34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

A Robust Classification Scheme for Detection of Food Intake Through Non-Invasive Monitoring of Chewing*

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

Abstract—**Automatic methods for food intake detection are needed to objectively monitor ingestive behavior of individuals in a free living environment. In this study, a pattern recognition system was developed for detection of food intake through the classification of jaw motion. A total of 7 subjects participated in laboratory experiments that involved several activities of daily living: talking, walking, reading, resting and food intake while being instrumented with a wearable jaw motion sensor. Inclusion of such activities provided a high variability to the sensor signal and thus challenged the classification task. A forward feature selection process decided on the most appropriate set of features to represent the chewing signal. Linear and RBF Support Vector Machine (SVM) classifiers were evaluated to find the most suitable classifier that can generalize the high variability of the input signal. Results showed that an average accuracy of 90.52% can be obtained using Linear SVM with a time resolution of 15 sec.**

I. Introduction

A chronic imbalance between the energy consumed in foods and the energy expended in physical activities is one of the leading causes of overweight and obesity. Among adults, the prevalence of obesity reached a total of 35.5% in 2009-2010 in United States whereas prevalence of obesity and overweight combined raised up to 68.8% [1]. Additionally, obesity in adolescence was strongly associated with the risk of developing severe obesity in adulthood [2]. These facts clearly indicate that obesity is a problem that needs to be promptly and carefully addressed.

Monitoring of Ingestive Behavior (MIB) of individuals under free living conditions is particularly important to detect and correct specific patterns of food intake that cause weight gain. Current self-reporting methods for dietary assessment such as food questionnaires [3], meal diaries, food recall [4] and multimedia diaries [5] lead to inaccurate measurements of food intake due to subjects are inclined to underreport and miscalculate food consumption [6]. Consequently, objective and more accurate methods are necessary for monitoring eating behavior.

Automatic methods of MIB based on wearable sensors have been proposed as a potential solution to replace manual self-reporting methods. Most of these systems integrate a wearable sensor for monitoring physiological changes associated with food intake and signal processing and/or pattern recognition algorithms for determining when and how food is consumed. In [7], an in-ear microphone was used to detect characteristic sounds generated during chewing and swallowing of food. Two different signal processing algorithms were developed to detect food intake activity from the acoustic signal with 83.3% detection accuracy. Another study implemented an earpad sensor to capture air-conducted vibrations originated during chewing of food [8]. Data collected from 2 subjects under varied environmental conditions was used to develop a pattern recognition system that classified intake of four different food types with 86.6% accuracy.

Our research group is working towards the development of a non-invasive wearable device for automatic and objective MIB in free-living conditions. Our approach is based on the monitoring of chewing and swallowing activities as indicators of food intake (when, how and how much food is consumed) [9], [10]. Swallowing information was used in [11] and [12] to create individual and group models to detect periods of food intake. Higher detection rates were observed for subject dependent models suggesting the need for individual calibration. Food intake detection through monitoring of chewing alone was introduced in [13]. A piezoelectric film strain sensor was used to sense characteristic jaw motions during chewing. A food intake detection accuracy of 80.98% was achieved by a group model based on Support Vector Machines (SVM) that included information from 20 subjects collected during resting, reading and eating activities. The development of a group model eliminated the need for individual calibration.

The main goal of this study was to create a more robust group model that can be suitable for food intake detection in free-living conditions. For that reason, more variability was added to the chewing signal by collecting signals during several daily living activities. Linear and RBF SVM models were implemented using time and frequency domain features extracted from the filtered signal. Experimental results showed that a robust classification model discriminated periods of food intake from no intake with an accuracy of 90.52% with 15 sec time resolution.

* The project described was supported by Grant Number R21DK085462 from the National Institute of Diabetes and Digestive and Kidney Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive and Kidney Diseases or the National Institutes of Health.

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

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

978-1-4577-1787-1/12/$26.00 ©2012 IEEE

II. METHODS

A. Data Collection

A total of 7 healthy subjects (5 males and 2 females, aged 18 to 34 years old) were recruited for this study. Subjects did not present any dental problem that would hinder normal food intake. Each subject participated in three different visits during which they performed the following activities:

a) *Talking* for 10-min, which was divided into two parts: reading aloud during the first 5-min and participating of a conversation for the next 5-min;

b) *Walking* for 10-min, where subjects walked on a hard surface (pavement) during the first 5-min and on a soft surface (grass) for the remaining 5-min;

c) *Eating*, where subjects had unlimited time to eat an entire meal of content and size selected by each participant according to their own preferences and

d) *Resting* for 10-min, which was also divided into two parts: subject sitting quietly for 5-min and browsing the internet on a laptop computer for the next 5-min.

In general, each visit resulted in approximately 50 minutes of data and in 2.5 hours of monitored activity from each subject. Around 20-30% out of the total collected data belonged to food intake whereas the remaining data belonged to activities that were labeled as "no intake". These activities were selected as they represent the most commonly observed activities in a free-living environment.

A wearable system was used to monitor the subject's activities during each experiment. This system comprised a jaw motion sensor and a self-report push button. The jaw motion sensor was a non-invasive piezoelectric film strain gauge sensor (MSI, Inc) that monitored the motion of the jaw of each subject during the experiments. Previous studies showed that the most suitable location for the sensor is the area below the outer ear where jaw motion can be easily detected by monitoring changes in the skin curvature [10], [13]. Medical adhesive was used to attach the sensor to the skin. Collected chewing signal was first buffered with unity gain using a ultra-low power operational amplifier (1 GOhm differential input resistance) and then amplified using a custom-built differential amplifier of gain 2. The resultant signal was sampled at 1000 Hz, quantized with 10 bits using a portable data logger (Logomatic V3.0, Sparkfun Electronics) and then stored into a microSD memory card.

Subjects were asked to report all instances of food intake using handheld push-button, which provided a pulse of 1.0 V and 0.5 V for solid and liquid intake respectively. A solid food intake instance consisted of three parts: bite, chewing sequence and swallows whereas a liquid food intake instance consisted of only two parts: sip and swallows (chewing is not observed during liquid consumption). Subjects were instructed to press and hold the button during each entire food intake instance. Self-report was used to compute the true food intake labels for the pattern recognition algorithm. Push-button signal was sampled at 10 Hz, quantized with 10 bits and stored into the microSD using the same data logger. Chewing and self-report signals were simultaneously acquired with no drift between signals. Top graph of Fig. 1 illustrates an example of the chewing collected during one entire visit whereas bottom plot shows the food intake labels.

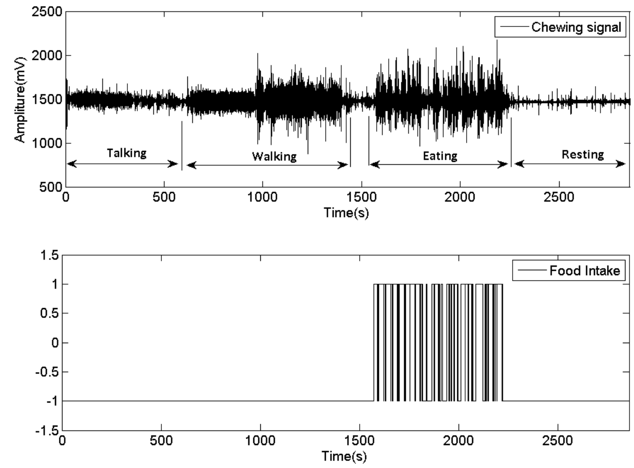

Figure 1. Top: Example of the strain sensor signal collected during one visit. Bottom: Food intake labels as reported by subject.

B. Classification scheme

The DC component of the strain sensor signal was removed and the resulting signal was normalized with respect to the median to adjust differences in signal amplitude between subjects. The processed signal was divided into non-overlapping epochs of fixed duration. Previous studies determined that an epoch of length 30 sec presented high food intake detection rate [9][13]. However, the shorter the epoch size, the better the time resolution of food intake detection. Also shorter epochs would help to detect snacking instances which are not usually self-reported. For that reason, epoch sizes of 30 sec and 15 sec were evaluated in this study in order to improve time resolution.

The classification scheme assigns a class label $C_k \epsilon$ {"no intake"; "intake"} to an epoch by classifying the state of chewing in the signal as "chewing" or "no chewing". As a result, each epoch e_i was associated with a label $t_i \epsilon \{-1, 1\}$, where $t_i = -1$ and $t_i = 1$ represented a "no intake" and a "intake" epoch, respectively. The labels for each epoch were derived from the self-report signal. An epoch was labeled as $t_i = 1$ if the self-report signal indicated food intake for at least 50% of the total epoch duration, and it was labeled $t_i = -1$ otherwise.

Features were extracted from each epoch in order to train the classifier. Our previous study analyzed spectral differences between "chewing" and "no chewing" signals indicating a pronounced difference in the range of 1.25-2.5Hz [13]. In this study, a higher sampling frequency allowed an analysis at higher frequency differences. Fig. 2 shows the averaged power spectra of the talking period (left) and of the chewing period (right). In this case, the frequency interval 100-200 Hz presented the largest difference between talking and chewing suggesting that features from that range could be used to improve classification performance. The reason for this difference is that the sensor is capturing the fundamental frequency of voice which could be found in the 85-180Hz range for adult men and in the 165-255Hz range for adult female.

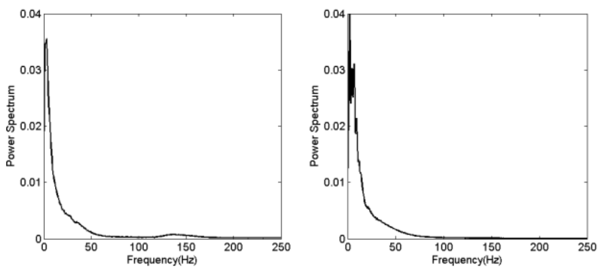

Figure 1. Average power spectra of talking (left) and chewing (rigth) shows a difference at 100-200Hz range due to fundamental frequency of voice was captured by the jaw motion sensor.

To compute the features, each epoch was filtered using a band-pass filter with 1.25-2.5 Hz resulting in a filtered epoch ef1. The original epoch e_{nf} was also band-pass filtered with 100-250 Hz cutoff frequencies to give a second filtered epoch e_{f2}. A set of 29 scalar features was extracted from each filtered epoch e_{f1} and e_{f2} and from the non-filtered epoch e_{nf}. The set of 29 features included time and frequency domain features as presented in [13]. The final feature vector combined information from e_{f1}, e_{f2} and e_{nf} in linear and logarithmic scale. To account for the time-varying structure of the chewing process, features from one neighboring epoch before and one neighboring epoch after were concatenated to the original epoch feature vector to create a final feature vector $F_i \in \Re^{2088}$.

A forward feature selection process [14] was implemented to extract a subset with the most relevant features based on classification accuracy. In this iterative procedure, the feature presenting the highest classification accuracy was added to the subset on each iteration. The algorithm stopped after no accuracy improvement was observed.

A group model incorporating chewing information from all subjects to account for the inter-subject variability was proposed to detect periods of food intake. Support Vector Machines (SVM) is a well-known supervised machine learning technique having two valuable properties: robustness and high generalization [15]. These properties become very important when training group models due to the training set encompasses high variability. In this study, SVM was the algorithm selected for chewing classification. LibSVM software package was used to implement the classifiers [16]. Linear and Radial Basis Function (RBF) were evaluated as potential kernel functions of the SVM classifiers. For the Linear kernel the penalty parameter C was optimized through a grid search procedure varying C as $C = 10^n$ for $n \in (-2, -1, 0, 1)$. For the RBF kernel, the parameters C and gamma value γ were also optimized through a grid search procedure varying C as in the Linear case and γ as $\gamma = 2^n$ for $n \in (-4, -3, \ldots, 0, 1, \ldots, 3)$.

Classification accuracy was the metric used to compare the ability of the classifiers to discriminate between "food intake" and "no intake". Accuracy was defined as the average between precision and recall:

$$\text{Accuracy} = \frac{(\text{Precision} + \text{Recall})}{2} = \frac{1}{2}\left(\frac{T_+}{T_+ + F_+} + \frac{T_+}{T_+ + F_-}\right) \quad (1)$$

True positive (T_+) was the number of correctly classified "intake" epochs, false negative (F_-) was the number of times that the model failed to classify "intake" and false positive (F_+) was the number of times the model incorrectly classified an epoch as "intake".

A 7-fold cross-validation was implemented to train and validate Linear-SVM and RBF-SVM models. This procedure allowed to train the model with data from 6 subjects and validate the model with data from the remaining subject. The classification accuracy was calculated as the average accuracy across all subjects.

III. Results

A total of 15 visits from 7 subjects were used to create Linear-SVM and RBF-SVM group models. Features from the 1.25-2.5 Hz and 100-250 Hz frequency ranges were combined into the training dataset. Results from 7-fold cross validation indicated that groups models achieved average accuracy values greater than 87% as illustrated in Table I.

TABLE I - Classification results for Linear-SVM and RBF-SVM for two different epoch sizes: 15 seconds and 30 seconds

	Linear SVM	
	Epoch size = 15 sec	***Epoch size = 30 sec***
Presicion	87.98% (± 11.85%)	88.04% (± 11.11%)
Recall	93.06% (± 5.37%)	87.56% (± 17.57%)
Accuracy	**90.52% (± 5.11%)**	**87.80% (± 11.57%)**

	RBF SVM	
	Epoch size = 15 sec	***Epoch size = 30 sec***
Presicion	85.85% (± 10.10%)	81.42% (± 12.62%)
Recall	92.05% (± 6.26%)	94.62% (± 4.06%)
Accuracy	**88.95% (± 4.66%)**	**88.02% (± 6.31%)**

Linear-SVM presented higher accuracy values than RBF-SVM although the differences were not statistically significant. The best performance was achieved by Linear-SVM with 90.52% average accuracy using an epoch size of 15 sec. When the epoch size was increased up to 30 sec the accuracy of Linear-SVM dropped to 87.80%. Feature selection procedure for Linear SVM yielded a total of 17 relevant features to describe epochs of 15 sec and a total of 6 features to describe epochs of 30 sec length.

RBF-SVM classifier showed less that 1% difference in the accuracy between epoch sizes. For 15 s epochs, RBF-SVM achieved 88.95% using a total of 10 features to represent signal epochs whereas for 30 s epochs, RBF-SVM achieved 88.02% accuracy with a feature vector having 6 elements.

IV. Discussion

A robust classification scheme was presented in this study to discriminate periods of food intake from no intake by classifying the state of the chewing signal. Implementation of Linear and RBF SVM group models achieved classification accuracies higher than 87%, with either 15 s and 30 s time resolution. Data used to train such models included high inter and intra-subject variability as chewing signals were collected from 7 participants when they performed several daily living activities: walking, talking, reading, internet navigation, food ingestion and resting. The main advantage of group models based on SVM was that suitable classification performances were achieved requiring no individual calibration. The best performance was achieved by a Linear SVM model with 90.52% (± 5.11%) accuracy with 15s time resolution. The reported accuracy value included solid and liquid intake detection which is also an advantage of the presented classification scheme.

When comparing with results from previous studies, the system improved in detection performance and in time resolution. Our previous study reported an average accuracy of 80.98% for food intake detection using chewing signals segmented into 30 sec epochs [13]. The Linear-SVM model proposed in this study showed a significant increment in the average accuracy up to a value of 90.52%. This result was achieved with an epoch sizes of 15 s which would potentially allow the detection of short periods of food intake, such as snacking. Other approaches also used chewing monitoring to detect periods of food intake [7], [8]. Models created in those studies achieved accuracies ranging from 83-87% but without considering variability caused by daily living activities.

The variability introduced into the training data by including several daily living activities may significantly affect the ability of the model to detect food intake. The number of false positive instances (reflected in Precision values) was an indication of whether the model was able to generalize for such variability. Higher Precision values were observed for Linear-SVM than for RBF-SVM models suggesting that the former models may be more suitable for food intake detection.

Result from the feature selection procedure confirmed that the frequency ranges 1.25-2.5 Hz and 100-250 Hz provided critical information for food intake detection. In all cases the subsets with the most relevant features comprised features from filtered (e_{f1}, e_{f2}) and unfiltered (e_{nf}) epochs. In addition, more information was required to classify chewing instances when using 15 s epochs since a higher number of features were selected to represent 15 s epochs than 30 s epoch in both Linear-SVM models (17 vs. 6 features) and RBF SVM (10 vs. 6 features) models.

The non-intrusive characteristic of chewing sensor plus the robust classification scheme presented in this study are suitable to be integrated into an automatic ingestion monitoring system for long-term applications (24hs) under free-living conditions. Additionally, the implementation of a group model would allow detection of food intake without requiring calibration for each individual.

References

[1] K. M. Flegal, M. D. Carroll, B. K. Kit, and C. L. Ogden, “Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999-2010,” *JAMA*, vol. 307, no. 5, pp. 491–497, Feb. 2012.

[2] N. S. The, C. Suchindran, K. E. North, B. M. Popkin, and P. Gordon-Larsen, “Association of Adolescent Obesity With Risk of Severe Obesity in Adulthood,” *JAMA*, vol. 304, no. 18, pp. 2042–2047, Nov. 2010.

[3] S. A. Bingham, C. Gill, A. Welch, K. Day, A. Cassidy, K. T. Khaw, M. J. Sneyd, T. J. A. Key, L. Roe, and N. E. Day, “Comparison of dietary assessment methods in nutritional epidemiology: Weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records,” *British Journal of Nutrition*, vol. 72, no. 4, pp. 619–643, 1994.

[4] C. A. Vereecken, M. Covents, W. Sichert-Hellert, J. M. F. Alvira, C. Le Donne, S. De Henauw, T. De Vriendt, M. K. Phillipp, L. Béghin, Y. Manios, L. Hallström, E. Poortvliet, C. Matthys, M. Plada, E. Nagy, and L. A. Moreno, “Development and evaluation of a self-administered computerized 24-h dietary recall method for adolescents in Europe,” *International Journal of Obesity*, vol. 32, no. SUPPL. 5, p. S26–S34, 2008.

[5] D.-H. Wang, M. Kogashiwa, and S. Kira, “Development of a new instrument for evaluating individuals’ dietary intakes,” *J Am Diet Assoc*, vol. 106, no. 10, pp. 1588–1593, Oct. 2006.

[6] M. B. E. Livingstone and A. E. Black, “Markers of the validity of reported energy intake,” *J. Nutr*, vol. 133 Suppl 3, p. 895S–920S, Mar. 2003.

[7] S. Passler and W.-J. Fischer, “Food Intake Activity Detection Using a Wearable Microphone System,” in *2011 7th International Conference on Intelligent Environments (IE)*, 2011, pp. 298–301.

[8] O. Amft, “A wearable earpad sensor for chewing monitoring,” in *Sensors, 2010 IEEE*, 2010, pp. 222–227.

[9] E. S. Sazonov, S. A. C. Schuckers, P. Lopez-Meyer, O. Makeyev, E. L. Melanson, M. R. Neuman, and J. O. Hill, “Toward Objective Monitoring of Ingestive Behavior in Free-living Population,” *Obesity*, vol. 17, no. 10, pp. 1971–1975, May 2009.

[10] 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,” *Physiological Measurement*, vol. 29, no. 5, pp. 525–541, 2008.

[11] P. Lopez-Meyer, O. Makeyev, S. Schuckers, E. Melanson, M. Neuman, and E. Sazonov, “Detection of Food Intake from Swallowing Sequences by Supervised and Unsupervised Methods,” *Annals of Biomedical Engineering*, vol. 38, no. 8, pp. 2766–2774, 2010.

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

[13] E. Sazonov and J. Fontana, “A Sensor System for Automatic Detection of Food Intake Through Non-Invasive Monitoring of Chewing,” *IEEE Sensors Journal, accepted for publication*.

[14] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997.

[15] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16] C. Chih-Chung and L. Chih-Jen, “LIBSVM: a library for support vector machines.” [Online] http://www.csie.ntu.edu.tw/~cjlin/libsvm.