Even for this volunteer, the swallowing identification accuracy was greater than 80%. Thus, these results suggest that the proposed methodology could be used for monitoring of food intake in obese individuals.

The reported experimental results were obtained on the dataset containing 64.5 h of data with 9966 swallows collected from 20 subjects with the experimental conditions resembling those of food consumption in free-living. To our knowledge, this is largest dataset collected to date. In addition, the MS of swallows used for training of the classifiers has known reliability metrics [25]. Overall, the proposed methodology showed good performance in testing on a more complicated dataset than any of the previous studies. The next step in the development of the acoustical method of the detection of swallowing is development of intervisit individual and group models that could be practically applied for automatic scoring of the swallowing sound recordings. The desired accuracy of the identification of swallowing is another question that needs further investigation. However, the methods for detection of food intake and prediction of ingested mass [9] should offers some tolerance to the errors in detection of swallowing instances, since they rely on multiple swallows and relatively long time windows (up to 2 min).

The results of this paper also have important implications for the original intent to use automatic recognition of swallowing sounds in a wearable device for monitoring of ingestion. The time sequence of swallows detected by the proposed method can be further processed by algorithms in [9] to detect and characterize food intake to achieve real-time monitoring of ingestion. With the rapid progression of computing power available in modern ubiquitous platforms [cell phones, personal digital assistant (PDA)], the proposed MIB methodology can be implemented as a wearable device allowing for real-time biofeedback to individuals. Such a wearable device may potentially find numerous applications in research, clinical nutrition, and self-monitoring of food intake by general population.

VII. Conclusion

In this paper, we describe two automatic acoustical swallowing detection methods for use in MIB applications. The methods were based on combination of msFS or WPD and SVM. The proposed methodology was tested on the data collected from 20 human subjects with 35% of the subjects being obese with BMI of at least 30 and the average BMI of 28.53 using a multimodal data collection system designed for noninvasive monitoring of chewing and swallowing. The total duration of data used for training and validation was 64.5 h including 9966 swallows, which makes it the largest dataset to date. Average weighted epoch-classification accuracy of 96.8% resulted in 84.7% average weighted accuracy in detection of swallowing events. Optimal duration of a sound time slice was found to be 1.5 s that corresponds well to the statistics of swallowing duration. The msFS decomposition with three lags clearly outperformed WPD in recognition accuracy. A study of impact of food intake, talking, and background noise on accuracy of swallowing detection suggests robustness of the proposed methodology to such events, as well as its ability to accurately separate swallowing sounds from sound artifacts that originate in respiration, talking, head movements, food ingestion, and ambient noise. The method was also demonstrated to work equally well for both obese and nonobese subjects. The described methodology and sensors may be implemented in a wearable monitoring device, thus enabling MIB applications in free-living individuals.

Acknowledgment

The authors would like to thank the anonymous reviewers for their suggestions on improving this manuscript.

References

[1] World Health Organization. (2008, Oct.). Global Infobase for Overweight and Obesity, http://www.who.int/mediacentre/factsheets/fs311/en/index.html

[2] S. J. Olshansky, D. J. Passaro, R. C. Hershow, J. Layden, B. A. Carnes, J. Brody, L. Hayflick, R. N. Butler, D. B. Allison, and D. S. Ludwig, “A potential decline in life expectancy in the United States in the 21st century,” *New Eng. J. Med.*, vol. 352, pp. 1138–1145, Mar. 2005.

[3] J. P. Flatt, “Substrate utilization and obesity,” *Diabetes Rev.*, vol. 4, pp. 433–449, 1996.

[4] J. O. Hill, H. R. Wyatt, G. W. Reed, and J. C. Peters, “Obesity and the environment: Where do we go from here?,” *Science*, vol. 299, pp. 853–855, Feb. 2003.

[5] P. Ainslie, T. Reilly, and K. Westerterp, “Estimating human energy expenditure: A review of techniques with particular reference to doubly labelled water,” *Sports Med. (Auckland, N.Z.)*, vol. 33, pp. 683–698, 2003.

[6] J. L. Weber, P. M. Reid, K. A. Greaves, J. P. DeLany, V. A. Stanford, S. B. Going, W. H. Howell, and L. B. Houtkooper, “Validity of self-reported energy intake in lean and obese young women, using two nutrient databases, compared with total energy expenditure assessed by doubly labeled water,” *Eur. J. Clin. Nutr.*, vol. 55, pp. 940–950, Nov. 2001.

[7] J. M. D. Castro, “Methodology, correlational analysis, and interpretation of diet diary records of the food and fluid intake of free-living humans,” *Appetite*, vol. 23, pp. 179–192, Oct. 1994.

[8] C. H. Kaczkowski, P. J. Jones, J. Feng, and H. S. Bayley, “Four-day multimedia diet records underestimate energy needs in middle-aged and elderly women as determined by doubly-labeled water,” *J. Nutr.*, vol. 130, pp. 802–805, Apr. 2000.

[9] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, E. Melanson, M. Neuman, and J. Hill, “Toward objective monitoring of ingestive behavior in free living population,” *Obesity*, vol. 17, no. 10, pp. 1971–1975, 2009.

[10] C. Ward, *Compulsive Eating: The Struggle to Feed the Hunger Inside*. New York: The Rosen Publishing Group, 1998.

[11] A. J. Stunkard, W. J. Grace, and H. G. Wolff, “Night eating syndrome,” in *Eating Disorders and Obesity. A Comprehensive Handbook*. New York: Guilford Press, 2002, pp. 183–188.

[12] A. K. Kant, R. Ballard-Barbash, and A. Schatzkin, “Evening eating and its relation to self-reported body weight and nutrient intake in women, CSFII 1985–86,” *J. Amer. College Nutr.*, vol. 14, pp. 358–363, Aug. 1995.

[13] P. S. Haines, M. Y. Hama, D. K. Guilkey, and B. M. Popkin, “Weekend eating in the United States is linked with greater energy, fat, and alcohol intake,” *Obesity Res.*, vol. 11, pp. 945–949, Aug. 2003.

[14] C. Ertekin, I. Aydogdu, Y. Seçil, N. Kiylioglu, S. Tarlaci, and T. Ozdemirkiran, “Oropharyngeal swallowing in craniocervical dystonia,” *J. Neurol. Neurosurg. Psychiatry*, vol. 73, pp. 406–411, Oct. 2002.

[15] E. Stellar and E. E. Shrager, “Chews and swallows and the microstructure of eating,” *Amer. J. Clin. Nutr.*, vol. 42, pp. 973–982, Nov. 1985.

[16] M. Pehlivan, N. Yüceyar, C. Ertekin, G. Celebi, M. Ertaş, T. Kalayci, and I. Aydoĝdu, “An electronic device measuring the frequency of spontaneous swallowing: Digital phagometer,” *Dysphagia*, vol. 11, pp. 259–264, 1996.

[17] L. J. Lazareck and Z. K. Moussavi, “Automated algorithm for swallowing sound detection,” in *Proc. Can. Med. Biol. Eng. Conf.*, 2002, pp. 1–4.

[18] M. Aboofazeli and Z. Moussavi, “Automated classification of swallowing and breadth sounds,” in *Conf. Proc.: Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Conf.*, 2004, pp. 3816–3819.

[19] M. Aboofazeli and Z. Moussavi, “Automated extraction of swallowing sounds using a wavelet-based filter,” in *Proc. Conf. Proc.: Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Conf.*, 2006, pp. 5607–5610.

[20] M. Aboofazeli and Z. Moussavi, “Analysis of swallowing sounds using hidden Markov models,” *Med. Biol. Eng. Comput.*, vol. 46, pp. 307–314, Apr. 2008.

[21] A. Das, N. P. Reddy, and J. Narayanan, “Hybrid fuzzy logic committee neural networks for recognition of swallow acceleration signals,” *Comput. Methods Prog. Biomed.*, vol. 64, pp. 87–99, 2001.

[22] O. Makeyev, E. Sazonov, S. Schuckers, P. Lopez-Meyer, T. Baidyk, E. Melanson, and M. Neuman, “Recognition of swallowing sounds using time-frequency decomposition and limited receptive area neural classifier,” in *Proc. AI-2008, 28th SGAI Int. Conf. Innovative Tech. Appl. Artif.*, Cambridge, MA, 2008, pp. 33–46.

[23] O. Amft and G. Tröster, “Recognition of dietary activity events using on-body sensors,” *Arif. Intell. Med.*, vol. 42, pp. 121–136, 2008.

[24] J. B. Palmer, N. J. Rudin, G. Lara, and A. W. Crompton, “Coordination of mastication and swallowing,” *Dysphagia*, vol. 7, pp. 187–200, 1992.

[25] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman, “Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior,” *Physiol. Meas.*, vol. 29, pp. 525–541, 2008.

[26] I. Turkoglu, A. Arslan, and E. Ilkay, “An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks,” *Comput. Biol. Med.*, vol. 33, pp. 319–331, Jul. 2003.

[27] Y. Liu, C. Zhang, and Y. Peng, “Neural classification of lung sounds using wavelet packet coefficients energy,” in *Proc. PRICAI 2006: Trends. Artif. Intell.*, 2006, pp. 278–287.

[28] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[29] N. Addison, *The Illustrated Wavelet Transform Handbook*. New York: Taylor & Francis, 2002.

[30] S. Fu, X. Liu, B. Muralikrishnan, and J. Raja, “Wavelet analysis with different wavelet bases for engineering surfaces,” in *Proc. 16th Annu. Meeting Amer. Soc. Precision Eng.*, Raleigh, NC, 2001, pp. 249–252.

[31] J. B. Buckheit and D. L. Donoho, “WaveLab and reproducible research,” in *Lecture Notes in Statistics*. New York: Springer-Verlag, 1995, pp. 55–55.

[32] R. Moddemeijer, “On estimation of entropy and mutual information of continuous distributions,” *Signal Process.*, vol. 16, pp. 233–248, Mar. 1989.

[33] G. Wu and C. Lin, “Word boundary detection with mel-scale frequency bank in noisy environment,” *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 541–554, Sep. 2000.

[34] D. Meyer, F. Leisch, and K. Hornik, “The support vector machine under test,” *Neurocomputing*, vol. 55, pp. 169–186, 2003.

[35] C. Chang and C. Lin. (2001). LIBSVM: A library for support vector machines, [Online]. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm

[36] M. Vaiman and O. Nahlieli, “Oral vs. pharyngeal dysphagia: Surface electromyography randomized study,” *BMC Ear Nose Throat Disord.*, vol. 9, p. 3, 2009.

[37] N. D. Belie, M. Sivertsvik, and J. D. Baerdemaeker, “Differences in chewing sounds of dry-crisp snacks by multivariate data analysis,” *J. Sound Vib.*, vol. 266, pp. 625–643, Sep. 2003.

Edward S. Sazonov (M’02) received the Diploma of Systems Engineer degree from Khabarovsk State University of Technology, Khabarovsk, Russia, in 1993, and the Ph.D. degree in computer engineering from West Virginia University, Morgantown, in 2002.

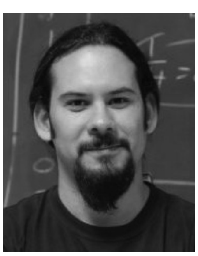
He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY, and the Head of the Clarkson Laboratory of Ambient and Wearable Devices, Clarkson University. His research interests include bioengineering, computational intelligence, wireless, ambient and wearable devices, development of methods and wearable sensors for noninvasive monitoring of ingestion, methods and devices for monitoring of physical activity and energy expenditure, wearable platforms for rehabilitation of stroke patients and monitoring of the risk of falling in elderly, and self-powered ambient sensors for structural health monitoring. His work has been supported by national (National Science Foundation, National Institutes of Health, and National Academies of Science) and state agencies, and private industry.

Oleksandr Makeyev (M'03) received the B.Sc. degree in mathematics and the M.Sc. degree in statistics from Taras Shevchenko National University of Kyiv, Kiev, Ukraine, in 2003 and 2005 respectively. He is currently working towards the Ph.D. degree in engineering science with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY.

His current research interests include development and application of computational intelligence-based pattern recognition methods to engineering problems, signal processing and pattern recognition for monitoring of ingestive behavior in humans.

Stephanie Schuckers (M'95–SM'07) received the B.S. degree in electrical engineering from the University of Iowa, Iowa City, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, in 1994 and 1997, respectively.

She is currently an Associate Professor with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY. Her research interests include processing and interpreting signals that arise from the human body, analysis of real data collected from human, cadaver, and animal studies. Her work has been supported from various sources, including National Science Foundation, American Heart Association, National Institute of Health, Department of Homeland Security, Center for Identification Technology, and private industry, among others. She is the author of more than 30 journal publications, as well as many conference papers and book chapters.



Paulo Lopez-Meyer received the Bachelor's degree in telecommunications engineering and the Master's degree in instrumentation engineering from the National Autonomous University of Mexico, Mexico City, Mexico, in 2003 and 2005, respectively. He is currently working toward the Ph.D. degree in biomedical engineering from Clarkson University, Potsdam, NY.

His research interests include the applications of machine learning and pattern recognition in the solution of real-life problems.

Edward L. Melanson received the M.S. and Ph.D. degree in exercise science from the University of Massachusetts, Amherst, in 1994 and 1999, respectively.

He began his research career as a Postdoctoral Fellow in the Center for Human Nutrition, University of Colorado Denver, in 1998, and joined the faculty in the Division of Endocrinology, Metabolism, and Diabetes, in 2003. His current research interests include the effects of diet, exercise, and obesity on substrate metabolism and energy expenditure. Currently, he is performing studies on the effects of different intensities of exercise and manipulations in dietary fat on fat oxidation. These studies are performed using whole-room indirect calorimetry. He also is currently performing an exercise intervention aimed at determining the effects of exercise on the different components of energy expenditure, particularly nonresting energy expenditure. In these studies, a variety of approaches are used to assess energy expenditure including doubly labeled water, indirect calorimetry, and accelerometry. His research is funded by the National Institutes of Health.

Michael R. Neuman (M'62–SM'93–LSM'04) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Case Institute of Technology, Cleveland, OH, and the M.D. degree from Case Western Reserve University, Cleveland.

In August 2003, he joined the Department of Biomedical Engineering, Michigan Technological University, Houghton, MI, as a Professor and Chair. He previously held the Herbert Herff Chair at the Memphis Joint Program in Biomedical Engineering and served for thirty-two years at Case Western Reserve University in the Departments of Biomedical and Electrical Engineering and the Departments of Reproductive Biology and Obstetrics and Gynecology. His current research interests include the application of microelectronic technology to problems in clinical medicine.

Dr. Neuman has served as the President of the International Society on Biotelemetry and was the Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING during 1989–1996. He also served as the Editor-in-Chief of the international journal, Physiological Measurement during 2002–2007, and is currently the Editor-in-Chief of the IEEE ENGINEERING IN MEDICINE AND BIOLOGY MAGAZINE.