AutoDietary: A Wearable Acoustic Sensor System for Food Intake Recognition in Daily Life

Yin Bi, Mingsong Lv, Chen Song, Wenyao Xu, Nan Guan, *Member, IEEE*, and Wang Yi, *Fellow, IEEE*

***Abstract*— Nutrition-related diseases are nowadays a main threat to human health and pose great challenges to medical care. A crucial step to solve the problems is to monitor the daily food intake of a person precisely and conveniently. For this purpose, we present AutoDietary, a wearable system to monitor and recognize food intakes in daily life. An embedded hardware prototype is developed to collect food intake sensor data, which is highlighted by a high-fidelity microphone worn on the subject's neck to precisely record acoustic signals during eating in a noninvasive manner. The acoustic data are preprocessed and then sent to a smartphone via Bluetooth, where food types are recognized. In particular, we use hidden Markov models to identify chewing or swallowing events, which are then processed to extract their time/frequency-domain and nonlinear features. A lightweight decision-tree-based algorithm is adopted to recognize the type of food. We also developed an application on the smartphone, which aggregates the food intake recognition results in a user-friendly way and provides suggestions on healthier eating, such as better eating habits or nutrition balance. Experiments show that the accuracy of food-type recognition by AutoDietary is 84.9%, and those to classify liquid and solid food intakes are up to 97.6% and 99.7%, respectively. To evaluate real-life user experience, we conducted a survey, which collects rating from 53 participants on wear comfort and functionalities of AutoDietary. Results show that the current design is acceptable to most of the users.**

***Index Terms*— Food intake recognition, wearable sensor, acoustic signal processing, embedded system.**

I. INTRODUCTION

A KEY factor in maintaining healthy life is balancing energy intake and expenditure. Abnormalities in this balance can lead to diseases, such as obesity, anorexia, and other eating disorders, which may furthermore deteriorate into chronic diseases if not seriously treated [1]. A crucial step to solve the problems is to *continuously* measure daily calorie balance [2]. There are many off-the-shelf solutions to measure calorie expenditure, such as Fitbit, Philips DirectLife, etc. However, continuously and non-invasively monitoring calorie *intake* remains a challenge. Currently, the common solutions rely on users' self-reports, which are neither convenient nor precise since food intakes are versatile and energy contained in different food may vary significantly [3]. It is highly desirable to develop accurate and easy-to-use methods to monitor the food eaten and predict energy intake. Although some food intake monitoring methods exist, they are either inaccurate or involve complex sensor systems, which limits their adoption in daily life.

To address this challenge, we propose AutoDietary, a wearable system to recognize food types by monitoring eating process. The system is mainly composed of two parts: (i) Embedded Hardware System and (ii) Smartphone Application. An embedded hardware is developed to collect and pre-process food intake data. The highlight is the necklace-like acoustic sensors to pick up high-quality sound signals of eating behaviors in a convenient and non-invasive manner. The data are then transmitted via bluetooth to a smartphone, where food types are recognized. We also developed an application which not only aggregates food recognition results but also provides the information in a user-friendly way and offers suggestions on healthier eating, such as the user should chew slower or should intake adequate hydration.

Specifically, food types can be distinguished from chewing information, since the energy exerted in mastication basically depends on the structural and the textural properties of the food material and can be extracted from the chewing sound. Food type recognition consists of several steps. The acoustic signals are firstly framed. Then, the sound frames are processed by the hidden Markov model (HMM) [4] based on the Mel Frequency Cepstrum Coefficients [5] to detect the chewing events. Moreover, we also detect fluid intake by swallowing events. Then each event is processed to extract several key features containing both time/frequency-domain and non-linear information. A light-weight decision tree based algorithm is adopted to recognize the type of food intake.

To evaluate our system, experiments are conducted involving 12 subjects to eat 7 different types of food. More than 4,000 food intake events are collected and evaluated. Results show that the accuracy of identifying chewing/swallowing events is 86.6%, based on which an accuracy of 84.9% can be achieved for food type recognition. To classify liquid and solid food, the accuracy can be up to 97.6% and 99.7%, respectively. We also conducted a survey to investigate user experience of AutoDietary. Results show that the current design (regarding wear comfort and functionalities) is acceptable to most users. By continuously monitoring eating behavior, AutoDietary can provide customized suggestions to healthier eating habits, and

Manuscript received July 24, 2015; accepted July 31, 2015. Date of publication August 17, 2015; date of current version January 12, 2016. This work was supported by the National Science Foundation of China under Grant 61300022 and Grant 61370076. The associate editor coordinating the review of this paper and approving it for publication was Prof. Subhas C. Mukhopadhyay. *(Corresponding author: Mingsong Lv.)*

Y. Bi, M. Lv, N. Guan, and W. Yi are with the College of Information Science and Engineering, Northeastern University, Shenyang 110004, China (e-mail: biyin0125@gmail.com; lvmingsong@ise.neu.edu.cn; guannan@ise.neu.edu.cn; wangyi@ise.neu.edu.cn).

C. Song and W. Xu are with the Department of Computer Science and Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260-2500 USA (e-mail: csong5@buffalo.edu; wenyaoxu@buffalo.edu).

Digital Object Identifier 10.1109/JSEN.2015.2469095

can be used as medical auxiliaries in therapy of nutrition-related diseases.

The article is organized as follows. Section II surveys existing methods for food intake recognition and their applications. We present the system architecture of AutoDietary in Section III and explain the core algorithms in Section IV. Experimental evaluation is given in Section V. Discussion & Future Work are presented in Section VI, and Section VII concludes the article. The survey results on the user experience of AutoDietary are provided in Appendix.

II. Related Work

Recognizing food types in food intake monitoring has been of great interests to the research community [6]. Meyer et al. developed a methodology to study the ingestive behavior by non-invasive monitoring of swallowing and chewing [7]. The main objective is to research the behavioral patterns of food consumption and producing volumetric and weight estimates of energy intake. The monitoring is conducted by a sound sensor located over laryngopharynx and by a bone conduction microphone detecting chewing through a below-the-ear strain sensor. Obviously, the composite microphone system reduces the wearability and comfortability.

Amft [8] presented an acoustic ear-pad sensor device to capture air-conducted vibrations of food chewing. To recognize the food types, Amft derived spectral features from all continuous chewing sounds, then averaged these features using multiple sliding windows. A combination of a Fisher discriminant filter and a naive Bayes classifier was used to perform feature reduction and food classification respectively. Amft record 375 chewing sequences with 4 different foods totally, and an overall classification accuracy of 86.6% was obtained. A major drawback of their system is that it requires multiple microphones to collect acoustic signals and some of the microphones are placed in the ear canal. Therefore, wearing such a system is very unpleasant for the users. In contrast, we use a single necklace-like device to collect acoustic signals, which is more convenient and comfortable to wear. Moreover, the experiments in [8] only include 4 types of food while in our paper 7 types are included.

Pabler and Wolff [9] proposed an approach based on the sound produced during food intake. The sound signals are recorded non-invasively by miniature microphones in the outer ear canal. In this work, hidden Markov models are used for the recognition of single chewing or swallowing events. Food intake cycles are modeled as event sequences in finite-state grammars, and recognition of consumed food is realized by a finite-state grammar decoder based on the Viterbi algorithm [10]. A database of 51 participants eating seven types of food and consuming one drink has been developed. The event detection accuracy is 83% and the food recognition accuracy is 79% on a test set of 10% of all records. While the method in [9] uses hidden Markov models for food classification, our paper use hidden Markov models for event detection and use decision trees for food classification to achieve higher food recognition accuracy. Similar to the approach in [8], microphones placed in the ear canals are used in [9] to collect sound signals, which is less comfortable to wear.

Radio frequency identification (RFID) tags on food packages are used to detect and distinguish how and what people eat [11]. The amount of food eaten should be recorded by a dining table with integrated weight scales, which records bite weights and assigns these weights to each person at the table. While this sensor device is still under development, several limitations have been mentioned. A fatal draw of this approach is the restriction to only one location where meals could be recorded, which makes it impractical for eating monitoring under free living conditions and not applicable for many applications such as energy balance monitoring. Moreover, the system is more expensive to deploy and requires extra efforts to attach RFID tags on every food package available.

Bite Counter invented by researchers of Clemson University are used to identify food intake gestures, chewing and swallowing to provide timing and food category information [12]. They used a watch-like configuration of sensors to continuously track wrist motion throughout the day and automatically detect periods of eating. Food category and eating habits are obtained by analyzing periods of vigorous wrist motion. This method successfully discriminates different events by the gesture behavior such as eating with fork and knife, drink from a glass, eat with a spoon, or the use of hands to approach food to the mouth. The motion sensor jacket developed was a research prototype and less complex sensors are being developed.

Lester and Tan [13] presented a method that utilizes optical, ion selective electrical pH, and conductivity sensors to sense and recognize daily fluid intake. They used a smart cup which combines pH, conductivity, and light spectrum to fingerprint different liquids and allow them to distinguish different beverages for long-term fluid intake monitoring. They described feasibility experiments that suggest that it is possible to reliably recognize specific drinks with up to 79% recognition accuracy for 68 different drinks.

Video fluoroscopy and electromyography (EMG) are considered the gold standard in studies of deglutition [14]. Video fluoroscopy depends on bulky and potentially unsafe equipments, while EMG is too invasive due to frequently used subcutaneous placement of electrodes in the masseter, suprahyoid, and infrahyoid muscles to avoid interference from the muscles of the neck. Gao [15] used video analysis to record eating behavior of people. Typical food intake movements are tracked in video recordings. The meal duration can be logged for every inhabitant of the retirement home. This solution is restricted to special locations where cameras are installed. In a study by Wu and Yang [16], intake of fast food was recorded using a wearable camera to recognize food by frequently taking images and comparing them with reference images in the database. Both visual-based solutions suffer from a common problem: the recognition accuracy drops significantly with changing ambient light and in the cases in which relevant objects are hidden behind other objects such as furniture or people.

Zhou et al. presented a smart table surface system to support nutrition monitoring [17]. The system is based on a smart table cloth equipped with a fine grained pressure textile matrix and

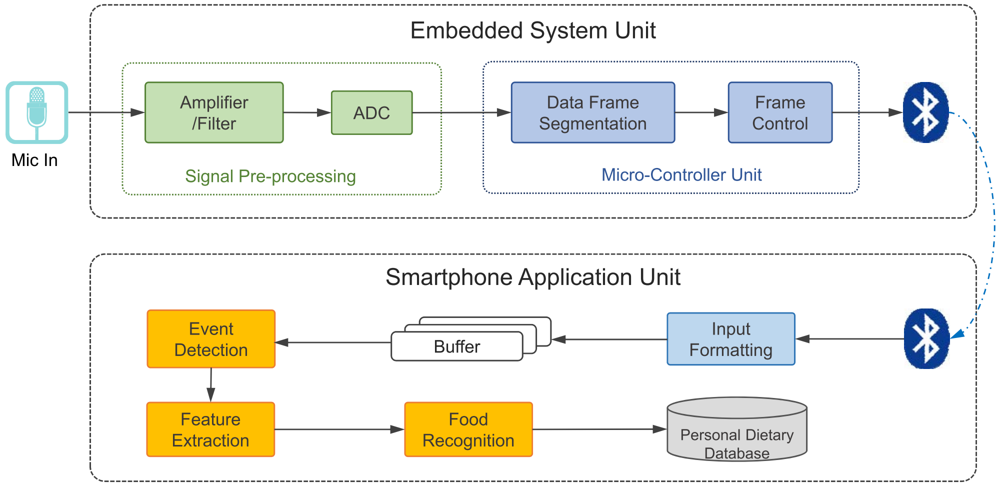


Fig. 1. System architecture of AutoDietary.

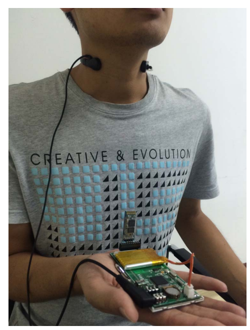
a weight sensitive tablet. Food intake related actions, such as cutting, scooping, stirring, poking, etc., are detected and recognized. Based on these actions, the food content and weight can be estimated. The system does not work well in food type recognition and food amount estimation, since (i) different food of similar types may lead to similar actions, and (ii) to use cutting force to estimation food amount is not accurate and does not provide stable estimations.

III. System Architecture

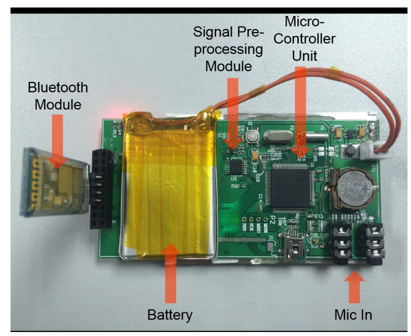
AutoDietary is mainly composed of two components: an embedded system unit for acoustic data acquisition and pre-processing, and an application running on the smartphone that implements food type recognition and provides an information interface for the users. The main architecture of the system is shown in Fig. 1. The rest of this section presents the hardware design of the embedded system unit and the smartphone application. We elaborate the details on the food type recognition algorithms in Section IV.

A. Acoustic Sensors

A high-precision and high-fidelity throat microphone is employed to pick up acoustic signals during eating. The microphone is worn over the user's neck close to the jaw. The throat microphone converts vibration signals from the skin surface to acoustic signals rather than picking up sound wave pressure as most common microphones do. This principle enables very high quality signals to be collected for the specific purpose of AutoDietary by efficiently reducing the interference from ambient noise. Additionally, the throat microphone is comfortable to wear and can be better accepted by users (see the user experience survey in Appendix), compared with other high-quality acoustic sensors, such as ear-canal microphones. The throat microphone adopted in our data acquisition provides a dynamic range of 36 ± 3dB and a frequency range of 20Hz-20kHz, which is capable of acquiring chewing and swallowing sound. The effect of wearing the throat microphone is depicted in Fig. 2(a).



(a)



(b)

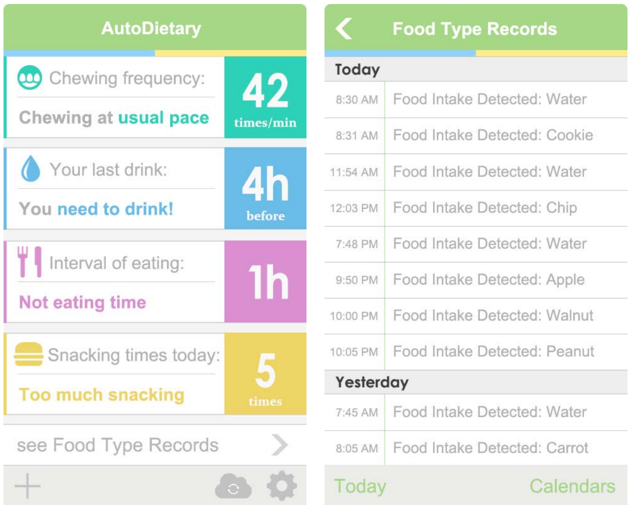
Fig. 2. System illustrations. (a) A user wearing AutoDietary; (b) Details of the hardware board.

Note that the throat microphone can be further miniaturized in the future to be more portable and comfortable.

B. Hardware Board

An embedded hardware board, shown in Fig. 2(b), is designed for data pre-processing and transmission. When acoustic data are collected from the throat microphone and input from Mic In, they are amplified and filtered for better signal quality. Then the analog signals are converted to digital signals for later steps. The amplifier adopted is LM358 [18], featured by its high common-mode rejection ratio, low noise and high gain. The total gain of the amplifier is 250, and the cutoff frequency of the low-pass filter is 3000Hz. The adopted AD converter is TLV2541 [19] with a sampling rate of 8000Hz and 12 bit resolution.

The digital signals are then sent to a micro-controller via the I^2C interface. Sound signals are segmented into frames for later processing. The micro-controller is also responsible for frame admission control of raw signals from the throat microphone. The adopted micro-controller is the ultra low-power MSP430F5438 [20], which is widely used in energy-constrained consumer electronic products and portable medical equipments. The data frames are sent to a bluetooth



(a) (b)

Fig. 3. Screenshots of the current smartphone application. (a) The screen displaying healthy eating suggestions; (b) The screen displaying food type recognition results corresponding each event.

module through UART using the SPI transport protocol, and further sent to the smartphone by the bluetooth module with a SPP profile at a data rate of 150KBit/s. Bluetooth ensures reliable wireless data transmission within a distance of 10m, which is adequate for our system. A rechargeable LiPo battery is adopted to power the whole hardware board.

C. Smartphone Application

The application developed on the smartphone side has two major roles. First, it performs food type recognition by implementing the main algorithms detailed in the next section. Second, it serves as a data manager and provides an interface to the user. Fig. 3 gives some screenshots of the application. To use our system, a user simply wear and power on AutoDietary, and start the application. When the user starts to eat, the system will perform food type recognition and store the detailed data into a database. The user can not only check the detailed records (as shown in Fig. 3(b)), but also go through the suggestions on healthier eating habits which are obtained by analyzing the data collected. Eating guidance currently includes: a) more regular and balanced diets, b) alerts on abnormal chewing speed, c) suggestions on hydration intake, d) alerts on excessive snacking in a day, and e) suggested intervals between meals. Based on the key information and main framework provided by AutoDietary, developers can further expand the application with new features on personal health management.

IV. Food Type Recognition

Food type recognition takes the continuous sound frames during eating as input and produces a recognized food type for each identified chewing or swallowing event, which is realized by three consecutive steps shown in Fig. 4. The first step uses the hidden Markov model based on the Mel frequency cepstrum coefficients to detect chewing or swallowing events from the continuous sound frames. Frames within an event is maintained together and those not involved are discarded. In the second step, each event is processed to extract the key features that best distinguish different food types. The last step takes the feature values for each event and evaluates them with prior knowledge represented by a decision tree to predict the food type corresponding to the event. The results are stored in a personal dietary database for future analysis. This helps to reduce the computation time and memory usage, which leads to longer battery lifetime.

A. Event Detection

A recording sample contains a sequence of chewing and swallowing events separated by silence periods during which no event occurs. In this part, we use hidden Markov model (HMM) to automatically detect the chewing and swallowing events from each continuous recording sample. HMM has been widely used in many fields, such as speech recognition. In recent decades, many different acoustic event detection/classification methods based on HMM has been proposed [21]–[23].

A recording sample is framed into frames (every frame is 0.5s and the overlap is 0.25s). We formulate the goal of acoustic event detection in a way similar to speech recognition: to find the frame sequence that maximizes the posterior probability of the frame sequence $W = (W_1, W_2, \cdots, W_M)$, given the observations $O = (O_1, O_2, \cdots, O_T)$:

$$W = arg\ max_W P(W/O) = arg\ max_W P(O/W)P(W). \tag{1}$$

The model $P(W/O)$ is the HMM for acoustic events and silence periods, with 4 emitting states and left-to-right state transitions. Observations O are composed of 32 Mel Frequency Cepstrum Coefficients for event sequence or silence sequence. According to O and the original model, the HMMs for event and silence are trained using the Baum-Welch algorithm [10]. The Viterbi algorithm is used to compute the posterior probability of every observation under event and silence HMMs, respectively. A frame belongs to some acoustic event if its posterior probability under the event HMM is larger than that under the silence HMM.

To be used for food recognition in the next steps, we label each frame belonging to an event with bit 1 and each non-event frame with bit 0. Obviously, a consecutive sound frames all labeled by 1 margined by zeros correspond to a chewing or swallowing event among the sequence of frames of a recording sample.

B. Feature Extraction

The accuracy of food type recognition heavily depends on the selection of event features which can best distinguish different food types. In this work, we extract time-domain features, frequency-domain features and non-linear features for each event, listed in TABLE I, II and III, respectively.

In the time domain, statistical features are computed for each chewing event, including *high peak value*, *low peak*

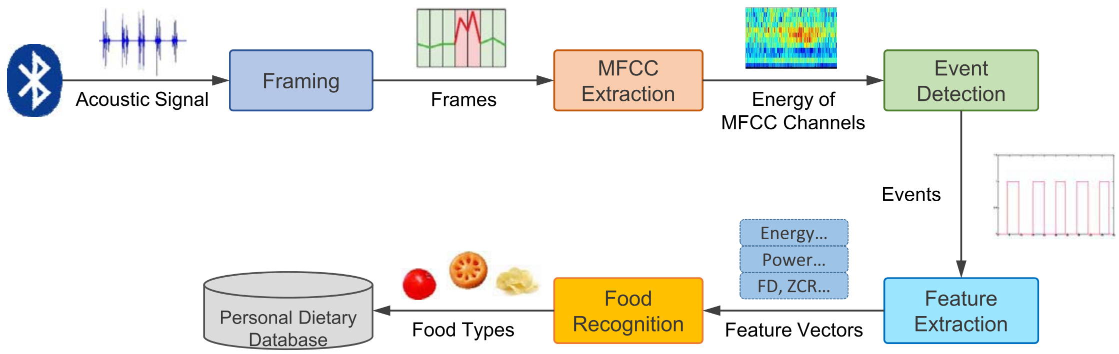


Fig. 4. Food intake signal processing data flow.

TABLE I
TIME DOMAIN FEATURES

Features	Descriptions
High Peak	Maximum value of a event
Low Peak	Minimum value of a event
Mean	Average value of a event
Variance	The square of Std_variance
std_variance	Measure of spreadness of event
ZCR	Measure related with frequency
Skewness	The degree of asymmetry of the data distribution
Kurtosis	whether signal is peaked or flat relative to a normal distribution
Interquartile	Measure of statistical dispersion

TABLE II
FREQUENCY DOMAIN FEATURES

Features	Descriptions
Pmax	Maximum power
Pmean	Mean power of event
P250*(i-1)-250*i, (i=1,2,,,10)	Power at 250*(i-1) - 250*i Hz
Ei (i=1,2,,,8)	Energy of 8 sub-bands after WPD

TABLE III
NON-LINEAR FEATURES

Features	Descriptions
DetrenFlu	Quantify fractal scaling properties
AppEn	Measure of regularity & complexity
FraDimen	Index of complexity how detail in a event changes with the scale
Hurst-E	Measure of the smoothness
CorDimen	Measure of fractal dimension

value, *mean value*, *variance* and *standard deviation* of the signals in the event. Most of these features have been intensively used in related studies and are demonstrated to be useful for pattern recognition [7], [24], [25]. Besides, we add 4 features, *zero crossing rate*, *skewness*, *kurtosis* and *interquartile range*, to better represent the geometry characteristics of the signals.

Frequency domain features can describe the distribution of the signals over a given range of frequencies. In this study, Power Spectrum Density (PSD) of the signal in each segment is estimated based on Welch's method with a Hamming window [26]. With respect to PSD, the maximal power (Pmax) and mean power (Pmean) for a specific frequency are computed. The energy for each 250Hz frequency band ranging from 0 to 2.5kHz is computed using numerical integration [27]. We used Wavelet Packet Decomposition (WPD) [27] to extract frequency-domain features. WPD decomposes signal energy on different time-frequency plains, and the power of the signal can be computed, which is proportional to the integration of the square of amplitude by WPD [28]. In our implementation, "db3" wavelet is used to decompose the signal into 3 levels, and Shannon Entropy is used to measure sub-band energy. Therefore, 8 sub-band energy values are extracted to form the feature vector.

It is now generally acknowledged that non-linear techniques are able to describe the progress of signals generated by biological systems in a more effective way [29]. These techniques are based on the concept of chaos and have been applied in many fields including medicine and biology. Non-linear features, such as slope of detrended fluctuation analysis (DetrenFlu), approximate entropy (AppEn), fractal dimension (FraDimen), Hurst exponent (Hurst-E) and correlation dimension (CorDimen), have been demonstrated useful to describe a signal [29], [30]. Therefore, we add these 5 non-linear features to describe each event.

A total of 34 features for each event are extracted to represent its acoustic characteristics. Figure 5 gives the signals and the corresponding spectrograms for four food intake events, each of which corresponds to a different food.

C. Recognition and Classification of Food Types

In this step, we intend to predict the food type corresponding a chewing or swallowing event, which is actually achieved by evaluating the feature values based on prior knowledge. In this study, prior knowledge are vast experimental feature values obtained from thousands of pre-recorded events for different types of known food. These data are used as the guidance to classify the food type of a new event.

A major design issue is how to efficiently evaluate the given feature values with the vast prior knowledge. In our work, we employ the decision tree [31], an approach widely used in activity recognition [32], text classification [33], etc., to compactly represent the key information contained in our prior knowledge.

An exemplary decision tree used by AutoDietary is shown in Fig. 6, which is actually a chunk of the whole tree to recognize cookie and water. To determine the food type of a given event, a decision process starts from the root node.

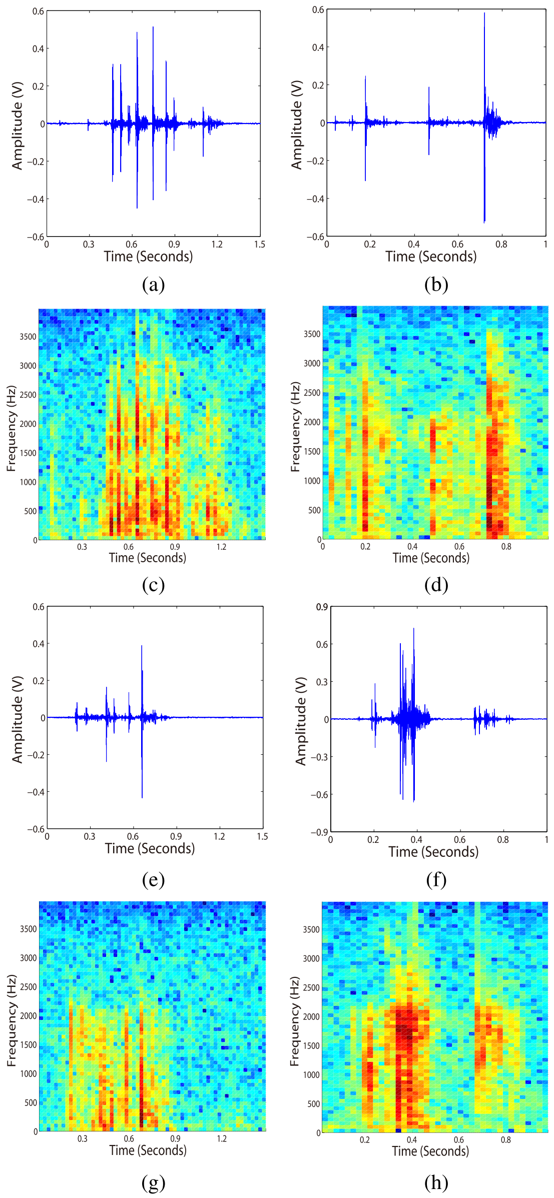


Fig. 5. Examples of time domain and frequency domain features for food intake events: chewing (a) cookie, (b) carrot, (e) walnut and (f) swallowing water, (c), (d), (g) and (h) are the corresponding spectrograms.

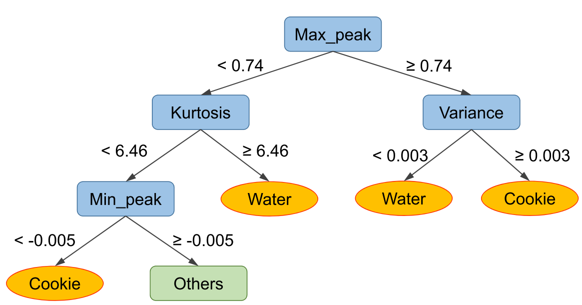


Fig. 6. Part of the decision tree to recognize cookie and water.

The *Max_peak* feature value is first checked, and the branch satisfying the constraint is taken. The process proceeds and more feature values may be checked in the intermediate nodes.

TABLE IV
PHYSICAL CHARACTERISTICS OF THE TWELVE SUBJECTS

Subject No.	**Gender**	**Age**	**Height(cm)**	**Weight(kg)**	**BMI**(kg/m^2)
1	Female	13	162	48	18.3
2	Male	14	172	57	17.9
3	Male	14	165	53	19.4
4	Male	24	170	60	20.7
5	Female	23	163	51	19.2
6	Male	24	184	63	18.6
7	Male	23	176	68	21.9
8	Female	23	169	60	21.0
9	Male	48	178	81	25.6
10	Female	47	166	67.5	24.5
11	Male	49	174	71	23.5
12	Female	44	164	63	23.4

Once a leaf node is reached, a final decision on the food type is returned. Our decision tree is created using the Matlab built-in function `classregtree`, which applies Ginis diversity as the separation criterion [34]. Multiple decision paths for a single type is possible. This is because during the chewing process of one bite, the sizes of the food chunks in the mouth become smaller and smaller, which causes variations in the feature values. However, we can still classify the chewing events for the same food type by looking at different feature combinations (i.e., different decision paths). When the food chunks are small enough, in fact, it is almost impossible to distinguish different foods. In such cases, since the major acoustic features are so weak, the corresponding chewing events are already ruled out in the event detection step.

V. EVALUATION

A. Experimental Setup and Data Collection

Experimental data are collected from 12 subjects, the details of whom are listed in TABLE IV. Each subject wearing the throat microphone is required to eat food in single pieces with his/her usual pace. The experiments were conducted in a laboratory environment with low noise. Food types are excluded only if the participant exhibits a strong dislike. The subjects are suggested to reduce the movement of head, speaking, coughing and other activities during the experiments.

7 different types of food, including apples, carrots, cookies, potato chips, walnuts, peanuts and water, are used to evaluate our systems. Note that drinking water is also treated as eating a special type of food, which is also used in classifying solid food from liquid food. In total, we have collected 171 samplings composed of 4047 events (including 54 bite events, 3433 chewing events, and 560 swallowing events). Fourfold cross validation was performed in the following experiments with three folds used for training the model and one fold used for validation.

B. Event Detection Accuracy

We evaluate the accuracy of using the HMM approach for event detection (Sec. IV-A) with two different metrics. For all the sound frames contained in a sample, we define:

- TP: the number of event frames correctly recognized;
- FN: the number of event frames falsely recognized;
- FP: the number of non-event frames falsely recognized;
- TN: the number of non-event frames correctly recognized.

TABLE V
EVENT DETECTION ACCURACY

	TP+TN	TP+FN+FP+TN	Accuracy
Subject 1	275	323	85.1%
Subject 2	334	407	82.1%
Subject 3	273	330	82.7%
Subject 4	890	1017	87.5%
Subject 5	919	986	93.2%
Subject 6	968	1029	94.1%
Subject 7	958	1073	89.3%
Subject 8	1058	1189	89.0%
Subject 9	270	318	84.9%
Subject 10	231	268	86.2%
Subject 11	268	328	81.7%
Subject 12	218	261	83.5%
Average			86.6%

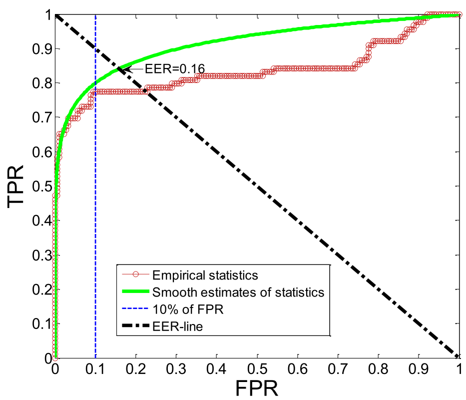


Fig. 7. Receiver operating characteristic curve (Empirical statistics is computed based on detection results, and smooth estimate is fitted by binormal model using detection results.)

And we use the true positive rate (TPR), the false positive rate (FPR), and $Accuracy$ as follows:

$$Accuracy := (TP + TN)/(TP + FN + FP + TN) \quad (2)$$

$$TPR := TP/(FN + TP) \quad (3)$$

$$FPR := FP/(FP + TN) \quad (4)$$

Table. V shows the event detection accuracy results for each of the 12 subjects. The smallest witnessed $Accuracy$ is 81.7% and the overall $Accuracy$ is 86.6%.

To further explore the accuracy w.r.t different FPR, we construct the receiver operating characteristic curve (ROC) [35], which illustrates the performance of a binary classifier system as its discrimination threshold is varied. ROC is created by plotting the TPR against the FPR, as shown in Fig. 7. The red curve is constructed from the empirical data. However, this curve may be imprecise to describe the receiver operating characteristics, due to inadequate data or uneven data distribution. Therefore, we fit the smooth estimation (the green curve in Fig. 7) out of the empirical data by the binormal model which assumes that the detection results follow two independent Gaussian distributions. Basically, the ROC exhibits the accuracy of event detection with different levels of tolerance to false. For example, given a false positive threshold of 0.1, the TPR achieves around 80%. It can be computed that the area under the curve (AUC) is around 0.9, which indicates that our approach can accurately detect events across the whole range of false positive threshold. The back-diagonal in Fig. 7 represents the Equal Error Rate (EER) line, which is defined as the set of points on which FPR euqals FNR ($FNR = 1 - TPR$). The smaller the EER, the better the detection performance. The EER regarding the fitted curve is 0.16. The performance can meet our requirement: for food intake events, if there is at least 1 frame is recognized by the classifiers, the entire food intake acoustic event including this event frame can be detected. Note that after detecting the chewing and swallowing events, some index values, such as chewing frequency, can already be calculated to quantitatively evaluate the eating patterns of the subject.

TABLE VI
FOOD TYPE RECOGNITION PERFORMANCE OF EACH SUBJECT

	Recall	Precision	Accuracy
Subject 1	85.2%	88.2%	86.7%
Subject 2	86.6%	82.8%	84.7%
Subject 3	84.3%	78.8%	81.5%
Subject 4	86.3%	84.3%	85.3%
Subject 5	88.8%	89.9%	89.4%
Subject 6	84.4%	84.3%	84.3%
Subject 7	89.5%	89.7%	89.6%
Subject 8	90.1%	89.1%	89.6%
Subject 9	85.7%	83.4%	84.5%
Subject 10	91.5%	88.7%	90.1%
Subject 11	88.1%	86.9%	87.5%
Subject 12	89.4%	88.1%	88.8%
Average	87.5%	86.2%	87.1%

C. Food Type Recognition Accuracy

- *Evaluation w.r.t. Subjects:* To evaluate the recognition accuracy regarding different individuals, we use each subject's chewing and swallowing events to build a specific decision tree to evaluate other events of the same subject. Three metrics, $Precision$, $Recall$ and $Accuracy$, are used for the evaluation. The following equations give their definitions, where i is the index of each food type, and $n(i)$ is the number of all food intake samplings of type i.

$$Precision := \sum_i TP(i) / \sum_i (TP + FP)(i) \quad (5)$$

$$Recall := \sum_i TP(i) / \sum_i (TP + FN)(i) \quad (6)$$

$$Accuracy := \sum_i (TP + TN)(i) / \sum_i n(i) \quad (7)$$

TABLE VI gives the results for the 12 subjects, the accuracy of which ranges from 81.5% to 90.1%. Compared to Fig. IV, the results show that evaluation for subject with small *body mass index* (BMI) has the lower accuracy, which indicates that these subjects are thin with a small neck circum feretory and it is possible that when wearing the throat microphone, it does not perfectly fit on the skin of the subject. Thus, the quality of the sampled signals might be compromised.

- *Evaluation w.r.t. Food Types:* We also evaluate the recognition accuracy regarding each food type. For this part, similarly fourfold cross validation was performed in this experiment with three folds used for building the tree and one fold used for validation. Totally, we extract 579 apple chewing events, 607 carrot chewing events, 471 potato chips chewing

TABLE VII

CONFUSION TABLE OF FOOD TYPE RECOGNITION

Prediction / **Actual Food**	**Apple**	**Carrot**	**Chip**	**Cookie**	**Peanut**	**Walnut**	**Water**	**Precision**	**Accuracy**
Apple	512	21	11	16	9	9	1	88.4%	86.3%
Carrot	23	528	17	19	11	9	0	87.0%	84.9%
Chip	23	32	389	13	10	4	0	82.6%	82.9%
Cookie	18	20	23	517	6	8	0	87.3%	87.7%
Peanut	12	15	12	6	171	10	0	75.6%	75.5%
Walnut	19	17	9	9	19	295	0	80.2%	83.4%
Water	2	4	6	7	1	3	158	87.3%	93.3%
Recall	84.1%	82.9%	83.3%	88.1%	75.3%	87.3%	99.4%		

TABLE VIII

LIQUID/SOLID FOOD CLASSIFICATION

Prediction / **Actual Food**	**Solid**	**Liquid**	**Precision**	**Accuracy**
Solid	1647	5	99.7%	99.7%
Liquid	4	177	97.8%	97.6%
Recall	99.8%	97.3%		

events, 592 cookie chewing events, 221 peanut chewing events, 368 walnut chewing events and 221 water swallowing events. The results are listed in TABLE VII, in terms of *Precision*, *Recall* and *Accuracy*.

Our method achieves an average recognition accuracy of 84.9%. This result confirms that the recognition algorithm performed sufficiently well to be used in the exhibition. Specifically, recognition precision of peanut is the lowest (0.755). This implies that sounds of chewing peanut are similar to chewing other solid food. A possible reason is, for solid food, varying physical properties directly impact the sounds of mastication, and consequently influence food type recognition. Similarly, the sound of swallowing water is characterized with high frequency signal compared to chewing solid food, which makes swallowing water effectively recognized, with highest recall of 0.994. We plan to integrate adaptive methods into our recognition approach to further improve accuracy in the future.

• *Liquid/Solid Food Classification:* To enable the suggestion for users to intake adequate hydration, it suffices to distinguish liquid food from solid food. To test the recognition accuracy, we conduct another set of experiments which includes 1652 solid food chewing events and 181 liquid swallowing events from 12 subjects. For these experiments, we construct a decision tree specific for this purpose.

TABLE VIII shows the results of classification between solid and liquid foods. Only 4 liquid food events are incorrectly recognized as solid food events among 181 liquid food events; 5 solid food events are incorrectly recognized as liquid food events among 1652 solid food events, resulting in the accuracy for solids and liquids of 99.7% and 97.6% respectively, which is more than enough to remind the user with proper hydration intake.

D. Feature Sensitivity Analysis

In our experiments, we are interested in how much the 34 features contribute to the accuracy of food type recognition, which is quantitatively evaluated by the *information gain*. Information gain originates from information theory, and it

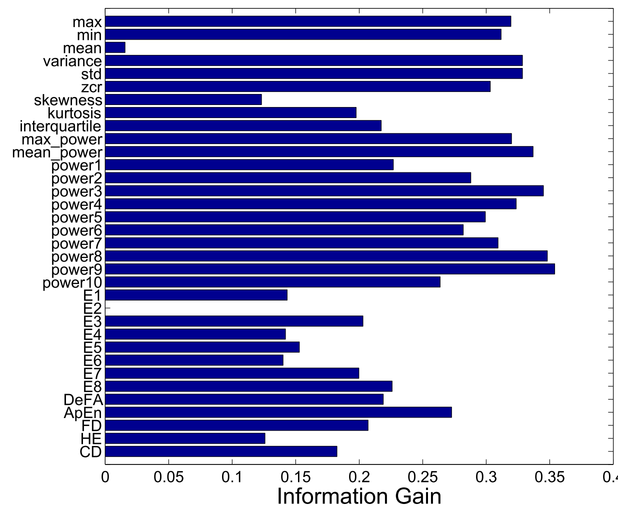


Fig. 8. Information gain of each feature (the larger the value, the more contribution the feature offers in food recognition).

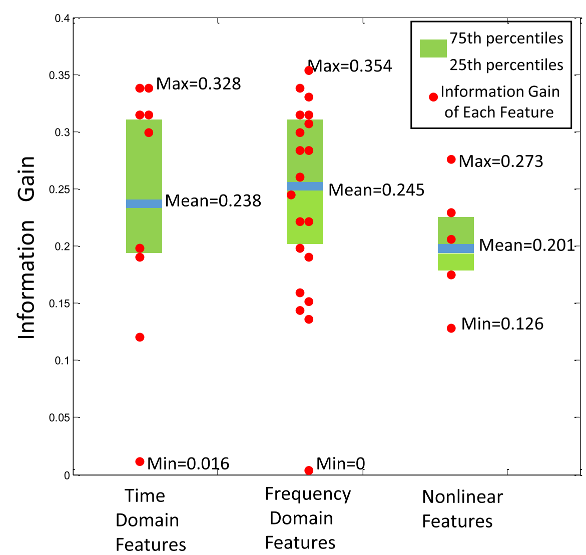


Fig. 9. Information gain of the three classes of features.

is one of the widely used approaches as a term importance criterion for text data [36]. The information gain of an outcome O from an attribute A is defined as the expected decrease in entropy of O conditioned on A. The following equations can be used to calculate the information gain about a discrete outcome O from a discrete attribute A, denoted by $IG(O, A)$. We use $H(O)$ to denote the entropy of O, $H(O/A)$ to denote the entropy of O given A, and $P(a)$ to denote the probability

AutoDietary Questionary

ID________
Age________
Gender________

We are interested in your comments and suggestions. Please take a few minutes to answer the following questions. The results of this evaluation will be used to improve future design of AutoDietary. Please score 1 to 5 to evaluate the questions. 1: the worst grade; 5: the highest grade.

Q1: Overall evaluation of comfort?	☐1	☐2	☐3	☐4	☐5
Q2: Convenience and portability of AutoDietary?	☐1	☐2	☐3	☐4	☐5
Q3: Do you mind the size of AutoDietary?	☐1	☐2	☐3	☐4	☐5
Q4: Are you satisfied with the tightness of throat microphone?	☐1	☐2	☐3	☐4	☐5
Q5: Overall evaluation of functions?	☐1	☐2	☐3	☐4	☐5
Q6: Do you believe the suggestions provided by AutoDietary?	☐1	☐2	☐3	☐4	☐5
Q7: Do the functions of AutoDietary meet your requirements?	☐1	☐2	☐3	☐4	☐5
Q8: AutoDietary does not influence daily eating and life?	☐1	☐2	☐3	☐4	☐5
Q9: Usefulness of AuroDietary?	☐1	☐2	☐3	☐4	☐5
Q10: Accepted price range?	☐0-100¥	☐100-200¥	☐200-300¥	☐>300¥	
Q11: Firstly considered factor?	☐Price	☐Appearance	☐Quality	☐Function	
Q12: Optimal charging frequency?	☐3 days	☐1 week	☐2 weeks	☐>2 weeks	

Fig. 10. The Questionary.

that attribute A takes on value a in the data.

$$IG(O, A) := H(O) - H(O/A) \tag{8}$$

$$H(O) := -\sum_{o \in (O)} P(o) \log P(o) \tag{9}$$

$$H(O/A) := \sum_{a \in (A)} P(a) H(O/A = a) \tag{10}$$

The information gains for the 34 features are computed and listed in Fig. 8. We can see some features, such as max_peak, ZCR, mean_power and power9 (power in 2250-2500Hz), exhibit higher information gain, which means these features have closer relevance to food type recognition.

We are also interested in which class of features are most relevant. The information gains of time-domain features, frequency-domain features and non-linear features are grouped and shown in Fig. 9. The mean information gains for the three classes are 0.238, 0.245, 0.201, respectively. Clearly, the frequency-domain features are comparably more relevant to food type recognition. However, the range of information gain regarding frequency-domain features is the largest, which indicates that their contributions are uneven. A closer look into the frequency-domain features shows that power features, such as max power, mean power and power at specific frequencies, are more relevant than energy features. With the results of information gain, we consider to remove unrelated features from food type recognition in the future, which may reduce computation overhead but introduce negligible accuracy loss.

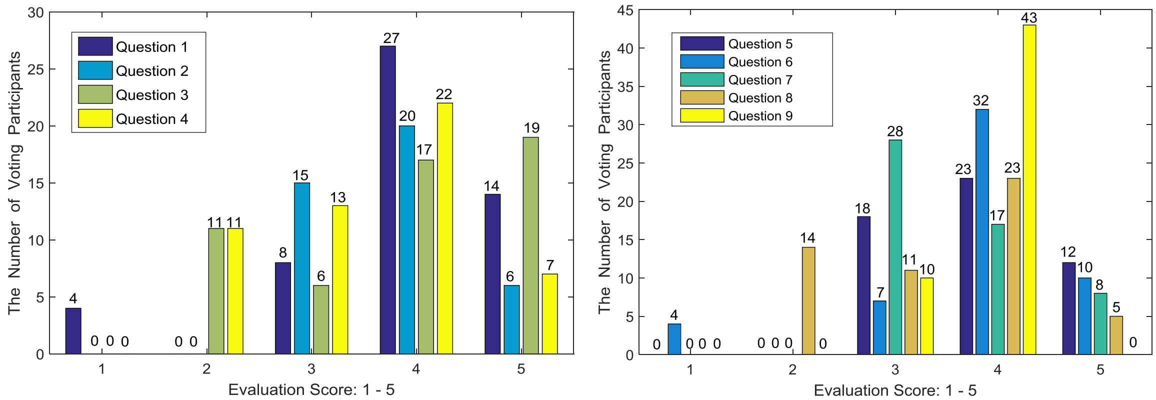
VI. Discussion and Future Work

From the experimental results we can see that AutoDietary has high performance in food type recognition, especially in distinguishing solid food from liquid food. These advantages enables us to provide trustable suggestions on proper eating. A major reason for the good performance is the high-precision and high-fidelity throat microphone adopted by AutoDietary. The throat microphone guarantees high quality signal samplings by effectively reduing the reference noise. A survey on user experience (regarding wear comfort and functionalities) shows that the current design of AutoDietary is acceptable by most users for daily use. AutoDietary now can be used in medical or physiotherapy studies with special interests in food intake behavior. For example, in diet control for diabetes patients, precisely monitoring daily food intake and thus provide proper eating suggestions can be very helpful to alleviate the disease; by precisely identifying bad eating habits and suggesting good ones, AutoDietary can help to reduce bowel disorders due to improper chewing and swallowing speed. Besides, AutoDietary can be especially useful for disabled and very sick people, for whom daily food intake monitoring that involves too much human intervention is not practical.

In the future, we plan to further improve AutoDietary in several aspects. First, experiments in this study are conducted in a lab environment with low noise; head movement, speaking and coughing are reduced as much as possible. The algorithms in sound signal processing and food type recognition will be improved to allow environmental noise and still maintain the same level of accuracy as the current version of AutoDietary. Second, the food recognition capability will be enhanced to handle a broader range of food types, even including food with composite textures. Third, we also plan to add new capabilities to AutoDietary, such as identifying the volume and weight of food intake, so as to precisely estimate daily calorie intake. Last but not least, we intend to further reduce the size of the microphone and the embedded system unit to optimize user experience. The specific target is to design an embedded system unit in a USB key size, which can be worn like a necklace pendant or easily put into the chest pocket. All these improvements will be validated in long-term real life scenarios.

VII. Conclusion

In this paper, we have presented AutoDietary, a comprehensive and preliminary solution for food intake recognition in daily life. We developed an embedded hardware to collect food intake sensor data, which is highlighted by a throat microphone comfortably worn on the subjects neck to precisely record acoustic signals during eating in a non-invasive manner.



(a) (b)

Fig. 11. Results of the Survey. (a) Results regarding wear comfort; (b) Results regarding functionalities.

TABLE IX
RESULTS ON QUESTION 10-12

Accepted price range	0-100¥	100-200¥	200-300¥	> 300¥
	28	25	0	0
Firstly considered factor	Price	Appearance	Quality	Function
	0	4	21	28
Optimal charging frequency	3 days	1 week	2 weeks	> 2 weeks
	6	29	7	11

The sensor data are then sent to a smartphone via bluetooth, where food types are recognized. Specifically, we use hidden Markov models to identify chewing or swallowing events, which are then processed to extract their time/frequency-domain and non-linear features. A light-weight decision tree based algorithm is adopted to recognize the type of food intake. We also developed an application on the smartphone which not only aggregates food intake recognition results but also displays the information in a user-friendly way and provides suggestions on healthier eating. Extensive experiments are conducted to evaluate the accuracy of our approach. The average accuracy of food type recognition by AutoDietary is 84.9%, and those to classify liquid and solid food intakes are up to 97.6% and 99.7%, respectively. A survey regarding wear comfort and functionalities of AutoDietary is conducted. The results show that the current design of AutoDietary is acceptable to most users for daily use.

APPENDIX
A SURVEY TO INVESTIGATE THE USER EXPERIENCE OF AUTODIETARY

We conducted a survey involving 53 participants to investigate the user experience of AutoDietary. The survey contains 12 questions (listed in Fig. 10), with Question 1-4 focusing on wear comfort, Question 5-9 focusing on functionalities, and the rest on other aspects.

Results on the two main aspects, i.e., wear comfort and functionalities, are presented in Fig. 11(a) and Fig. 11(b), respectively. A higher score indicates higher satisfaction. It can be seen that the current design of AutoDietary is acceptable to most users for daily use. The results for other aspect (Question 10-12) are listed in Table IX in terms of the number of votes for each class. These results exhibit user's preference or expectations on AutoDietary, which can help us to further improve our system.

REFERENCES

[1] *Obesity and Overweight: What are Overweight and Obesity. Fact Sheet*, World Health Organization, Geneva, Switzerland, 2006, (311).
[2] E. S. Sazonov and S. Schuckers, "The energetics of obesity: A review: Monitoring energy intake and energy expenditure in humans," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 1, pp. 31–35, Jan./Feb. 2010.
[3] L. E. Burke *et al.*, "Self-monitoring dietary intake: Current and future practices," *J. Renal Nutrition*, vol. 15, no. 3, pp. 281–290, 2005.
[4] J. Beh, D. K. Han, R. Durasiwami, and H. Ko, "Hidden Markov model on a unit hypersphere space for gesture trajectory recognition," *Pattern Recognit. Lett.*, vol. 36, pp. 144–153, Jan. 2014.
[5] S. R. M. S. Baki, Z. M. A. Mohd, I. M. Yassin, A. H. Hasliza, and A. Zabidi, "Non-destructive classification of watermelon ripeness using mel-frequency cepstrum coefficients and multilayer perceptrons," in *Proc. IJCNN*, Jul. 2010, pp. 1–6.
[6] O. Amft and G. Troster, "On-body sensing solutions for automatic dietary monitoring," *IEEE Pervasive Comput.*, vol. 8, no. 2, pp. 62–70, Apr./Jun. 2009.
[7] E. Sazonov *et al.*, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiol. Meas.*, vol. 29, no. 5, pp. 525–531, 2008.
[8] O. Amft, "A wearable earpad sensor for chewing monitoring," *IEEE Sensors*, vol. 1, no. 4, pp. 222–227, Nov. 2010.
[9] S. Päßler, M. Wolff, and W.-J. Fischer, "Food intake monitoring: An acoustical approach to automated food intake activity detection and classification of consumed food," *Physiol. Meas.*, vol. 33, no. 6, pp. 1073–1093, 2012.
[10] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
[11] K.-H. Chang *et al.*, "The diet-aware dining table: Observing dietary behaviors over a tabletop surface," in *Pervasive Computing*. Berlin, Germany: Springer-Verlag, 2006, pp. 366–382.
[12] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, "Detecting periods of eating during free-living by tracking wrist motion," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1253–1260, Jul. 2013.
[13] J. Lester, D. Tan, and S. Patel, "Automatic classification of daily fluid intake," in *Proc. IEEE 4th Int. Conf. Pervas. Comput. Technol. Healthcare (PervasiveHealth)*, Mar. 2010, pp. 1–8.
[14] K. Sato and T. Nakashima, "Human adult deglutition during sleep," *Ann. Otol., Rhinol. Laryngol.*, vol. 115, no. 5, pp. 334–339, 2006.
[15] J. Gao, A. G. Hauptmann, A. Bharucha, and H. D. Wactlar, "Dining activity analysis using a hidden Markov model," in *Proc. 17th ICPR*, vol. 2, 2004, pp. 915–918.
[16] W. Wu and J. Yang, "Fast food recognition from videos of eating for calorie estimation," in *Proc. IEEE ICME*, Jun./Jul. 2009, pp. 1210–1213.

[17] B. Zhou *et al.*, "Smart table surface: A novel approach to pervasive dining monitoring," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, Mar. 2015, pp. 155–162.

[18] Texas Instruments. *TI Homepage: LM358*. [Online]. Available: http://www.ti.com/product/lm358, accessed Sep. 24, 2014.

[19] Texas Instruments. *TI Homepage: TLV25431*. [Online]. Available: http://www.ti.com/product/tlv2541, accessed Sep. 24, 2014.

[20] Texas Instruments. *TI Homepage: MSP430*. [Online]. Available: http://www.ti.com.cn/product/MSP430F169, accessed Sep. 30, 2014.

[21] X. Zhou *et al.*, "HMM-based acoustic event detection with adaBoost feature selection," in *Multimodal Technologies for Perception of Humans*. Berlin, Germany: Springer-Verlag, 2007.

[22] C. Zieger, "An HMM based system for acoustic event detection," in *Multimodal Technologies Perception Humans*. Berlin, Germany: Springer-Verlag, 2008, pp. 338–344.

[23] A. Temko, R. Malkin, C. Zieger, D. Macho, C. Nadeu, and M. Omologo, "Acoustic event detection and classification in smart-room environments: Evaluation of CHIL project systems," *Cough*, vol. 65, no. 48, p. 5, 2006.

[24] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, Berlin, Germany: Springer-Verlag, 2004, pp. 1–17.

[25] T. Huynh and B. Schiele, "Analyzing features for activity recognition," in *Proc. Joint Conf. Smart Objects Ambient Intell., Innov. Context-Aware Services*, 2005, pp. 159–163.

[26] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[27] A. Yadollahi and Z. Moussavi, "Feature selection for swallowing sounds classification," in *Proc. 29th Annu. Int. Conf. IEEE EMBC*, Aug. 2007, pp. 3172–3175.

[28] E. S. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. 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, Mar. 2010.

[29] A. Metin, *Nonlinear Biomedical Signal Processing Volume II: Dynamic Analysis and Modeling*. New York, NY, USA: Wiley, 2000, pp. 83–92.

[30] M. E. Cohen, D. L. Hudson, and P. C. Deedwania, "Applying continuous chaotic modeling to cardiac signal analysis," *IEEE Eng. Med. Biol. Mag.*, vol. 15, no. 5, pp. 97–102, Sep./Oct. 1996.

[31] U. Kumar *et al.*, "Mining land cover information using multiplayer perception and decision tree from MODIS data [J]" *Indian Soc. Remote Sens.*, vol. 38, no. 4, pp. 592–602, Dec. 2010.

[32] C. Chien and G. J. Pottie, "A universal hybrid decision tree classifier design for human activity classification," in *Proc. 34th Annu. Int. Conf. IEEE EMBS*, San Diego, CA, USA, Aug./Sep. 2012, pp. 1065–1068.

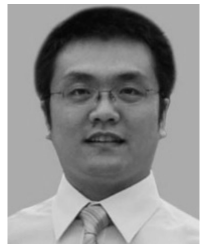
[33] Y. Sakakibara, K. Misue, and T. Koshiba, "Text classification and keyword extraction by learning decision trees," in *Proc. 9th Conf. Artif. Intell. Appl.*, Mar. 1993, p. 466.

[34] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Tree*. London, U.K.: Chapman & Hall, 1984.

[35] M. M. Bundele and R. Banerjee, "ROC analysis of a fatigue classifier for vehicular drivers," in *Proc. 5th IEEE Int. Conf. Intell. Syst.*, Jul. 2010, pp. 296–301.

[36] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Proc. 14th Int. Conf. Mach. Learn.*, 1997, pp. 412–420.

Yin Bi received the B.S. degree in electrical engineering and the M.S. degree in signal and information processing from Northeastern University, China, in 2013 and 2015, under the supervision of Prof. W. Yi. His research interests include signal processing, pattern recognition, embedded and wearable sensor systems, and mobile healthcare.



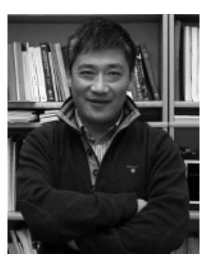
Mingsong Lv received the B.S., M.S., and Ph.D. degrees from Northeastern University, China, in 2002, 2005, and 2010, respectively. He is currently an Associate Professor with Northeastern University. His research interests include program analysis, multi-core embedded systems, and new energy vehicles. He received the Design Automation, Test in Europe Best Paper Award in 2013 (C. Song).

Chen Song received the B.S. degree in optic science and engineering from Fudan University, and the M.S. degree in electrical engineering from the State University of New York at Buffalo, where he is currently pursuing the Ph.D. degree with the Department of Computer Science, under the direction of Prof. W. Xu. His current research focuses on human–computer interaction, compressed sensing, and health-care and security in biomedical applications.

Wenyao Xu received the Ph.D. degree from the Electrical Engineering Department, University of California, Los Angeles, CA, USA, in 2013. He is currently an Assistant Professor with the Computer Science and Engineering Department, State University of New York at Buffalo, New York, NY, USA. His current research foci include embedded system, computer architecture, wireless health, low-power technologies, and their applications in biomedicine, healthcare, and security. He owned five licensed U.S. and international patents, and has authored over 60 peer-reviewed journal and conference papers.

Dr. Xu received the Best Paper Award at the IEEE Conference on Implantable and Wearable Body Sensor Networks in 2013, and the Best Demonstration Award of the ACM Wireless Health Conference in 2011.

Nan Guan (M'12) received the bachelor's and master's degrees from Northeastern University, China, and the Ph.D. degree from Uppsala University, Sweden. He is currently a Full Professor with Northeastern University. His research interests include cyber-physical systems, real-time scheduling theory, and program analysis. He received the European Design Automation Association Outstanding Dissertation Award in 2014, the IEEE Real-Time Systems Symposium Best Paper Award in 2009, the Design Automation, Test in Europe Best Paper Award in 2013.



Wang Yi (F'15) received the Ph.D. degree in computer science from the Chalmers University of Technology in 1991. He is currently a Full Professor with Northeastern University, China. He also holds the Chair of Embedded Systems with Uppsala University, Sweden. He received the CAV Award 2013 for the development of UPPAAL. His current interests include models, algorithms and software tools for modeling and verification, timing analysis, real-time scheduling, and their application to the design of embedded and real-time and mixed-criticality systems.

He received the Outstanding Paper Award at ECRTS 2012 and Best Paper Awards at RTSS 2009, DATE 2013, and ECRTS 2015 with P. Ekberg, N. Guan, and M. Stigge. He has been an Editor of several journals, including the IEEE TRANSACTIONS ON COMPUTERS, the PC Chair of TACAS 2001, FORMATS 2005, EMSOFT 2006, HSCC 2011, and LCTES 2012, and Track/Topic Chair of RTSS 2008 and DATE 2012-2014. He is also serving on the Steering Committee of FORMATS, EMSOFT (Co-Chair), LCTES, and ESWEEK.