

Automatic Ingestion Monitor: A Novel Wearable Device for Monitoring of Ingestive Behavior

Juan M. Fontana, *Member, IEEE*, Muhammad Farooq, and Edward Sazonov, *Senior Member, IEEE*

Abstract—Objective monitoring of food intake and ingestive behavior in a free-living environment remains an open problem that has significant implications in study and treatment of obesity and eating disorders. In this paper, a novel wearable sensor system (Automatic Ingestion Monitor, AIM) is presented for objective monitoring of ingestive behavior in free living. The proposed device integrates three sensor modalities that wirelessly interface to a smart phone: a jaw motion sensor, a hand gesture sensor and an accelerometer. A novel sensor fusion and pattern recognition method was developed for subject-independent food intake recognition. The device and the methodology were validated with data collected from 12 subjects wearing AIM during the course of 24 hours in which both the daily activities and the food intake of the subjects were not restricted in any way. Results showed that the system was able to detect food intake with an average accuracy of 89.8%, which suggests that AIM can potentially be used as an instrument to monitor ingestive behavior in free living individuals.

Index Terms—Automatic ingestion monitor (AIM), chewing, eating disorders, food intake detection, obesity, pattern recognition, wearable sensors.

I. INTRODUCTION

EFFECTIVE interventions are required to reduce the incidence of obesity and eating disorders and their life-threatening complications. Obesity is a major health problem that affects not only adult population but also adolescents and children. A reduction in life expectancy of individuals with severe obesity is plausible [1]. In the United States, the prevalence of obesity reached a total of 35.5% among adults and 16.9% among adolescents in 2009-2010 [2]. On the other hand, eating disorders are serious mental disorders that cause disturbances on eating habits or weight-control behavior of individuals [3]. Anorexia nervosa, bulimia nervosa and binge

eating are the most common eating disorders with lifetime prevalence ranging from 0.6 to 4.5% in the United States [4]. Both obesity and eating disorders are medical conditions highly resistant to treatment and can have severe physical and physiological health consequences [5]. Thus, the implementation of accurate methods for Monitoring of Ingestive Behavior (MIB) is critical to provide a suitable assessment of intake particularly in individuals who would most benefit from professional help.

Current methodologies used to understand and analyze food intake patterns associated with obesity and eating disorders largely rely on laboratory studies and on self-report rather than on objective observations [6], [7]. The doubly-labeled water [8] is the most precise method to measure energy intake over a period of several days, however it is not capable of identifying individual eating episodes. Other methods for MIB such as food frequency questionnaires and diet diaries are inaccurate due to subjects tending to underreport and miscalculate food consumption [9]. Thus, new approaches for objective and accurate assessment of free-living food intake patterns in humans are necessary for monitoring of eating behavior [10].

Recent advances in the area of food intake monitoring focused on the development of systems that attempt to address research and clinical needs by replacing manual self-reporting methods [11]–[19]. Several authors explored the use of chewing sounds captured through an in-ear microphone to detect and characterize food intake activity [14], [16]. Specialized algorithms were developed to process the acoustic signal achieving acceptable results for single meal experiments in laboratory settings, where the number of food types consumed was restricted. The recognition of gestures using wearable sensors was also proposed for MIB [12], [18]. Recently, a watch-like device incorporating a miniature gyroscope was developed for measuring intake by an automatic tracking of wrist motions during hand-to-mouth gestures or “bites” [18]. This device showed high sensitivity for “bite” counting but may carry the limitations of self-reported food intake as subjects need to turn it on and off at every meal to avoid spontaneous hand gestures registering as “bites”. A novel wearable sensor platform was recently presented [19]. It consisted of a microphone and a camera for detecting and characterizing food intake with high recognition rate. Although these technologies presented satisfactory performances in the laboratory, the accuracy of the methodologies for detecting unrestricted food intake in free-

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

Juan M. Fontana, Muhammad Farooq and Edward Sazonov (esazonov@bama.ua.edu phone: 205-348-1981) are with the Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, AL 35487 USA. Juan M. Fontana is also with the Facultad de Ingenieria, Universidad Nacional de Rio Cuarto, Rio Cuarto, Cordoba, Argentina.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

living environments remains to be tested.

Our research group is working towards the development of a non-invasive wearable device for automatic and objective MIB under free-living conditions. Our approach proposes the use of hand-to-mouth gestures, chews and swallows as indicators of food intake (objectively detecting timing and duration of each instance of food intake, number of bites and chews in each eating episode, characterizing eating frequency and mass of ingestion) [11], [13]. Monitoring of swallowing activities by acoustical means was integrated with machine learning algorithms in [20] and [21] to create classification models, which detected periods of food intake with more than 85% accuracy. The drawback of this methodology is the need for individual calibration due to an apparent uniqueness of swallowing sounds for each individual. Monitoring of chewing activities by sensing characteristic jaw motions was integrated with pattern recognition algorithms in [17] and [22] to create group models that achieved >80% accuracy for food intake detection in laboratory settings. The implementation of such group models eliminated the need for individual calibration.

In this paper we present the design and validation of a novel wearable sensor system (Automatic Ingestion Monitor, AIM) for objective detection of food intake in free-living. To the best of our knowledge, AIM is the first wearable sensor system that is capable of objective 24-hr monitoring of ingestive behavior without relying on any input or self-report from the subject, only compliance with wearing of the device. AIM wirelessly integrates three different sensors with a smart phone: a jaw motion sensor to monitor chewing, a hand gesture sensor to monitor hand-to-mouth (HtM) gestures and an accelerometer to monitor body motion. A novel approach to sensor information fusion and pattern recognition based on Artificial Neural Networks (ANN) is used for robust and accurate detection of food intake in free living conditions that presents a substantially more challenging environment than the laboratory. The device and methodology were validated with data collected from 12 subjects who wore AIM in free living during 24 hours without any restrictions on their food intake and daily activities (except showering).

II. AUTOMATIC INGESTION MONITOR (AIM)

A. Device description

The wearable sensor system of AIM consists of 4 key parts: a jaw motion sensor; a wireless module; an RF transmitter of the hand-to-mouth sensor; and an Android smartphone. Fig.1 right shows a picture of a subject wearing AIM.

The jaw motion sensor (Fig.1. left-a) was attached by medical adhesive below the earlobe and used to capture characteristic motion of the jaw during food intake [17]. The sensor was the LDT0-028K piezoelectric film element which was interfaced to the microcontroller through a buffering, level shifting and differential amplifying op-amp circuit.

The wireless module (Fig.1. left-b) was worn on a lanyard around the neck. It contained a custom-built electronic circuit powered by a Li-Polymer 3.7V 680mAh battery (Tenenergy). This circuit incorporated: 1) MSP430F2417 processor with an 8-channel 12-bit ADC used to sample analog sensor signals; 2)

RN-42 Bluetooth module with serial port profile; 3) M25P64 64Mbit serial flash memory for data buffering in situations where wireless Bluetooth connection is temporarily unavailable; 4) preamplifier for jaw motion sensor (sampled at 1 kHz); 5) RF receiver for HtM gesture sensor (sampled at 10 Hz) operating in RFID frequency band of 125 kHz; 6) ADXL335 low power 3-axis accelerometer for capturing body acceleration (sampled at 100 Hz); and 7) a self-report push button (sampled at 10 Hz) that was used in this study for pattern recognition algorithm development and validation and use of which will not be required in the future. The sensor signals were delivered in near real time via Bluetooth to an Android smart phone that acted as a data logger.

RF transmitter module was worn on the inner side of the dominant arm at the wrist (Fig.1. left-c). It interacted with the RF receiver in the wireless module to implement a proximity sensor for detection of characteristic HtM gestures during food intake [23]. The response of the hand gesture sensor was within a range of 0-20 cm, saturating at its maximum amplitude from 0-10 cm and reaching its minimal value at 20 cm.

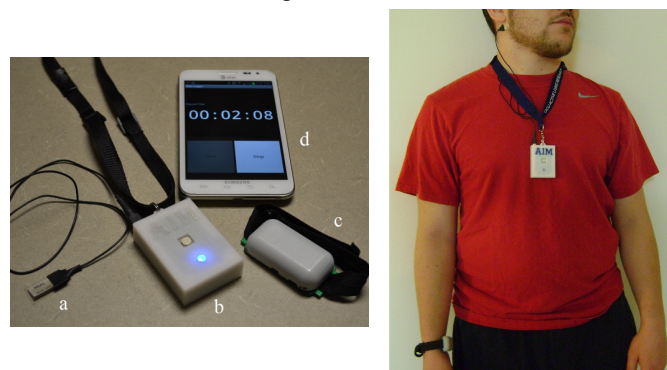


Fig. 1. Left: Wearable sensor system: a) jaw motion sensor, b) wireless module, c) RF transmitter, and d) Smart phone. Right: Subject wearing AIM.

B. Data Collection and Signal Preprocessing

A total of 12 subjects (6 male, 6 female) were recruited to participate in this study. The average age was 26.7 y (SD \pm 3.7) and the average body mass index (BMI) was 24.39 kg/m² (SD \pm 3.81). Subjects did not present any medical condition that affected normal food intake. This study was approved by the Internal Review Board at The University of Alabama and subjects read and signed an informed consent document.

Subjects were asked to wear AIM in free living for a period of 24 hrs, usually starting in the morning before breakfast and finishing in the morning of the following day. The experiments were initiated in the laboratory where a member of the research team helped subjects to put on the AIM device. The lanyard was then adjusted to the proper length making sure that subjects were comfortable while the AIM was in a good position for capturing hand gestures. An Android smart phone that included a data logger application was provided to subjects and they were asked to keep the phone with them at all times to ensure proper data transmission. Subjects were then dismissed from the lab and asked to continue with their regular activities of daily living (except showering/swimming and other activities including water immersion) without restrictions. During the experiments, subjects accomplished

ad-libitum intake meaning that they were able to eat any kind of food at any time of the day according to their own preferences and without any restrictions on the content and size of each eating episode. Subjects were asked to come back to the lab after 24 hours of data collection.

The push button included in AIM was used as the primary method for self-reporting food intake. Subjects marked each eating episode by pressing and holding the button with their non-dominant hand (not equipped with the proximity sensor) during the chewing process. As a secondary method of self-report, subjects completed a food journal indicating what type of foods they ate during the experiment as well as the start and end time of each eating period. The push button signal was used as the gold standard for identifying, timing, and marking food intake events in the sensor signals. The food journal was used to corroborate button data as well as to identify and remove accidental presses of the button. The annotated sensor data was used for developing the food intake detection algorithm.

Selecting 24 hours as the recording period for this study was based on two main reasons. First, it was long enough to cover a full cycle of daily meals and activities, including potential night eating [24]. Second, it was short enough to avoid introducing a significant error related to self-reporting intake over several days [25]. Also, the goal was to demonstrate recognition of food intake in realistic conditions of daily living without the high costs related to a longer observational study.

Fig. 2 shows the raw signals collected for one subject during a 24-hr period: the jaw motion signal (top graph), the hand gesture signal (2nd graph), the accelerometer signal for the y-axis (3rd graph) and the self-reported food intake (bottom graph). The self-report indicates the consumption of 4 major meals: breakfast (at the beginning of the experiment), lunch (at ~3.5 hrs after beginning), dinner (at ~13 hrs) and another breakfast before the end of the experiment (at ~23 hrs). Snacking was also reported to occur at approximately 2 hrs, 6.5 hrs and 12.5 hrs after the beginning of the experiment. The hand gesture signal shows an increment in the hand gesture activity during food intake intervals due to the presence of a high number of HtM gestures in the process of bringing the food to the mouth. Also, it is possible to visually identify the interval where the subject was sleeping (i.e. from 16 hrs to 22 hrs) by looking at the three sensor signals together. The jaw motion, hand gesture and body acceleration signals present a very low standard deviation during sleeping compared to the rest of the activities due to the subject resting quietly with minimal activity.

AIM captured and stored the jaw motion signal $JM(t)$, the accelerometer signals for each axis $ACC_x(t)$, $ACC_y(t)$, and $ACC_z(t)$, the hand gesture signal $HG(t)$ and the push button signal $PB(t)$. All sensor signals were time-synchronized and had exactly the same time duration. $JM(t)$ and $ACC(t)$ were high-pass filtered to remove the DC component (0.1 Hz cutoff frequency). These signals were then normalized to compensate for variations in the signal amplitude between subjects. $HG(t)$ was normalized to provide a value of “1” at saturation. HtM gestures outside the 0.25-7.5 s range were removed as most of the hand gestures related to food intake were shorter than 7.5 s

and having hand gestures shorter than 0.25 s is highly unlikely and it most probably is due to artifacts in the signal.

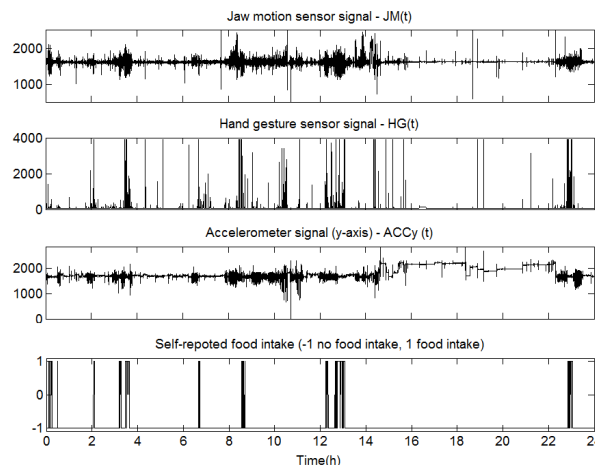


Fig. 2. Example of the signals collected by AIM in a 24-hr experiment.

III. SENSOR FUSION AND PATTERN RECOGNITION METHODOLOGY

Monitoring of food intake under free-living conditions over an extended period of time generates more complex datasets than monitoring in laboratory. The intra- and inter-subject variability of the sensor signals increases due to the presence of artifacts in the signals caused by activities from real life situations that are not possible to reproduce in a laboratory. The intra-subject variability was stipulated by the different activities performed by a subject during the 24-hr period (i.e. walking, talking, eating, sleeping, etc.) and the inter-subject variability was given by subjects having different eating patterns and lifestyles. This makes it difficult to apply models created with laboratory data to free-living data and maintain an acceptable performance. For example, a group (population-based) model created in our previous study with a database encompassing chewing information from 20 subjects achieved an accuracy of 81% under laboratory settings [17]. When the same model was applied to the free-living data collected in the present study, the accuracy decreased to 62%. This is a clear indication that lab experiments may not provide data that are representative of food intake in the community and food detection models should be created using innovative methodologies based on free living data. For that reason, a novel sensor fusion and pattern recognition approach was implemented in this work. This approach was divided into three steps, which are explained in detail in this section: a) the Sensor Fusion steps for reducing the size of the original dataset and eliminating the outliers, b) the Feature Extraction step for extracting features from the reduced dataset, and c) the Classification step for training a classifier to detect food intake episodes.

A. Sensor Fusion

The data collected after 24 hrs contained only about 3% of data related to food intake, which generated a highly unbalanced dataset. The sensor fusion approach combined information from the sensor signals to identify and remove periods of 'no food

intake' within the signals for balancing the dataset and removal of the signal artifacts and outliers. It involved two major steps. First, the product between the absolute values of $JM(t)$ and $HG(t)$ was computed as:

$$SF_1(t) = |JM(t)| \cdot |HG(t)| \quad (1)$$

$SF_1(t)$ was divided into non-overlapping epochs e_i of 30 s duration with $i = 1, 2, \dots, M_S$ total number of epochs for each subject S . The size selected for the epoch was found to present the best trade-off between the frequency of physiological events such as bites, chewing and swallowing and time resolution of food intake monitoring [11], [13]. The Mean Absolute Value (MAV) of the signal $SF_1(t)$ within e_i was computed as:

$$MAV_{e_i} = \frac{1}{N} \sum_{k=1}^N |x_k| \quad (2)$$

where x_k is the k -th sample in an epoch e_i of $SF_1(t)$ containing a total of N samples. The self-report signal, $PB(t)$, was also divided into 30 s epochs and used to assign a class label $c_i \in \{\text{'food intake' (FI), 'no food intake' (NFI)}\}$ to each e_i . An epoch was labeled as FI if at least 10s of self-report within the i -th epoch was marked as food intake; otherwise it was labeled as NFI. The 10s was chosen based on an estimation of the shortest duration of the physiological sequence that generates intake: bite, chews and swallows [26]. $SF_1(t)$ epochs would have higher MAV during food intake due to the presence of HtM gestures (associated with bites and use of napkins) and jaw motion activity (chewing) during eating. For that reason, a threshold level T_l was set to remove epochs in $SF_1(t)$ belonging to activities that did not present a combination of jaw motion and hand gestures (i.e. sleeping, sitting quietly, working on a computer, watching TV, etc.). Fig. 3 illustrates the cumulative distribution function (CDF) of the MAV for FI and NFI epochs in $SF_1(t)$ for one subject. The CDF represents the probability that an epoch will have a MAV less than or equal to a certain number in the x-axis. The CDF for NFI epochs grows faster than the CDF for FI epochs (Fig. 3), meaning that there is a high probability to find a NFI epoch with low MAV but a low probability to find a FI epoch with the same MAV and vice versa. Consequently, a common threshold value, T_l , was determined for all subjects and the indexes of the i -th epochs having a MAV below T_l were stored in a vector Idx_{SF_1} for further processing.

In the second step, the mean of the acceleration signals was computed as:

$$SF_2(t) = 1/3 \cdot (|ACC_X(t)| + |ACC_Y(t)| + |ACC_Z(t)|) \quad (3)$$

$SF_2(t)$ was divided into M_S non-overlapping epochs of 30 s duration and a class label c_i was assigned to each epoch e_i as in the first step. The MAV of $SF_2(t)$ within each epoch was calculated as in (2). Since most of the individuals consumed foods in a sedentary position, it was reasonable to anticipate that $SF_2(t)$ epochs would have higher MAV during activities involving body acceleration (i.e. walking, running, etc.) than during food intake. An example of this rationale is presented in Fig. 4, which shows a clear difference in the amplitude of $SF_2(t)$ (middle) during eating and during walking. This difference is not that clear in the jaw motion signal (top). Thus,

a common threshold value T_2 was set for all subjects and the indexes of the i -th epochs in $SF_2(t)$ with a MAV above T_2 were stored in a vector Idx_{SF_2} for further processing.

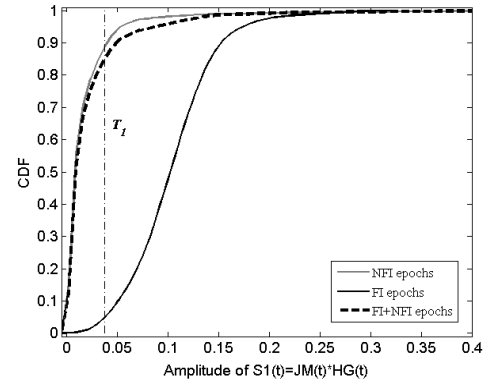


Fig. 3. Cumulative distribution function (CDF) of the MAV for food intake (FI), no food intake (NFI) epochs in $SF_1(t)$. Dashed line represents the CDF for all epochs combined.

The two steps of the Sensor Fusion algorithm were performed independently, so the epoch indexes in Idx_{SF_1} and Idx_{SF_2} were grouped into a new vector $Idx_{SF} = \{Idx_{SF_1} \cup Idx_{SF_2}\} \in \mathcal{R}^{D_S}$ with $D_S < M_S$ total number of epochs for each subject S . Finally, the signals $JM(t)$, $HG(t)$, $ACC_X(t)$, $ACC_Y(t)$, $ACC_Z(t)$, and $PB(t)$ for each subject were divided into M_S non-overlapping epochs of 30 s duration, which were synchronized in time with $SF_1(t)$ and $SF_2(t)$ epochs. Thus, the epoch indexes stored in Idx_{SF} were used to label the sensor signals epochs as NFI and remove them from the dataset used in the pattern recognition task. As a result, a total of D_S epochs were removed from the initial M_S epochs for each subject S . Implementation of this procedure allowed obtaining a more balanced dataset with approximately 36% of epochs labeled as food intake.

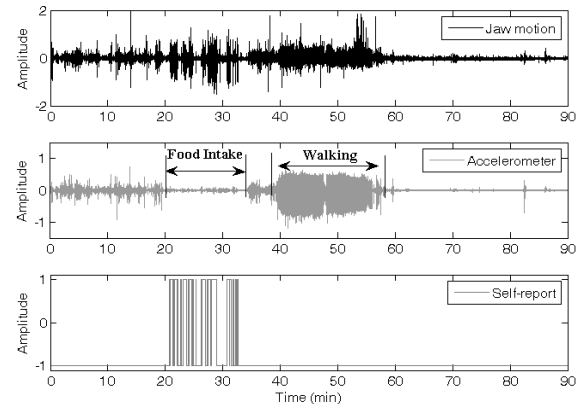


Fig. 4. Food intake and walking intervals captured by the jaw motion and accelerometer sensors. Reduced body acceleration is seen during food intake.

B. Feature Extraction

Time and frequency domain features were extracted from the remaining epochs of the sensor signals and combined to create a feature vector $f_i \in \mathcal{R}^{68}$ that represented a 30 s interval. Each vector f_i was formed by combining features from sensor signals as: $f_i = \{f_{JM}, f_{HG}, f_{ACC}\}$, where $f_{JM} \in \mathcal{R}^{38}$, $f_{HG} \in \mathcal{R}^9$, and $f_{ACC} \in \mathcal{R}^{21}$ represented the subsets of features extracted from

$JM(t)$, $HG(t)$, and the accelerometer signals respectively [27].

The subset f_{JM} included time and frequency domain features extracted from each epoch of the jaw motion signal (Table I). Frequency domain features were computed from different ranges of the frequency spectrum of $JM(t)$ within each epoch. Previous studies determined that features in the 1.25-2.5 Hz range contained information about chewing, which was successfully used to discriminate between FI and NFI epochs [17]. Also, the frequency spectrum in the 100-300 Hz range presented valuable information to identify talking intervals due to fundamental frequencies of voice for adults were found on that range [22]. Finally, features in the 2.5-10 Hz frequency range were also included in f_{JM} as they may contain important information related to other activities (i.e. walking) that could help to discriminate between intake and no intake.

The subset f_{HG} included time domain features extracted from the HtM gestures observed within each epoch (Table II). HtM gestures were detected when the amplitude of $HG(t)$ exceeded a predefined threshold value above the electronic noise.

The subset f_{ACC} contained time domain features computed from the accelerometer signals from each axis (Table III). Features included MAV, SD and the median value of the signal as well as number of zero crossings, mean time between crossings and entropy of the signal within the epoch. The means of the MAV, SD and entropy across the 3 axes were computed to obtain a total 21 features.

TABLE I. FEATURES EXTRACTED FROM THE JAW MOTION SIGNAL

#	Description	#	Description
1	Mean Absolute Value (MAV)	20	Energy spectrum in chewing range ² (chew_ene)
2	Root Mean Square (RMS)	21	Entropy of spectrum chewing range (chew_entr)
3	Maximum value (Max)	22	Ratio: chew_ene / spectr_ene
4	Median value (Med)	23	Energy spectrum in walking range ³ (walk_ene)
5	Ratio: MAV / RMS	24	Entropy of spectrum walking range (walk_entr)
6	Ratio: Max / RMS	25	Ratio: walk_ene / spectr_ene
7	Ratio: MAV / Max	26	Energy spectrum in talking range ⁴ (talk_ene)
8	Ratio: Med / RMS	27	Entropy of spectrum talking range (talk_entr)
9	Signal entropy (Entr)	28	Ratio: talk_ene / spectr_ene
10	Number of zero crossings (ZC)	29	Ratio: chew_ene / walk_ene
11	Mean time between ZC	30	Ratio: chew_entr / walk_entr
12	Number of peaks (NP)	31	Ratio: chew_ene / talk_ene
13	Average range	32	Ratio: chew_entr / talk_entr
14	Mean time between peaks	33	Ratio: walk_ene / talk_ene
15	Ratio: NP/ZC	34	Ratio: walk_entr / talk_entr
16	Ratio: ZC/NP	35	Fractal dimension
17	Wavelength	36	Peak frequency in chewing range (maxf_chew)
18	Number of slope sign changes	37	Peak frequency in walking range (maxf_walk)
19	Energy of the entire frequency spectrum ¹ (spectr_ene)	38	Peak frequency in talking range (maxf_talk)

¹Frequency range: 0.1-500 Hz; ²Chewing range: 1.25-2.5 Hz; ³Walking range: 2.5-10 Hz; ⁴Talking range: 100-300 Hz.

TABLE II. FEATURES EXTRACTED FROM THE HAND GESTURE SIGNAL

#	Description	#	Description
1	Num. of HtM gestures within epoch (num_HtM)	6	Wavelength (WL)
2	Duration of HtM (D_HtM)	7	Ratio: WL / Duration HtM
3	MAV of HtM	8	Ratio: D_HtM / num_HtM
4	Standard Deviation of HtM	9	Ratio: MAV_HtM / D_HtM
5	Maximum value (Max_HtM)		

TABLE III. FEATURES EXTRACTED FROM THE ACCELEROMETER SIGNALS

#	Description	#	Description
1	MAV of ACC_x (MAV _x)	12	Entropy of ACC_y (Entr _y)
2	SD of ACC_x (SD _x)	13	MAV of ACC_z (MAV _z)
3	Median of ACC_y	14	SD of ACC_z (SD _z)
4	Num. of zero crossings (ZC) for ACC_x	15	Median of ACC_z
5	Mean time between ZC for ACC_x	16	Num. of ZC for ACC_z
6	Entropy of ACC_x (Entr _x)	17	Mean time between ZC for ACC_z
7	MAV of ACC_y (MAV _y)	18	Entropy of ACC_z (Entr _z)
8	SD of ACC_y (SD _y)	19	Mean of {MAV _x , MAV _y , MAV _z }
9	Median of ACC_y	20	Mean of {SD _x , SD _y , SD _z }
10	Num. of zero crossings for ACC_y	21	Mean of {Entr _x , Entr _y , Entr _z }
11	Mean time between ZC for ACC_y		

Finally, each feature vector f_i was associated with a class label $t_i \in \{1, -1\}$, where $t_i = 1$ and $t_i = -1$ represented FI and NFI, respectively. The same rule used in the Sensor Fusion step was used here to assign class labels to each f_i vector. A dataset containing the pairs $\{f_i, t_i\}$ was presented to a classification algorithm for training, validation and testing.

C. Classification

ANN is a supervised learning technique that has shown excellent results for many pattern recognition and classification problems [28]. ANN is robust and flexible, can analyze complex patterns and can handle noisy data. In this study, a population-based classification model based on ANN was trained to discriminate between food intake and no food intake epochs. Implementation of a population-based model (group model) was preferred over subject-dependent models (individual models) to achieve a robust model that included intra- and inter-subject variability and required no individual calibration.

A three layered (input layer, hidden layer and output layer) feed-forward neural network with the back-propagation training algorithm was the network topology implemented. The input layer consisted of 68 predictors (one for each feature) whereas the hidden layer contained a total of 10 neurons. The output layer consisted of one output neuron, which indicated the final class label t_i assigned to the input vector f_i in the test set. The hyperbolic tangent sigmoid was the transfer function used for the hidden and output layers. Training, validation and testing of the model was done using the Neural Network toolbox available in Matlab R2011b (The Mathworks Inc).

A leave-one-out cross-validation procedure was used to evaluate the performance of the ANN model. This allowed training and validating the model with data from 11 subjects (80 % of the data for training and 20% for validation) and testing with data from an independent subject. This procedure allowed each subject to be the test subject once. The final result was obtained by averaging the results across all subjects. Per-epoch classification accuracy was the metric used to evaluate the performance of the classification model [17]. For each subject, the accuracy value was computed using all subject's epochs available (including epochs labeled as "no food intake" in the Sensor Fusion stage). This was done to illustrate the performance the entire Sensor Fusion and Pattern Recognition methodology. Accuracy was computed as the average between Precision (P) and Recall (R) to account for the high number of true negatives that are typical in monitoring of food intake over long periods of time. These

metrics measured the ability of the model to recognize FI epochs while rejecting NFI epochs. Finally, a weighted average accuracy was computed to determine whether the size of the subject's datasets will impact on the final results. This average was weighted by the proportion of epochs of a subject's dataset in the total amount of epochs

IV. RESULTS

Information collected from the food diaries indicated that the foods consumed were mostly solids and liquids (a total of 40 different foods). There was only one event associated to semi-liquid intake (yogurt). Tea/coffee (consumed 12 times), rice (9 times), toast (6), oranges (6), chips (5), banana (5), granola bar (5), cereal (4), chicken (2), and waffles (2) were the foods of choice for participants.

The dataset used to create the food intake detection model contained approximately 10 hours of data labeled as food intake. Results of the leave-one-out cross-validation procedure for the different sets of features are showed in Table IV. The values in the table represent the mean and standard deviations of the food intake detection accuracy as well of the precision and recall results. The best performance was achieved by combining features from the jaw motion and the accelerometer signals (89.8%). Using such features, one half of the subjects presented accuracies above 90%, being 75.82% and 97.7% the lowest and highest accuracy values obtained, respectively. Inclusion of hand gesture features did not improve the results. Weighted average accuracy values showed no major differences when compared to the original average accuracy.

TABLE IV. CROSS-VALIDATION RESULTS FOR DIFFERENT FEATURE SETS¹

Feature set	Precision (%)		Recall (%)		Accuracy (%)		Weighted accuracy (%)
	Mean	SD	Mean	SD	Mean	SD	
f_{JM}	88.1	8.5	84.8	14.4	86.4	8.8	86.1
f_{HG}	38.0	35.4	7.1	11.1	22.6	21.4	22.3
f_{ACC}	74.1	12.6	67.3	20.5	70.7	10.7	70.8
$f_{JM} + f_{HG}$	87.7	9.4	86.3	9.0	87.0	6.6	86.7
$f_{JM} + f_{ACC}$	89.8	8.8	89.9	9.0	89.8	6.7	89.7
$f_{HG} + f_{ACC}$	73.4	12.0	70.7	18.3	72.0	10.0	72.3
$f_{JM} + f_{HG} + f_{ACC}$	89.1	8.5	90.4	4.9	89.7	5.5	89.5

¹ f_{JM} : jaw motion features; f_{HG} : hand gesture features; f_{ACC} : accelerometer features

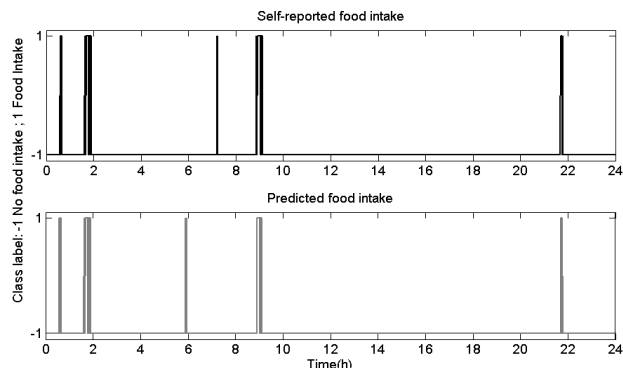


Fig. 5. Food intake prediction results for an average performance of the group model. It was able to predict (bottom graph) all major meals (2 hrs, 9 hrs and 22 hrs) but failed to predict a snacking event at 7 hrs with a false prediction at 6 hrs.

Fig. 5 shows the food intake detection results for one subject. The model correctly detected all major meals (lunch, dinner and breakfast at 2 hrs, 9 hrs and 22 hrs, respectively) and 1 snacking episode (granola bar at 1 hr) but incorrectly predicted

food intake at 6 hrs. Also, the model failed to predict a snacking event (slice of bread) at 7 hrs. The difference in the physical properties of the snacks may have been the reason for correctly detecting one out of two snacking events. Overall, the model correctly detected all of the 30 major meals consumed by the participants while incorrectly predicting the occurrence of 1 major meal. Regarding to snacking, 18 out of 19 episodes were correctly detected while 27 episodes were incorrectly predicted. Although this number may appear to be high, it only represented a total of 15 minutes of ingestion (2.5% of the total food intake time).

V. DISCUSSION AND CONCLUSIONS

The development of new strategies for objective and accurate monitoring of ingestive behavior of free-living individuals is imperative to overcome the current limitations of self-reported food intake. This paper introduced the Automatic Ingestion Monitor (AIM) as a wearable device for objective monitoring of food intake under free-living conditions. Compared to the state of the art, this work presents three new contributions: 1) the design and implementation of a novel wearable multi-sensor device that has the ability to monitor 24 hours of ingestive behavior without relying on self-report or any other actions from subjects, 2) the implementation of a robust methodology for reliably detecting food intake episodes in the presence of real-life artifacts in the sensor signals, and 3) the validation of the device and methodology in an objective study where 12 subjects wore AIM in free living during 24 hours without any restrictions on their eating behavior and activities.

The proposed wearable device presented several benefits that make it suitable for food intake monitoring without a conscious effort from the subjects. First, AIM integrated three different sensor modalities for monitoring of jaw motion, hand-to-mouth gestures, and body motion. Second, AIM was designed as a pendant worn on a lanyard around the neck which intended to satisfy the need for a socially acceptable device. Although further quantitative evaluation of necklace and sensors wear convenience is needed, preliminary studies showed that the jaw motion sensor presented high levels of wearing comfort and did not significantly affect the way subject eat their meals [29]. Third, the electric circuit embedded in the wireless module contained mostly low power components permitting 24 or more hours of data collection without need for recharging the battery. Finally, AIM presented a reliable data transmission system using Bluetooth technology. Minimal data loss was observed during the experiments. Only about 1.6% of the total data was lost most probably due to subjects walking away from the phone. All of these properties of AIM would theoretically keep subjects away from the burden of self-reporting their intake and provide objective measures of food consumption. Although AIM in its present form is rather a tool for studying of ingestive behavior over a period of several days, in the future it can be miniaturized into a less intrusive device suitable for long-term studies and personalized everyday monitoring.

An accurate detection of food intake in free living is highly desirable to obtain reliable information about dietary intake of individuals. Wearable systems integrating sensor signals with

complex pattern recognition algorithms have been implemented as a potential solution to self-reported intake [12], [14], [16]–[19]. Such systems achieved recognition rates ranging from 80% to 90%. However, direct comparisons with AIM results cannot be quantified due to they were obtained under different scenarios. The methodologies used in those wearable systems were developed based on data collected in the laboratory, which may result in high accuracies but with a poor generalization that would make them unreliable for free-living data. On the other hand, AIM incorporated a novel methodology based on free-living data that correctly detected around 90% of food intake. The advantage is that AIM can accurately detect ingestion events in a challenging scenario, where the real-world variability directly affects the eating behavior and is usually missing in laboratory settings, e.g. food selection, timing of meals (i.e. determined by external schedules, work, etc.), food intake environments and food intake behavior, which may vary during the course of a day [30], [31].

The methodology proposed consisted of three steps: sensor fusion, feature extraction, and classification. The sensor fusion step was implemented to balance the dataset and to remove potential outliers in the signals. Sensor fusion used information from the sensor signals to identify and remove more than 85% of NFI epochs for each subject (i.e. resting, walking, sleeping) while retaining most of the FI epochs (< 5% of FI epochs were removed per subject). In the feature extraction step, a set of time and frequency domain features were computed every 30 s of sensor signals and feed to the classifier. Table IV demonstrates that features from the jaw motion signals were the most important for food intake detection. The classification step presented a robust group model based on ANN. Although various advanced classification techniques and feature selection were previously evaluated by this research team [17], [22], [27], a computationally lightweight ANN algorithm was used because it can be easily implemented in a processor of a wearable system. This ANN model was able to identify when the major meals were consumed and when most of the snacking periods occurred.

In the proposed methodology it was assumed that food intake is mutually exclusive with vigorous physical activity (> 6 Metabolic Equivalents [32], for example, running). Indeed, although the study did not restrict or specify the way in which the food is to be consumed and involved individuals with origins from 5 different countries, having different lifestyles and ingestive behaviors, all of the participants demonstrated a tendency to eat while remaining sedentary. It is possible that some individuals may consume foods during physical activity of moderate intensity (3 to 6 MET) and the methodology will need to be validated in such scenarios. Lastly, the methodology requires no individual calibration meaning that it can potentially be used to detect food intake in a wide population. Another observation is that the optimization of the threshold values in the sensor fusion algorithm was based on the population as a whole. However, due to the diversity of subject population, the threshold values should be close to optimal and should not bias the recognition results.

The main limitation of this study is related to the use of self-report as the gold-standard for developing the food intake

detection methodology. Subjects may provide inaccurate information about dietary intake, thus leading to unreliable results. However, self-report error is mostly related to the amount of food consumed and the misreporting bias becomes significant over longer periods of time [25]. To minimize the reporting error, subjects were asked to report their intake in two different ways: by pushing a button during the chewing process and by writing down in a food journal the times at which foods were consumed. Push button data were analyzed in 30 s epochs and used to label sensor signals. This windowing procedure impacted on the accuracy of the timing reference. However, for the rule selected to label the epochs (see section III.a), the worst case scenario would have an epoch labeled as food intake when it contains 10 s of data associated to an ingestion event while the remaining 20 s will not be related to food intake. Even with this error, the timing of ingestion events using the windowed push button signal is significantly more accurate than timing events using the conventional food journal, which are typically filled post-factum and contain only approximate times.

The use of the push button and the monitoring system itself may change the eating behavior of individuals. However, there are two points to consider. First, even if the behavior was changed by using a push button in this study, AIM was able to accurately detect food intake episodes. Also, the use of push button is only required at the stage of algorithm development and in the future will not affect the behavior. Second, AIM is likely to affect behavior less than other methods as it does not rely on the self-report. The testing of this hypothesis is left to the future as it requires a much larger study and methods to assess behavior change under observation by different methods.

Another limitation of the methodology may be related to the capability of detecting liquid intake through the monitoring of jaw motion. The reported accuracy value corresponded only to solid food intake. Our previous studies suggested the presence of the characteristic jaw motion during a continuous intake (gulping) of liquids that is similar to the chewing present in the intake of solid foods [17]. Further studies are needed to determine the feasibility of the proposed methodology to detect liquid intake in free-living. In addition, effects of sensor positioning on the jaw and their impact on the accuracy of solid and liquid intake detection need to be studied.

Finally, the collected diary data was not enough to analyze the impact of certain activities on accuracy of food intake recognition. Our goal was to minimally load the participants with reporting burden, thus information about daily activities was restricted to what they reported in the diary. An observational study would be needed to obtain greater detail about subject's activities during the day and to better understand the behavior of AIM under daily life conditions.

The results presented in this study indicated a satisfactory level of robustness of the system for recognizing food intake events in free-living population, although testing on a larger and diverse population will be required to demonstrate performance of the device for individuals with uncharacteristic eating behaviors such as binge eating, bulimia nervosa, anorexia nervosa, compulsive eating disorders, etc. Lastly, future studies of the system will evaluate social acceptability

and subject compliance with wearing the device.

Although the prediction was done offline, in the future AIM is intended to perform real-time recognition and characterization of food intake (i.e. how much food is consumed) as well as to provide on-time feedback to individuals about their intake behavior. Estimation of the mass ingested could be achieved through the counts of chews [13] and the type and caloric density of food could potentially be determined by adding a camera triggered by the detection algorithm [19]. Ingested mass could also be predicted by acoustic recognition of chewing cycles and food types. A prior study showed that the bite weight of a reduced number of food types may be predicted using chewing sound features [15]. However, the main challenge facing this sound-based approach is the generalization of the bite weight prediction models for the broad variety of foods encountered in everyday life.

The development of AIM as a wearable device for monitoring food intake would intend to serve as a behavioral modification tool for correcting known ingestive behaviors leading to weight gain (snacking, night eating, weekend overeating) and would help to advance the study of free-living food consumption in obesity and in other eating disorders.

REFERENCES

- [1] S. J. Olshansky et al., "A potential decline in life expectancy in the United States in the 21st century," *N. Engl. J. Med.*, vol. 352, no. 11, pp. 1138–1145, Mar. 2005.
- [2] 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 J. Am. Med. Assoc.*, vol. 307, no. 5, pp. 491–497, Feb. 2012.
- [3] C. G. Fairburn and P. J. Harrison, "Eating disorders," *The Lancet*, vol. 361, no. 9355, pp. 407–416, Feb. 2003.
- [4] J. I. Hudson, E. Hiripi, H. G. Pope Jr, and R. C. Kessler, "The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication," *Biol. Psychiatry*, vol. 61, no. 3, pp. 348–358, Feb. 2007.
- [5] D. Sánchez-Carracedo, D. Neumark-Sztainer, and G. López-Guimerà, "Integrated prevention of obesity and eating disorders: barriers, developments and opportunities," *Public Health Nutr.*, vol. 15, no. 12, pp. 2295–2309, 2012.
- [6] F. E. Thompson and A. F. Subar, "Dietary assessment methodology," in *Nutrition in the Prevention and Treatment of Disease*, 2nd ed., Academic Press, San Diego, CA, 2008.
- [7] O. Amft, "Ambient, On-Body, and Implantable Monitoring Technologies to Assess Dietary Behavior," in *Handbook of Behavior, Food and Nutrition*, V. R. Preedy, R. R. Watson, and C. R. Martin, Eds. New York, NY: Springer New York, 2011, pp. 3507–3526.
- [8] D. A. Schoeller and P. Webb, "Five-Day Comparison of the Doubly Labeled Water Method with Respiratory Gas Exchange," *Am. J. Clin. Nutr.*, vol. 40, no. 1, pp. 153–158, 1984.
- [9] A. E. Black and T. J. Cole, "Biased Over- Or Under-Reporting is Characteristic of Individuals Whether Over Time or by Different Assessment Methods," *J. Am. Diet. Assoc.*, vol. 101, no. 1, pp. 70–80, Jan. 2001.
- [10] F. E. Thompson, A. F. Subar, C. M. Loria, J. L. Reedy, and T. Baranowski, "Need for technological innovation in dietary assessment," *J. Am. Diet. Assoc.*, vol. 110, no. 1, pp. 48–51, Jan. 2010.
- [11] 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, no. 5, pp. 525–541, May 2008.
- [12] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognit.*, vol. 41, no. 6, pp. 2010–2024, Jun. 2008.
- [13] E. 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, 2009.
- [14] O. Amft and G. Troster, "On-Body Sensing Solutions for Automatic Dietary Monitoring," *IEEE Perv. Comput.*, vol. 8, no. 2, pp. 62–70, 2009.
- [15] O. Amft, M. Kusserow, and G. Troster, "Bite Weight Prediction From Acoustic Recognition of Chewing," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 6, pp. 1663–1672, Jun. 2009.
- [16] 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.
- [17] E. Sazonov and J. M. Fontana, "A Sensor System for Automatic Detection of Food Intake Through Non-Invasive Monitoring of Chewing," *IEEE Sensors J.*, vol. 12, no. 5, pp. 1340–1348, 2012.
- [18] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Appl. Psychophysiol. Biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.
- [19] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G.-Z. Yang, "An Intelligent Food-Intake Monitoring System Using Wearable Sensors," in *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2012, pp. 154–160.
- [20] 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," *Ann. Biomed. Eng.*, vol. 38, no. 8, pp. 2766–2774, 2010.
- [21] 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," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 3, pp. 626–633, 2010.
- [22] J. M. Fontana and E. S. Sazonov, "A robust classification scheme for detection of food intake through non-invasive monitoring of chewing," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 4891–4894.
- [23] P. Lopez-Meyer, Y. Patil, S. Tiffany, and E. Sazonov, "Detection of Hand-To-Mouth Gestures Using a RF Operated Proximity Sensor for Monitoring Cigarette Smoking," *Int. J. Smart Sensors Intell. Syst.*, vol. in review, 2012.
- [24] R. H. Striegel-Moore, D. L. Franko, D. Thompson, S. Affenito, A. May, and H. C. Kraemer, "Exploring the typology of night eating syndrome," *Int. J. Eat. Disord.*, vol. 41, no. 5, pp. 411–418, 2008.
- [25] P. Suchanek, R. Poledne, and J. A. Hubacek, "Dietary intake reports fidelity - Fact or fiction?," *Neuroendocr. Lett.*, vol. 32, no. SUPPL. 2, pp. 29–31, 2011.
- [26] E. Stellar and E. E. Shrager, "Chews and swallows and the microstructure of eating," *Am. J. Clin. Nutr.*, vol. 42, no. 5 Suppl, pp. 973–982, Nov. 1985.
- [27] J. M. Fontana, M. Farooq, and E. Sazonov, "Estimation of Feature Importance for Food Intake Detection Based on Random Forests," in *2013 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.
- [28] M. Cohen and D. Hudson, *Neural Networks and Artificial Intelligence for Biomedical Engineering*, 1st ed. Wiley-IEEE Press, 1999.
- [29] J. M. Fontana and E. S. Sazonov, "Evaluation of Chewing and Swallowing Sensors for Monitoring Ingestive Behavior," *Sens. Lett.*, vol. 11, no. 3, pp. 560–565, 2013.
- [30] J. M. de Castro, "The control of food intake of free-living humans: putting the pieces back together," *Physiol. Behav.*, vol. 100, no. 5, pp. 446–453, Jul. 2010.
- [31] J. M. de Castro, "Eating behavior: lessons from the real world of humans," *Nutrition*, vol. 16, no. 10, pp. 800–813, Oct. 2000.
- [32] I.-M. Lee, C.-C. Hsieh, and R. S. Paffenbarger Jr., "Exercise intensity and longevity in men: The Harvard Alumni Health Study," *J. Am. Med. Assoc.*, vol. 273, no. 15, pp. 1179–1184, 1995.