Biomedical Signal Processing and Control 7 (2012) 649–656

Contents lists available at SciVerse ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Automatic food intake detection based on swallowing sounds

Oleksandr Makeyev[a,*], Paulo Lopez-Meyer[b], Stephanie Schuckers[c], Walter Besio[a], Edward Sazonov[b]

[a] Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, 4 East Alumni Ave., Kingston, RI 02881, USA
[b] Department of Electrical and Computer Engineering, University of Alabama, 801 University Boulevard, Tuscaloosa, AL 35487, USA
[c] Department of Electrical and Computer Engineering, Clarkson University, 8 Clarkson Ave., Potsdam, NY 13699, USA

ARTICLE INFO

Article history:
Received 19 December 2011
Received in revised form 11 March 2012
Accepted 17 March 2012
Available online 7 April 2012

Keywords:
Food intake
Ingestive behavior
Swallowing
Deglutition
Wearable sensors
Obesity

ABSTRACT

This paper presents a novel fully automatic food intake detection methodology, an important step toward objective monitoring of ingestive behavior. The aim of such monitoring is to improve our understanding of eating behaviors associated with obesity and eating disorders. The proposed methodology consists of two stages. First, acoustic detection of swallowing instances based on mel-scale Fourier spectrum features and classification using support vector machines is performed. Principal component analysis and a smoothing algorithm are used to improve swallowing detection accuracy. Second, the frequency of swallowing is used as a predictor for detection of food intake episodes. The proposed methodology was tested on data collected from 12 subjects with various degrees of adiposity. Average accuracies of >80% and >75% were obtained for intra-subject and inter-subject models correspondingly with a temporal resolution of 30 s. Results obtained on 44.1 h of data with a total of 7305 swallows show that detection accuracies are comparable for obese and lean subjects. They also suggest feasibility of food intake detection based on swallowing sounds and potential of the proposed methodology for automatic monitoring of ingestive behavior. Based on a wearable non-invasive acoustic sensor the proposed methodology may potentially be used in free-living conditions.

1. Introduction

This paper presents a novel fully automatic food intake detection methodology based on a wearable non-invasive swallowing sensor. Such methodology can be helpful in characterization of ingestive behaviors associated with a variety of eating disorders and for development of clinical interventions. According to World Health Organization, 1.5 billion adults, 20 and older, were overweight worldwide in 2008 [18]. Of these over 200 million men and nearly 300 million women were obese and monitoring of ingestive behavior (MIB) could potentially be used in active weight control programs providing the objective feedback needed for diet management [2,15]. The objectivity of MIB feedback is crucial as unhealthy and extreme weight-control were shown to predict outcomes related to obesity and eating disorders [10].

Most currently used self-reporting techniques demonstrate widespread underestimation of food intake [7]. Because of bias and imprecision, self-reported food intake should be interpreted with caution unless independent methods of assessing its validity are included in the experimental design [17]. Replacing paper-based reports with manually operated electronic devices to simplify tedious and error-prone logging did not improve the validity of self-reporting [19]. A potential solution is to replace or augment manual self-reporting with objective automatic sensor based monitoring where eating behavior is estimated without the individual's active participation. Such automatic sensor based monitoring has the potential to improve reporting accuracy.

Characterization of food intake behavior includes: detection of periods of food intake, differentiation of solid foods from liquids, recognition of food type, prediction of the mass of ingested food and evaluation of caloric intake [15]. In this paper we concentrate on detection of periods of food intake as the next fundamental step toward our long-term objective of creating an automatic, non-invasive and wearable MIB device suitable for use in free-living conditions [8,9,13–16]. Toward this objective we have already developed a sensor system for non-invasive monitoring of chewing and swallowing, validated its reliability based on manual scores [13,14], established a methodology for automatic detection of swallowing instances using acoustical signals [16], and developed methodologies for detection and characterization of food intake [8,15] as well as automatic identification of the number of food items in a meal [9] based on manual scores of chewing and swallowing. This paper now incorporates and expands our previous work concluding it with a demonstration of integration and validation of a fully automatic

* Corresponding author at: Department of Electrical, Computer, and Biomedical Engineering, University of Rhode Island, 4 East Alumni Ave., K111 Kelley Hall, Kingston, RI 02881, USA. Tel.: +1 315 2676016; fax: +1 401 7826422.
E-mail addresses: omakeyev@ele.uri.edu (O. Makeyev), plopezmeyer@bama.ua.edu (P. Lopez-Meyer), sschucke@clarkson.edu (S. Schuckers), besio@ele.uri.edu (W. Besio), esazonov@eng.ua.edu (E. Sazonov).

1746-8094/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2012.03.005

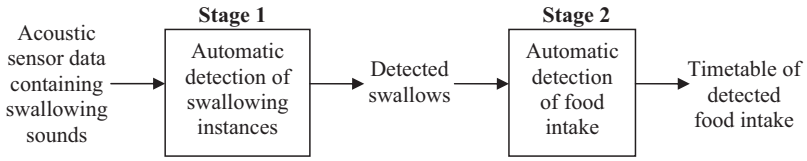

Fig. 1. Scheme of the two-stage automatic food intake detection.

food intake detection methodology. The proposed methodology consists of two stages: first, acoustic detection of swallowing instances based on mel-scale Fourier spectrum features, classification using support vector machines, principal component analysis (PCA) and a smoothing algorithm. Second, detection of periods of food intake based on frequency of swallowing. Scheme of the proposed methodology is presented in Fig. 1. To the best of our knowledge this is the first attempt of fully automatic detection of food intake based on swallowing data from wearable non-invasive sensors. A review of previous attempts to detect food intake such as intake gestures based approach proposed by Junker et al. in [6] and chewing based approach proposed by Nishimura and Kuroda in [11] is presented below. It is followed by a summary of contributions of this paper showing how it builds up on and extends the finding of our previous works integrating them into automatic food intake detection methodology. An extensive review of existing methodologies for automatic detection of swallowing instances and comparison of our methodology to related work in this area is out of scope of this paper and can be found in our previous work [16].

2. Related work

Three categories of non-invasive wearable sensors have been proposed as basis for creation of automatic food intake detection methodologies: intake gesture sensors [6], chewing sound sensors [11] and swallowing sensors. Activities that correspond to these sensor categories represent a temporal description of food consumption and thus can be used to identify periods of food intake.

Intake gestures are intentional arms and trunk movements related to food intake. In general, the task of gesture detection is difficult because relevant gestures occur sporadically in a continuous stream of data while often being embedded into other, partly arbitrary movements that are difficult to model due to their complexity and unpredictability [6]. An approach for intake gesture detection was proposed in [6] based on data from five inertial sensors attached to wrists (2), upper arms (2) and upper torso. This two-stage approach combines natural partitioning pre-selection with Hidden Markov Models classification. This approach was tested on frequently used human feeding movements of both arms and the trunk including: using fork and knife for intake of lasagna, using spoon for intake of cereals or soup, drinking from a cup, and intake of bread or chocolate bar using one hand only. The data was collected from four subjects; two sessions were performed for each subject on different days. A total of 4.7 h of data was collected with 34.7% of the data containing intake gestures. Average recall of 0.78 and average precision of 0.77 were obtained for subject-specific intra-subject prediction model. No results for non-personalized inter-subject model were reported.

The task of food intake detection based on chewing is virtually identical to the task of detection of chewing instances since the chewing sequence represents the solid food intake cycle. The chewing sequence consists of cyclic openings and closings of the jaw and arbitrary tongue movements. An approach for automatic detection of chewing instances was proposed in [11]. A wireless in-ear microphone is used to capture sound emissions generated by chewing and transmitted by bone conduction to the ear canal. A two-staged chewing detection algorithm first detects chew-like signals by applying the number of zero-crossings threshold to log energy regression coefficients. Then chewing sound verification is performed based on similarity of signals detected at the first stage to the chewing sound models derived from the training data. The authors report high average chewing detection accuracy of 98.7% for five food categories including chips, salad, rice, wafers and banana. However, limited details about the validation methods are provided including only the average number of test chews (516) and the number of training chews (100) per food category. It is not clear how many human subjects participated in the study, whether intra- or inter-subject model results are reported, how the data was divided into training and validation sets, what kind of validation technique was used, etc.

Detection of food intake based in swallowing differs from detection based on chewing since, unlike chewing, swallowing occurs sporadically and unconsciously throughout the day. Therefore, automatic detection of swallowing instances is only the first step toward food intake detection. Detected swallows need to be further classified as either spontaneous or food intake swallows. Our work on creation of automatic food intake detection based on swallowing started with the development of a non-invasive multi-modal monitoring system including a wearable acoustic swallowing sensor – a throat microphone located over the laryngopharynx [13,14]. This monitoring system comprised of hardware, software and a protocol for manual scoring of the collected data was used in a human study measuring chewing and swallowing in 21 subjects during food intake and resting periods. Next, in [15] we developed and validated methods for detection and characterization of food intake based on manual swallowing scores. In particular, we showed that the instantaneous swallowing frequency could serve as a predictor for accurate detection of food intake with an average accuracy of 87% for 30 s time windows. We also developed and validated methods for automatic detection of swallowing instances with acoustical sensors [16] yielding 84.7% average detection accuracy of swallowing events for intra-visit individual model. These results suggest the potential of using swallowing frequency for automatic detection of food intake.

However, two crucial issues remain unresolved. First, methodology of automatic swallowing detection was previously validated on intra-visit individual model only, i.e. training and validation of the algorithm were performed on different segments of the same recording rather than on separate recordings from the same or different human subjects. Such intra-visit model cannot be implemented in a MIB device as constant real-time re-training is not feasible in free-living conditions due to the absence of a gold standard manual score. Therefore, further validation is needed for intra- and inter-subject models which could be preprogrammed and implemented in such a device directly. Second, the food intake detection methodology was previously validated on gold standard manual scores only. Validation on automatically produced swallowing scores is needed to evaluate how sensitive the food intake detection algorithm is to errors in swallowing detection. The first contribution of this paper is that answers both of the aforementioned questions presenting a first fully automatic food intake detection methodology based on an acoustic swallowing sensor and validating it for intra- and inter-subject models on a database collected from 12 subjects during food intake and resting periods.

The second contribution of this paper is utilization of PCA and a smoothing algorithm to improve automatic swallowing detection accuracy for intra- and inter-subject models.

3. Methods

3.1. Human study

The automatic food intake detection methodology used for this paper was validated on a dataset that is a subset of the data collected during the human study reported in [13]. A short summary of the aspects of the original dataset that are the most relevant to the current study is presented below.

The original subject population included 21 generally healthy volunteers, 12 males and 9 females, with different degrees of adiposity. Thirty eight percent of human subjects had body mass index (BMI) greater than 30 (obese). Institutional Review Board approval was obtained for the study. Subjects read and signed the informed consent form. No subjects had dental problems that would interfere with normal food intake. Each subject participated in four separate visits scheduled for different days. Each visit consisted of three parts: (1) a 20-min resting period (10 min of silent inactivity and 10 min of talking where the subject was asked to read aloud), (2) a meal period of unlimited time to eat the meal of a fixed size, (3) a second 20-min resting period (10 min of silent inactivity and 10 min of talking). The following food items were included in the meal: a slice of cheese pizza, a can of 1% fat yogurt, an apple, and a peanut butter sandwich. The foods were selected to represent different physical properties of the food such as crispiness, softness/hardness and tackiness. The variability in physical properties of food ensured that the proposed methodology was tested on a sample that is representative of the variability in the properties of everyday food. The provided drink was clear water. All food items were to be consumed unmixed and completely. Water was consumed separately from food. Subjects ate in silence during half of the meals and were involved in a dialogue during the other half to evaluate the impact of a meal-time conversation on the accuracy of swallowing detection. Additionally, a mix of background noise was used during half of the visits to simulate realistic environments where people may be eating. Subjects were videotaped and monitored by a multimodal sensor system which included a miniature IASUS NT (IASUS Concepts Ltd.) throat microphone located over the laryngopharynx. The microphone provided a dynamic range of 46 ± 3 dB with a frequency range of 20–8000 Hz. Microphone signals were amplified by a custom-built pre-amplifier with a variable gain in the range 20–60 dB. The gain of the amplifier was set experimentally to reliably capture the subtle sounds of swallowing without saturating the amplification circuits during normal speech and fixed for the whole data collection process. 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 food intake periods and each swallowing instance. The scoring software was developed to allow assignment of corresponding labels using manual review and playback of acquired video and sensor data. To evaluate the accuracy of manual swallowing score a multi-rater reliability study was conducted with three raters on data from five out of the 21 subjects [13]. Comparing manual scores from three raters the study showed high reliability of the manual scores with average intra-class correlation (ICC) coefficient of 0.98 obtained for scores of swallowing instances. The range of the ICC is between 0 and 1 and high value of ICC means that there is little variation between the scores given by different raters. This indicates a high degree of agreement between raters suggesting that manual scores are reliable for use as a gold standard in validation of automatic swallowing and food intake detection algorithms on a large dataset. To the best of our knowledge the 65 h dataset with over 10 K swallows used in [13] is the largest dataset collected to date in a study of ingestive behavior monitoring based on data from wearable non-invasive sensors. It is also the most complex dataset with inclusion of a variety of sound artifacts and background noises of various origins, various foods and human subjects with different degrees of adiposity to create experimental conditions resembling those of free-living food consumption. Out of the 84 originally collected visits, 4 visits from one subject were used for the initial calibration of the multi-modal monitoring system and therefore discarded from further studies. Another 10 visits had partially incomplete data and were discarded from the dataset. All the cases of incomplete data can be traced to a single reason of operator's error during the data collection. The detailed description of those errors can be found in [13]. From the remaining 70 complete visits only 12 out of 20 subjects had complete datasets for all four visits. These 12 subjects comprise the dataset used to validate the methodology proposed in this paper. This derived dataset is large (44.1 h with a total of 7305 swallows) and as complex as the original dataset since it includes the same variety of data for the population with similar average degree of adiposity. Namely, the average BMI for the derived dataset is 29.2 ± 6.9 compared to 29 ± 6.4 of the original dataset. Average intra-visit swallowing detection accuracy calculated for the derived dataset is 96.7% (per-epoch) and 85.1% (per-swallow) compared to 96.8% and 84.7% respectively obtained for the original dataset. This indicates that the derived dataset is a representative subset of the original dataset.

3.2. Automatic detection of swallowing instances

The methodology for automatic detection of swallowing instances with acoustical sensors for this paper is based on the methods proposed in [16] with two major improvements: preprocessing of mel-scale Fourier spectrum (MSFS) features using principal component analysis (PCA) and postprocessing of the automatically detected swallowing instances using a smoothing algorithm. A summary of the original methodology and description of proposed improvements are presented below.

The original methodology was based on MSFS for time–frequency representation and support vector machines (SVM) for automatic detection of characteristic sounds of swallowing. First, the sound stream was split into a series of overlapping epochs and mel-scale Fourier transform was applied to each epoch. Second, the resultant epoch feature vectors were merged for a number of adjacent epochs to produce time-lagged feature vectors accounting for time-varying structure of a swallow. Assuming a feature vector f_i was formed for each epoch a time-lagged feature vector f_i' was produced by merging feature vectors of the $2 \times K + 1$ adjacent epochs: $f_i' = \{f_{i-K}, \ldots, f_i, \ldots, f_{i+K}\}$. These time-lagged vectors were used as inputs for training and validation of the SVM classifier using the Gaussian radial basis kernel function. Near-optimal values for the following parameters: epoch duration of 1.5 s, epoch step size of 0.2 s, eighth MSFS decomposition level, number of lags K equal to 1, SVM misclassification penalty parameter equal to 10 and Gaussian kernel width parameter equal to 0.05 were determined using a grid search procedure in our previous work and used in the current study [16].

Binary class labels assigned to each 1.5 s epoch for training and prediction of automatic score by SVM were produced in the following way: if any part of the epoch was scored manually as belonging to the swallow class then the epoch label was set to '1' (swallow epoch), otherwise it was set to '−1' (non-swallow epoch). These labels represent accuracy of the classifier on epoch level and do not correspond well to the accuracy of detection of swallowing events. To evaluate the classifier swallowing detection accuracy gold standard manual and automatically produced epoch

scores were used to identify the swallowing instances consisting of multiple epochs. Then the numbers of true positive (T_+), false positive (F_+), false negative (F_-), and true negative (T_-) detections were calculated. A swallow detection was considered a true positive if both manual and automatic scores contained continuous sequences of swallow epochs intersecting at one or more epochs or on the sequence boundary. These detections were later used to calculate sensitivity ($T_+/(T_+ + F_-)$), specificity ($T_-/(T_- + F_+)$), and prevalence ($(T_+ + F_-)/(T_+ + T_- + F_+ + F_-)$) to further calculate the overall accuracy ($(T_+ + T_-)/(T_+ + T_- + F_+ + F_-)$) of swallow detection as the weighted average of sensitivity and specificity, where sensitivity is weighted by prevalence and specificity is weighted by the complement of prevalence [1].

The methodology was tested on the original dataset containing 70 visits from 20 subjects. Each visit was divided into 55 equal segments with an average duration of 1 min. Three-fold cross-validation was performed with two folds (two of every three consecutive segments) used for training and one fold used for validation at each step. An average detection accuracy of 96.8% for epochs and 84.7% for swallowing instances was obtained for such intra-visit model [16]. Analysis of obtained results also revealed that detection accuracy seemed not to be dependent on subject's BMI and had substantial tolerance to sound artifacts resulting from food intake, intrinsic speech and background noise. Complete details on the methodology including justification of selection of feature extraction and classification mechanisms and assessment of near-optimal parameters can be found in [16].

As a first proposed improvement to the swallowing detection methodology PCA was used for preprocessing of MSFS features. The combination of two machine learning algorithms: supervised SVM and unsupervised PCA is widely used in biomedical engineering [4,12]. PCA is a multivariate non-parametric statistical technique that when applied to a number of possibly correlated variables reveals the internal structure of the data in a way that best explains its variance and to transform the data into a new set of orthogonal variables called principal components which are linear combinations of the original variables. The first principal component accounts for as much of the variance in the original data as possible, and each succeeding component accounts for as much of the remaining variance as possible. A detailed description of the PCA theory is out of the scope of this paper and can be found in [5]. Measuring variance along each principal component provides information on the relative importance of each component. Therefore, PCA is often used for dimensionality reduction of feature vectors with a smaller number of principal components being used compared to the original feature vector dimension. Since kernel methods like SVM are tolerant to high dimensionality of features we used PCA with the maximal number of principal components not losing any data from the original dataset.

As the second proposed improvement, a postprocessing smoothing algorithm was used to refine the automatic score produced by the SVM. Refinement of the automatic epoch score by a smoothing algorithm was performed in two steps. Two predefined thresholds whose optimal values were obtained using grid search were used. First, labels for short segments, of up to a first threshold, of a predefined number of epochs in duration that were automatically marked as '−1' (non-swallow epochs) but were surrounded by epochs marked as '1' (swallow epochs) on both sides were reset to '1'. This postprocessing step was needed to correct the situations in which a single swallow of more than two epochs may be split into several parts by a misclassified epoch. Second, labels for short segments, of up to a second threshold, of a predefined number of epochs in duration that were automatically marked as '1' but were surrounded by epochs marked as '−1' on both sides were reset to '−1'. This step eliminated the cases where accidental epochs were incorrectly classified as swallows. An illustration of the automatic score refinement with the smoothing algorithm is presented in Fig. 2.

The results of testing of both the original and improved swallowing detection techniques as a part of a food intake detection methodology in intra- and inter-subject models are presented below.

3.3. Automatic detection of food intake

Our food intake prediction approach assigns binary labels 'intake' or 'no intake' to predefined length time windows based on the average instantaneous swallowing frequency (ISF) calculated for the current window. The ISF is the inverse of the time between each two consecutive swallows and is expressed in swallows per minute: $\mathrm{ISF}_i^j = 60/(t_i - t_{i-1})$ (sw/min), where t_i is the temporal location of the swallow occurrence in seconds for swallow $i = 2, \ldots, N_j$ and N_j is the total number of swallows of the current window j. A higher ISF indicates shorter time between two consecutive swallows.

Selection of the intake detection window size defines the temporal resolution. As higher frequencies of swallowing indicate the presence of food ingestion the window should be long enough to reliably detect an increase in the number of swallows associated with food consumption compared to spontaneous swallowing, i.e. the smallest window size is limited by the minimal detectable change in swallowing frequency. At the same time the window should be short enough to detect such short food consumption events as snacking. In [15] we estimated the optimal trade-off between detection accuracy and temporal resolution to be a time window length of 30 s. The same window length was used for the current study.

Detection of food intake using swallowing frequency as a predictor was performed using a floating average prediction model in the following way: first, a decision threshold $T = \alpha \cdot (1/M)\sum_{j=1}^{M}\left((1/N_j)\sum_{i=2}^{N_j}\mathrm{ISF}_i^j\right)$ is calculated as a product of the average ISF of the training set multiplied by a scaling factor α where M is the number of time windows in the training set and $j = 1, \ldots, M$. In this way a decision threshold is a function of the average ISF. Food intake labels are determined for each time window in the training

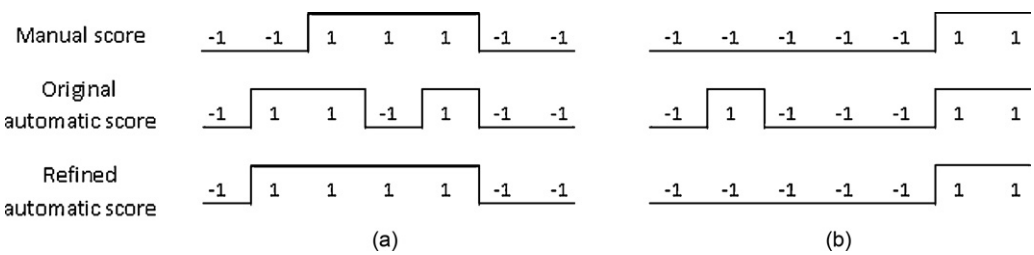

Fig. 2. Refinement of the automatic epoch score by a smoothing algorithm in two steps correcting the situations where: (a) a single swallow is split into several parts by a misclassified epoch; (b) accidental epochs are incorrectly classified as swallows.

set: if the average ISF for the current window is greater than or equal to the decision threshold then the current window is labeled 'intake':

$$L_j = \begin{cases} \text{'intake'} & \text{if } \overline{\text{ISF}}^j \geq T \\ \text{'no intake'} & \text{if } \overline{\text{ISF}}^j < T \end{cases}.$$

Training is repeated for a range of scaling factors α and the optimal scaling factor is selected based on the highest prediction accuracy achieved for the training set and further used for intake detection on the validation set. Food intake label determination for the validation set is determined similarly as on the training set using the optimal scaling factor to obtain the decision threshold T. Complete details on this approach can be found in [15] where a similar model was proposed and validated with an average accuracy of 87% obtained on the manual scores from the dataset described in [13].

3.4. Detection models

The validation of the proposed automatic food intake detection methodology based on swallowing sounds was performed for intra- and inter-subject models (as opposed to intra-visit swallowing detection model previously used in [16]). Descriptions of both models are presented below.

Individual (subject-specific) intra-subject food intake detection model was built separately for each subject so models for different subjects were independent from each other. Only data from the individual subject was included in each intra-subject model. Four-fold cross-validation was used with each of the four visits per subject used as a fold. At each cross-validation step both swallowing detection and food intake detection were performed with three visits of a subject used for training and the remaining one used for validation. Namely, each cross-validation step consisted of:

- Swallowing detection stage: acoustical sensor data and manual swallowing scores for three training visits were used to produce the automatic swallowing score for the validation visit.
- Food-intake detection stage: manual swallowing and food intake scores for three training visits were used to obtain an optimal scaling factor that was used for intake detection on the validation visit using the automatic swallowing score from the previous stage.
- Accuracy assessment: manual food intake score for the validation visit was used as the gold standard to calculate the accuracy of food intake detection.

Group (non-personalized) inter-subject food intake detection model was built for the entire population. Twelve-fold cross-validation was used with each fold representing all four visits of a certain subject. At each cross-validation step all the data from eleven subjects was used for training and all the data from the remaining one subject was used for validation. Namely, each cross-validation step consisted of:

- Swallowing detection stage: acoustical sensor data and manual swallowing scores for data from eleven subjects were used to produce the automatic swallowing scores for data from the remaining subject.
- Food-intake detection stage: manual swallowing and food intake scores for data from eleven subjects were used to obtain an optimal scaling factor that was further used for intake detection on the data from the remaining subject using the automatic swallowing scores produced at the previous stage.
- Accuracy assessment: manual food intake scores for data from the validation subject were used as the gold standard to calculate the accuracy of food intake detection for all of the subject's four visits.

4. Results

4.1. Intra-subject food intake detection model

The results of automatic detection of swallowing and food intake for intra-subject model with and without preprocessing with PCA and postprocessing with the smoothing algorithm are presented in Table 1. Distributions of per-subject swallowing and food intake detection accuracies versus the subject's BMI and corresponding linear fits of the data are presented in Fig. 3 for the case of the highest average food intake detection accuracy obtained for the intra-subject model highlighted with bold in Table 1. In Table 1 per-epoch and per-swallow swallowing detections correspond to classifier detection accuracies on epoch and swallowing instance levels respectively.

4.2. Inter-subject food intake detection model

The results of automatic detection of swallowing and food intake for inter-subject model with and without preprocessing with PCA and postprocessing with the smoothing algorithm are presented in Table 2. Distributions of per-subject swallowing and food intake detection accuracies versus the subject's BMI and corresponding linear fits of the data are presented in Fig. 4 for the case of the highest average food intake detection accuracy obtained for the inter-subject model highlighted with bold in Table 2.

Table 1
Effects of preprocessing (PCA) and postprocessing (smoothing algorithm) on average accuracy for intra-subject model.

Intra-subject model	Average accuracy [sensitivity, specificity] (%)			
	Baseline	Baseline + preprocessing	Baseline + postprocessing	Baseline + preprocessing + postprocessing
Per-epoch swallowing detection	95.1 [39.3, 98.7]	95.7 [42.6, 99.2]	95 [40.7, 98.5]	**95.7 [44, 99]**
Per-swallow swallowing detection	74.5 [66.2, 81.7]	79.2 [72.5, 84]	75.9 [64.9, 84.8]	**80.4 [71.3, 87]**
Food intake detection	76.4 [68.2, 79.2]	75.8 [71.8, 77.5]	77.5 [68.5, 80.4]	**80.3 [71.4, 84.2]**

Table 2
Effects of preprocessing (PCA) and postprocessing (smoothing algorithm) on average accuracy for inter-subject model.

Inter-subject model	Average accuracy [sensitivity, specificity] (%)			
	Baseline	Baseline + preprocessing	Baseline + postprocessing	Baseline + preprocessing + postprocessing
Per-epoch swallowing detection	91.5 [36.3, 95.2]	93.5 [25, 98]	**90.9 [39, 94.3]**	93.3 [26.5, 97.8]
Per-swallow swallowing detection	64 [62.1, 69.6]	65.3 [52.1, 72.9]	**66.4 [60, 73.1]**	66.7 [51.5, 75.6]
Food intake detection	74.2 [70.7, 75.7]	59.8 [66.1, 57.5]	**75.2 [67.9, 78.5]**	60.1 [64.7, 58.1]

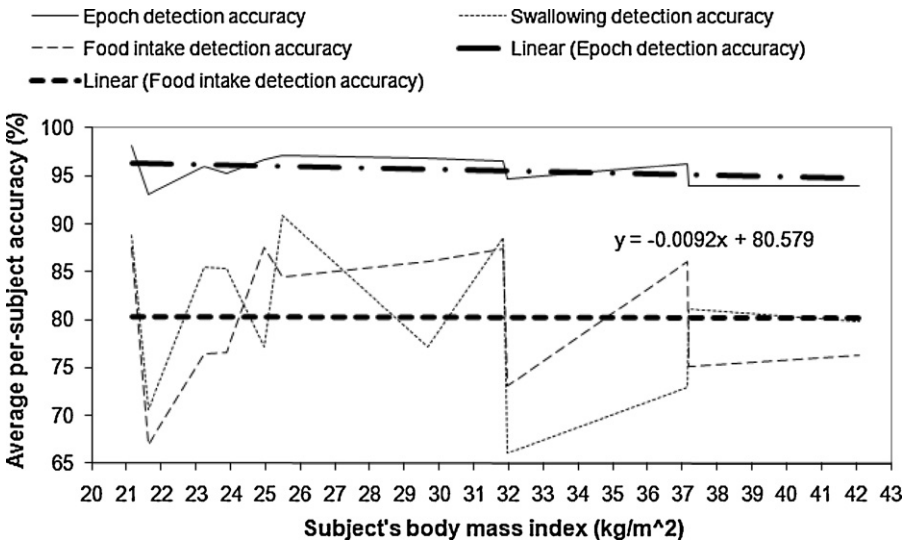

Fig. 3. Intra-subject model: distribution of accuracies for per-epoch and per-swallow swallowing detection and food intake detection versus subject's BMI with corresponding linear fits plotted for the case of the highest average food intake detection accuracy highlighted with bold in Table 1.

5. Discussion

As can be seen from Tables 1 and 2 the highest average food intake detection accuracies of 80.3% and 75.2% were obtained for intra- and inter-subject models respectively. Linear fits for distributions of per-subject food intake detection accuracy versus subject's BMI are presented in Figs. 3 and 4. While the linear fit for intra-subject model has a negative slope of less than 0.01 the one for inter-subject model has a negative slope of 0.62 with accuracy decreasing with an increase of BMI. Still, even for the volunteer with the highest BMI in this study (BMI of 42.1, severe obesity) the inter-subject food intake detection accuracy was 70.6% suggesting that the proposed food intake detection approach is suitable for monitoring of obese individuals even though more data is needed for conclusive proof.

It can be seen from Tables 1 and 2 that the effect of feature preprocessing using PCA was different for intra- and inter-subject models. For intra-subject model there was an improvement in the detection accuracy while for inter-subject model there was none. We believe that this difference may be attributed to the difference in PCA application for the two models. This difference stems from computational burden of the covariance matrix calculation needed to compute the principal components. Dimensionality of the covariance matrix is equal to the squared number of observations or, in our case, epochs in the dataset. The covariance matrix is calculated for the training set and is used to calculate the matrix of eigenvalues which is further used to project the validation data onto the new orthogonal basis. For the intra-subject model the training set is limited to the three visits from a particular subject and the associate covariance matrix can be calculated directly. For inter-subject model the training set includes a total of 44 visits from eleven subjects. With over fourteen times the calculations a complete covariance matrix is infeasible. Therefore, to apply a similar methodology in both cases, subsets are selected randomly from the training set and PCA training is performed on these subsets only. The largest feasible number of visits to use as a subset (three) was determined empirically. Representativeness of such a small subset is very limited. Allowing a better representation of the training set data in covariance matrix for inter-subject model could potentially result in increased food intake detection accuracy.

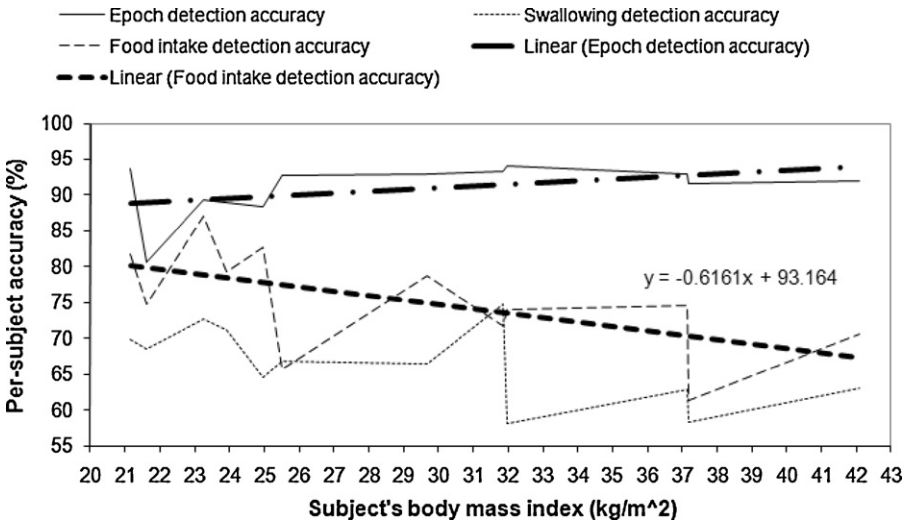

Fig. 4. Inter-subject model: distribution of accuracies for per-epoch and per-swallow swallowing detection and food intake detection versus subject's BMI with corresponding linear fits plotted for the case of the highest average food intake detection accuracy highlighted with bold in Table 2.

It can also be seen from Tables 1 and 2 that while in the case of intra-subject model application of PCA increased both sensitivity and specificity at the stage of swallowing detection, in the case of inter-subject model sensitivity decreased and specificity increased. The latter increased the existing imbalance between sensitivity and specificity for inter-subject model along with an underlying imbalance between false negative and false positive swallowing detections. Even though food intake detection may offer some tolerance to errors in automatic swallowing detection due to food intake detection being based on time intervals containing multiple swallows such tolerance depends on the balance between false negative and false positive swallowing detections partially cancelling each other. Therefore, even though the highest overall swallowing detection accuracy for inter-visit model was achieved with application of the PCA the highest overall food intake detection accuracy was obtained for the case of the best balance between sensitivity and specificity of swallowing detection. Further investigation is needed for more precise evaluation of the effect of automatic swallowing detection errors on accuracy of food intake detection.

The postprocessing smoothing algorithm improved the detection accuracy for both models as can be seen in Tables 1 and 2. The optimal thresholds of number of epochs for swallow and non-swallow gaps were equal to 5 and 0 respectively. The optimal thresholds were determined with grid search and suggest that situations where not all epochs belonging to a single swallow were classified correctly created additional false positive swallow detection errors.

Finally, from Tables 1 and 2 and Section 3.1 we can see that even with the proposed improvements the highest per-epoch and per-swallow swallowing detection accuracies for intra- and inter-subject models are lower than the ones obtained with training and validation performed on each visit separately as in [16]. Several factors could be contributing to satisfactory generalization capability of swallowing detection including intra- and inter-subject variability of swallowing sounds. Further investigation is needed to determine the sources.

While the swallowing based automatic food intake detection approach proposed in this paper cannot be directly compared with the intake gesture and chewing based approaches proposed in [6,11] respectively, we can compare the validation procedure to the one used in [6]. In [6] results comparable to the ones obtained in our study are reported for subject-specific intra-subject model validated on a dataset with a total of 784 intake gestures compared to a total of 7305 swallows in our study. No results for non-personalized inter-subject model were reported in [6] like our report. In [11] very limited detail is provided on the validation procedure limiting the interpretation of the reported high average detection accuracy. Furthermore, the proposed swallowing based approach may be less prone to some of the limitations inherent in intake gesture and chewing based approaches. Namely, limitations of intake gesture approach include: first, not all food items require intake gestures, e.g. a high-caloric milkshake can be consumed using a straw. Second, arm movements to the head that are not related to food intake, e.g. brushing teeth, smoking, etc., may result in misclassifications. Third, high intra-subject variability of intake gestures is reported in [6] caused by differences in size and consistency of food pieces and temporal aspects such as changes in food temperature and natural satiety subjects were developing during the intake sessions. Finally, even though inter-subject variability of intake gestures has not been evaluated there may be significant differences due to differences in human eating behaviors, e.g. eating with chopsticks versus cutlery. Chewing and swallowing seem to be less related to personal eating behaviors and, therefore, may offer smaller intra- and inter-subject variability for automatic detection of food intake. However, the usability of food intake detection based on chewing sensors is limited to solid foods since there is little to no chewing present during consumption of liquid and certain semisolid (yogurt, pudding, etc.) food items. This limitation makes chewing sensors more feasible for sensor fusion rather than for an independent food intake detection sensor. Furthermore, the absence of spontaneous chewing throughout the day as compared to, for example, swallowing gives no indication whether the MIB device based on the chewing sensor is being worn or not making the device vulnerable to intentional misreport of food intake. Based on these actual and potential limitations of intake gesture and chewing based approaches we can conclude that swallowing sensor may be the most promising option for creation of a single sensor MIB device even though it also has several actual and potential limitations. First potential limitation is situations where swallowing rate is elevated for other reasons than food intake, e.g. mental strain, emotional reactions, etc. For example, in a study conducted by Cuevas et al. on data from 38 generally healthy undergraduates pleasant low arousal, neutral, or aversive high arousal condition all resulted in significantly increased spontaneous swallowing rates with means of 7.9 ± 1.9 (standard error), 15.8 ± 2.4, and 23.7 ± 3.6 swallows per 30 min, respectively [3]. In this case even the high arousal condition resulted only in 23.7 ± 3.6 swallows per 30 min swallowing rate corresponding at most one swallow per minute. At the same time in our study on using swallowing frequency as predictor of food intake we have found that the Bayes optimal threshold which defined an optimal decision boundary between the classes of “food intake” and “no food intake” was equal to 4 swallows per minute [15], i.e. it was at least four times higher than spontaneous swallowing rate during high arousal emotional state. This is an indication that alteration of spontaneous swallowing frequency due to emotional state is unlikely to cause problems to our food intake detection methodology. Further investigation is needed to evaluate the impact of other potential causes of elevated swallowing rate on food intake detection accuracy. Second potential limitation of an approach based on swallowing sensor is related to differences in voluntary swallowing rates for different individuals. These differences are part of our motivation to evaluate the non-personalized inter-subject model in this work and to compare its performance to performance of the subject-specific intra-subject model. Even though the performance for the former model was worse the difference between average food intake detection accuracies for intra- and inter-subject groups (80.3% and 75.2% correspondingly) was not significant ($p = 0.093$, two-sample t-test for comparison and Ryan-Joiner test for normality of sample distributions) suggesting that aforementioned differences are not rendering the proposed approach useless. Analysis of larger subject populations for longer study durations is needed for a conclusive proof and is planned for future work on this project. Most importantly, third limitation is related to the effect of sound artifacts and background noises on accuracy of acoustical swallowing detection and therefore accuracy of food intake detection. In our previous work we have assessed this effect on intra-visit swallowing detection accuracy where the highest accuracy was observed for quiet (no talking or reading aloud) periods of no food intake (88%) and the lowest one was observed for periods of food intake combined with talking and background noise (82.9%) [16]. These results suggest that artifact sounds may negatively impact the detection accuracy. It was not feasible to perform a similar assessment in current study since cross-validation for both intra- and inter-subject models involved grouping together subject visits with and without artifacts and noise. Overall, further investigation of this effect is needed with application of noise cancellation techniques having the potential to improve the detection accuracy. Finally, further comparison between different sensor modalities needs to be drawn including aspects other than detection accuracy such as comfort, acceptance, scalability, etc.

Acknowledgment

This work was supported in part by National Institutes of Health grants R21DK085462 and R21HL083052.

References

[1] A.J. Alberg, J.W. Park, B.W. Hager, M.V. Brock, M. Diener-West, The use of overall accuracy to evaluate the validity of screening or diagnostic tests, J. Gen. Intern. Med. 19 (2004) 460–465.
[2] O. Amft, G. Tröster, On-body sensing solutions for automatic dietary monitoring, IEEE Pervas. Comput. 8 (2009) 62–70.
[3] J.L. Cuevas, E.W. Cook III, J.E. Richter, M. McCutcheon, E. Taub, Spontaneous swallowing rate and emotional state. Possible mechanism for stress-related gastrointestinal disorders, Dig. Dis. Sci. 40 (1995) 282–286.
[4] J. Jin, X. Wang, B. Wang, Classification of direction perception EEG based on PCA-SVM, in: J. Lei (Ed.), Proceedings of 3rd International Conference on Natural Computation ICNC2007, Haikou, China, 2007, pp. 116–120.
[5] I. Jolliffe, Principal Component Analysis, second ed., Springer, New York, 2002.
[6] H. Junker, O. Amft, P. Lukowicz, G. Tröster, Gesture spotting with body-worn inertial sensors to detect user activities, Pattern Recogn. 41 (2008) 2010–2024.
[7] M.B.E. Livingstone, A.E. Black, Markers of the validity of reported energy intake, J. Nutr. 133 (2003) 895–920.
[8] P. Lopez-Meyer, O. Makeyev, S. Schuckers, E. Melanson, M. Neuman, E. Sazonov, Detection of food intake from swallowing sequences by supervised and unsupervised methods, Ann. Biomed. Eng. 38 (2010) 2766–2774.
[9] P. Lopez-Meyer, S. Schuckers, O. Makeyev, J.M. Fontana, E. Sazonov, Automatic identification of the number of food items in a meal using clustering techniques based on the monitoring of swallowing and chewing, Biomed. Signal Process. Control, doi:10.1016/j.bspc.2011.11.004.
[10] D. Neumark-Sztainer, M. Wall, J. Guo, M. Story, J. Haines, M. Eisenberg, Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: how do dieters fare 5 years later? J. Am. Diet. Assoc. 106 (2006) 559–568.
[11] J. Nishimura, T. Kuroda, Eating habits monitoring using wireless wearable in-ear microphone, in: T. Stouraitis (Ed.), Proceedings of 3rd International Symposium on Wireless Pervasive Computing ISWPC2008, Santorini, Greece, 2008, pp. 130–132.
[12] G. Rong, X. Song-yun, C. Xi-na, Z. Hai-tao, Combined SVM and PCA to recognize the brain function from fMRI images, in: K.C. Chou (Ed.), Proceedings of International Conference in Bioinformatics and Biomedical Engineering ICBBE2009, Beijing, China, 2009, pp. 1–3.
[13] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. Melanson, M. Neuman, Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior, Physiol. Meas. 29 (2008) 525–541.
[14] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. Melanson, M. Neuman, Reply to 'Comment on non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior', Physiol. Meas. 30 (2009) L5–L7.
[15] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, E. Melanson, M. Neuman, J. Hill, Toward objective monitoring of ingestive behavior in free living population, Obesity 17 (2009) 1971–1975.
[16] E. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, M. Neuman, Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior, IEEE Trans. Biomed. Eng. 57 (2010) 626–633.
[17] D.A. Schoeller, Limitations in the assessment of dietary energy intake by self-report, Metab. Clin. Exp. 44 (1995) 18–22.
[18] World Health Organization, Obesity and overweight. Fact sheet No. 311. Updated March 2011. http://www.who.int/mediacentre/factsheets/fs311/en/ (accessed 19.04.11).
[19] B.A. Yon, R.K. Johnson, J. Harvey-Berino, B.C. Gold, The use of a personal digital assistant for dietary self-monitoring does not improve the validity of self-reports of energy intake, J. Am. Diet. Assoc. 106 (2006) 1256–1259.