Eating Detection and Chews Counting through Sensing Mastication Muscle Contraction

Shuangquan Wang*, Gang Zhou*, Yongsen Ma*, Lisha Hu†, Zhenyu Chen‡, Yiqiang Chen§, Hongyang Zhao*, Woosub Jung*

*Computer Science Department, College of William and Mary

†Institute of Information Technology, Hebei University of Economics and Business

‡China Electric Power Research Institute

§Institute of Computing Technology, Chinese Academy of Sciences

Email: swang10@email.wm.edu, {gzhou, yma}@cs.wm.edu, hulisha@heuet.edu.cn, chenzhenyu@epri.sgcc.com.cn, yqchen@ict.ac.cn, hyzhao@cs.wm.edu, wjung01@email.wm.edu

***Abstract*—Unhealthy dietary habits (eating disorder, eating too fast, excessive energy intake, etc.) are major causes of some chronic diseases such as obesity, digestive system disease and diabetes. Dietary monitoring is necessary and important for patients to break and change their unhealthy diet and eating habits. Existing audio or video based methods are often invasive and bring privacy concerns. Motion sensor based related works are popular for eating detection, but cannot count chews. This paper presents the first effort in using motion sensor to sense mastication muscle contraction for continuous dietary monitoring. We observe that during eating the mastication muscles contract and hence bulge in some degree. In addition, the bulge of the mastication muscles has the same frequency as chewing. These observations motivate us to detect eating activity and count chews through attaching a triaxial accelerometer on the temporalis. The proposed method does not record any personal privacy information (audio, video, etc.). The accelerometer is embedded into a headband. Therefore, it is comparatively noninvasive for the user's daily living. Experiments are conducted and the results are promising. For eating activity detection, the average *accuracy* and *F-score* of five classifiers are 94.4% and 87.2%, respectively, in 10-fold cross validation test using only 5 seconds of acceleration data. For chews counting, the average error rate of four users is 12.2%.**

I. Introduction

According to statistics of the World Health Organization (WHO), nowadays chronic diseases have become one of the most serious threats to human health [1]. As major causes of chronic diseases, unhealthy dietary habits lead to prevalence of obesity, digestive system disease and diabetes. A national health and nutrition examination survey of 9120 participants shows that "in 2011-2012, 16.9% of youth (2 to 19-year-olds) and 34.9% of adults aged 20 years or older in the United States are obese [2]." The rapid increase of chronic diseases in recent years forces people to pay more attentions on dietary monitoring. This helps people optimize their diet composition and change their unhealthy eating habits. In the eating process, chewing is one essential step. According to the report of Daily Mail, "people who thoughtfully chew their food and don't rush mealtimes not only avoid indigestion - they could be preventing diabetes as well [3]." However, chewing is often overlooked.

In recent years, several methods have been proposed to recognize a subject's eating activity and count chews. Self-report diary based methods [4] are simple and straightforward but tedious and inaccurate. Audio based methods need to deploy sensors in the outer ear [5] or at the throat area [6], which is invasive. Video based methods do not require a person to wear any sensor, but demand a camera to capture mouth movement [7] and hence bring privacy concerns. Motion based methods aim to recognize hand motions [8, 9] or head vibrations [10] and indirectly infer eating behaviors. But they cannot count chews. In addition, some other sensors are also utilized, such as physiological sensors (e.g. electromyography sensor) and physical sensors (e.g. piezoelectric strain gauge sensor in [11]). However, these sensors need to be tightly adhered to skin, which is invasive and discomforting.

How to detect a subject's eating activity and count the number of chews in an accurate and noninvasive way? To answer this question, we investigate the principle of eating activity and are inspired by following observations:

- Eating activity is activated through a collaborative effort of four mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid and the temporalis. The first three muscles are near mouth cavity and hence not convenient for sensor deployment. The temporalis is a broad, fan-shaped muscle located at the side of the skull and in front of the ear [12]. This is the area where people often wear a headband or hat. Therefore, the temporalis is suitable for noninvasive sensing of eating activity.
- During eating, the temporalis contracts to elevate the mandible, which results in the bulge of this muscle. We are hence motivated to recognize the eating activity through detecting the temporalis contractions and bulges. This is done through embedding an acceleromter into a headband and attaching the accelerometer on the temporalis.
- The bulge of the temporalis has the same frequency as chewing. Thus, the number of chews can also be counted through recognizing the frequency of periodic muscle bulges.

In this paper, we propose to detect eating activity and count chews simultaneously with a triaxial acceleromter. We embed the accelerometer into a headhand to sense the temporalis contractions and bulges, which is noninvasive and convenient. Compared with existing audio or video based methods, our method only records the acceleration data. Therefore, our method has less privacy concerns.

This paper makes the following main contributions:

- We propose to detect eating activity and count chews through attaching a triaxial accelerometer on the temporalis. To our best knowledge, this is the first work on motion sensor based sensing of mastication muscle contraction for continuous dietary monitoring.
- We design and develop an eating activity detection module. It accurately differentiates eating activity from six other daily activities (reading/speaking, sitting, walking, drinking, coughing and standing) using only 5 seconds of acceleration data.
- We design and develop a chews counting module. It identifies the primary periodicity of highly noisy acceleration data and accurately count the number of chews.
- We evaluate the performance of the proposed method on real world dataset. Experimental results show that the average *accuracy* and *F-score* are 94.4% and 87.2%, respectively, for eating activity detection in 10-fold cross validation test. The average error rate of chews counting for four users is 12.2%.

The rest of this paper is organized as follows. Section 2 introduces background and motivation of the proposed method. Section 3 describes the system architecture and implementation of each module. Experiment and evaluation are presented in Section 4. Related work is introduced in Section 5, and Section 6 presents discussion and future work. Conclusion is drawn in Section 7.

II. Background and Motivation

In this section, we first introduce four mastication muscles. Then we present the motivation of acceleration data based eating activity detection and chews counting.

A. Four Mastication Muscles

From a physiological point of view, there are four mastication muscles: the masseter, the medial pterygoid, the lateral pterygoid and the temporalis [13]. During eating, these four muscles work together, enabling jaw open-close movements, to cut and grind the food.

The masseter is located on each side of a face. It connects the maxillae and the mandible, and primarily serves for elevating the mandible while the deep tissues help protrude the mandible forward [14]. The masseter is the most superficial muscle. It is also one of the strongest mastication muscles.

The medial pterygoid and the lateral pterygoid are located on the inner surface of the mandible. Contraction of the medial pterygoid helps elevate the mandible, and thus contributes to jaw-closing. However, the lateral pterygoid helps lower the mandible and open the jaw. It is the only mastication muscle for jaw-opening [15].

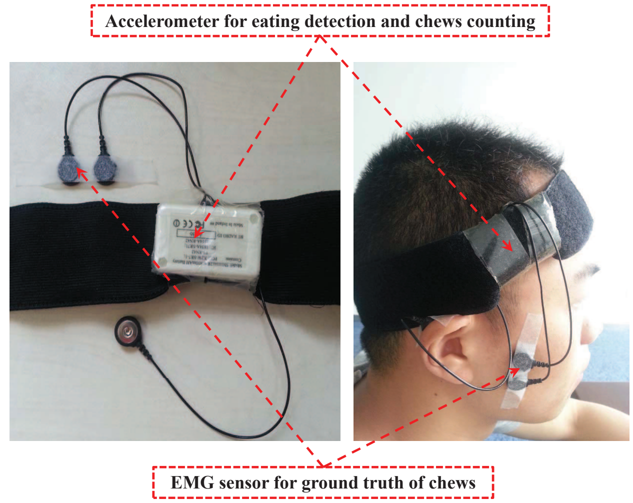

Fig. 1. Data collection device and deployment

The temporalis is a broad, fan-shaped muscle located on the side of the skull and in front of the ear [12]. It is one of three muscles that close the jaw and clench the teeth.

According to the above introduction, we see that the masseter, the medial pterygoid and the lateral pterygoid are in the face area and near mouth cavity. This area is inconvenient and invasive for sensor deployment. On the contrary, the temporalis covers a bigger area. This makes it convenient and noninvasive to attach a sensor, such as in a headband or hat.

B. Motivation

It is common knowledge that a muscle bulges when contracting. This motivates us to detect a subject's eating activity using a triaxial accelerometer attached on the temporalis. Moreover, as the temporalis contracts and bulges once for each chew, the number of chews can also be counted by recognizing the frequency of periodic muscle bulges. In this subsection, we demonstrate the potential of utilizing acceleration data to detect eating activity and count the number of chews.

In the experiment, Shimmer2r wireless sensor platform is used for data collection. Shimmer2r has an integrated accelerometer and can be connected to several types of external sensors, such as electrocardiogram (ECG), electromyography (EMG), GPS, etc [16]. We use triaxial accelerometer and EMG sensors to sample acceleration data and EMG data simultaneously. The acceleration data is utilized for eating activity detection and chews counting. The EMG data is used to obtain the ground truth of chews counting.

The sensor platform and its deployment are shown in Fig. 1. The EMG sensor has three electrodes: a positive electrode, a negative electrode and a neutral reference electrode. The sensing device is fixed in a headband using scotch tape. The user wears the headband and places the device near the right temple. The headband is elastic and adjustable. The X axis and Y axis of the accelerometer are perpendicular. These two axes form the tangent plane of the skull at the contact position. The Z axis of the accelerometer directs outward and is vertical to the X-Y plane. The accelerometer is calibrated as following: when one axis points downwards, its acceleration measurement is set to g (gravity); when it points upwards, the measurement is set to $-g$. For the EMG sensor, both positive and negative electrodes are attached on the right side of the face to detect contraction of the masseter. The neutral reference electrode is attached at the ear edge, where there is no muscle but just bone and skin. Hence, it is selected as the electrically neutral point of the body.

The sampling rates of the accelerometer and EMG sensors are the same and about 100Hz. All the collected data is wirelessly transmitted to the laptop through BlueTooth. Data of each continuous sampling process is stored in one file for post processing. The data of eating (while sitting) and six other non-eating daily activities (reading/speaking, sitting, walking, drinking, coughing and standing) are collected separately. The subject is served with multiple small pieces of watermelon. The reason for choosing watermelon will be discussed in Section 6.

Fig. 2 shows the acceleration data of eating and six other daily activities. The walking activity exhibits clear periodicity and large fluctuation range, which indicates that it could be easily differentiated. The fluctuation amplitude of the coughing activity is relatively large, but has no fixed periodicity. The other five activities, eating, speaking/reading, sitting, standing and drinking, have similar fluctuation amplitude. However, the acceleration data of the eating activity shows clear periodic pattern, especially for the Z axis data. This is because the bulge of the temporalis is in the same direction as the Z axis.

Fig. 3 shows an example window of raw Z axis acceleration data (m/sec^2) and EMG data (mVolts) during eating, indicated by a blue solid line and a red dashed line, respectively. We observe that: 1) both signals have the same periodic cycles. Each cycle corresponds to one chew; 2) the EMG signal has obvious spikes at the moments of masseter contraction. It almost equals to zero between two neighboring chews; 3) the Z axis acceleration data also has obvious increase during the muscle contraction, but there is some fluctuation between two neighboring chews.

According to the above observations, we believe that, despite many challenges, there is a high possibility of differentiating eating from other daily activities. Furthermore, it is feasible to count the number of chews only using Z axis acceleration data.

III. System Architecture and Implementation

Fig. 4 shows the system architecture of the proposed method. It contains two main modules: an eating activity detection module and a chews counting module.

The eating activity detection module includes two processes: offline training process and online testing process. They are marked with black arrows and red arrows, respectively. In the offline training process, with the offline data collected

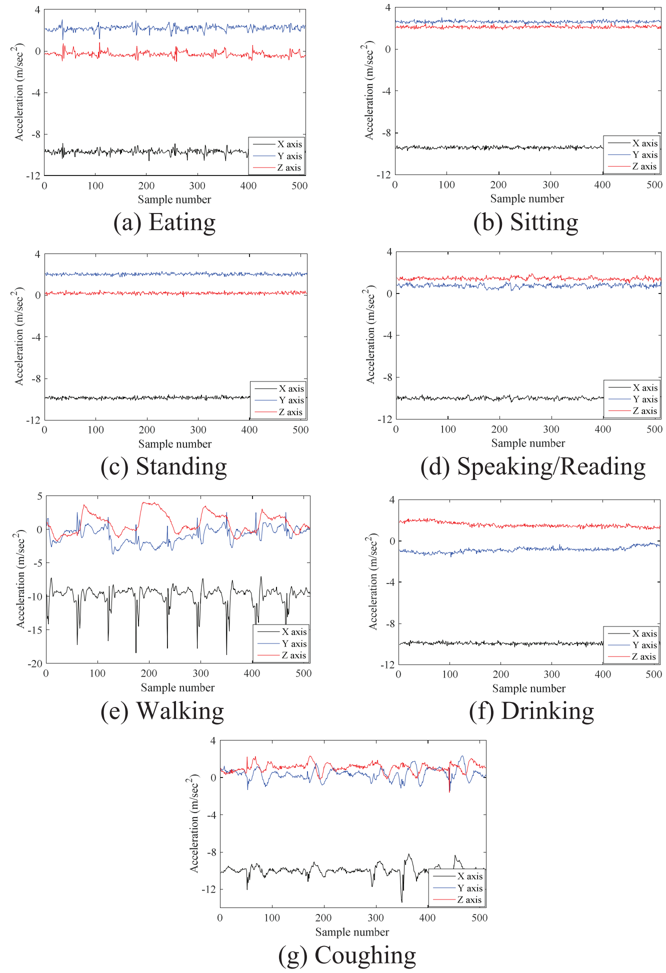

Fig. 2. Acceleration data of eating and other six activities

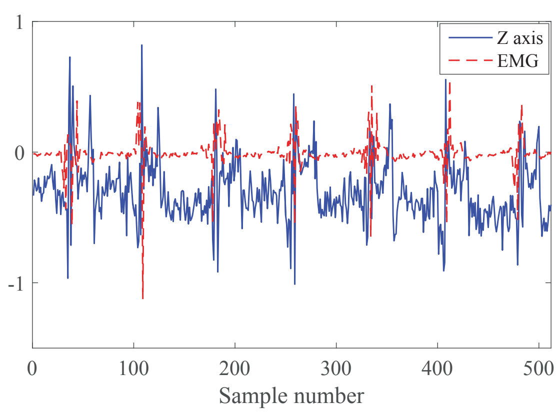

Fig. 3. Z axis acceleration and EMG data during eating

from the triaxial accelerometer, a sliding window of length L without overlap is used to segment the sensor data. For each window, the eating activity detection module first composes the acceleration data of three axes, and then extracts representative features. By combining normalized feature vectors with corresponding class labels, the training dataset is built to train

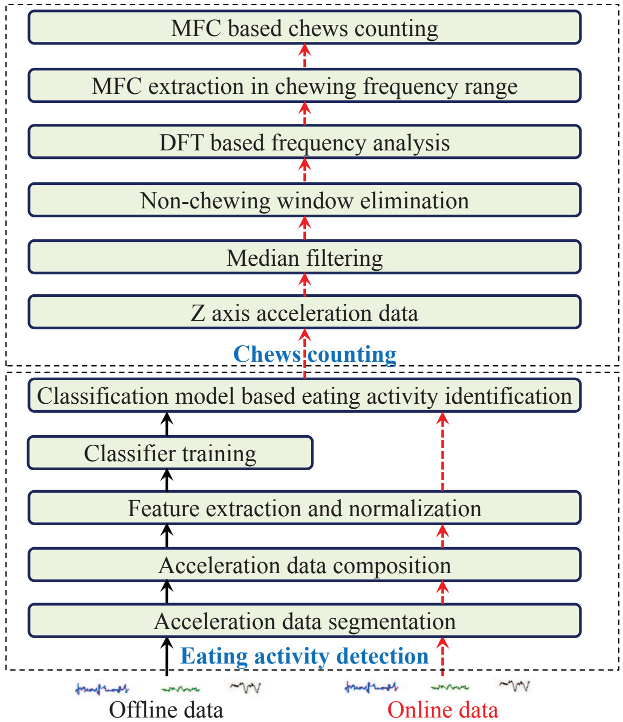

Fig. 4. System architecture of the proposed method

a two-class classifier for online recognition. The online testing process has the same operations of data segmentation, data composition, feature extraction and normalization with those of the offline training process. After that, unlabeled feature vectors are fed to the trained classification model. Then, the eating activity detection results are obtained.

If one window is identified as eating activity, the raw Z axis acceleration data is used to count the number of chews in this window. The Z axis acceleration data is first filtered using median filter to remove some sampling noise. Then, the non-chewing window is eliminated through checking the difference of acceleration magnitude. For a chewing window, discrete Fourier transform (DFT) based frequency analysis is applied, and the maximum frequency component (MFC) in the chewing frequency range is extracted. Finally, the number of chews in this window is estimated based on the frequency corresponding to MFC and the time length of the sliding window.

We introduce the detailed implementation of the above two modules as following.

A. Eating Activity Detection

After segmenting the triaxial acceleration data with a sliding window of length L without overlap, we compose the ith sensor readings of three axes, a_x^i, a_y^i and a_z^i into one scalar acceleration a_i:

$$a_i = \sqrt{(a_x^i)^2 + (a_y^i)^2 + (a_z^i)^2} \tag{1}$$

where $i = 1, \ldots, L$.

Based on the composed data, four groups of features are extracted to build the feature vector for this window of data. The first group consists of six time domain features: the maximum, the minimum, the 1st quartile, the 2nd quartile, the 3rd quartile and the number of mean cross.

The second group consists of four amplitude statistics features extracted from the composed window data. They are defined [17] as:

$$Amplitude: \mu_{amp} = \frac{1}{L}\sum_{i=1}^{L} a_i \tag{2}$$

$$Standard\ deviation: \sigma_{amp} = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(a_i - \mu_{amp})^2} \tag{3}$$

$$Skewness: \gamma_{amp} = \frac{1}{L}\sum_{i=1}^{L}\left(\frac{a_i - \mu_{amp}}{\sigma_{amp}}\right)^3 \tag{4}$$

$$Kurtosis: \beta_{amp} = \frac{1}{L}\sum_{i=1}^{L}\left(\frac{a_i - \mu_{amp}}{\sigma_{amp}}\right)^4 - 3 \tag{5}$$

The third group consists of four amplitude statistics features extracted from single-sided amplitude spectrum (without direct current component) after Fourier transform [18]. These features can be computed using the above four formulas after replacing L and a_i with $\frac{L}{2}$ and s_i, respectively. Here, s_i means the ith component of single-sided amplitude spectrum.

According to [19], chewing activity mainly occurs in the range between 0.94 Hz (5th percentile) and 2.17 Hz (95th percentile). In this paper, we define the chewing frequency range as 0.5 Hz to 2.5 Hz. Then, the single-sided amplitude spectrum (without direct current component) can be partitioned into three bands: $(0, 0.5)$ Hz, $[0.5, 2.5]$ Hz and $(2.5, SF/2]$ Hz. SF means the sampling frequency of accelerometer. Three features are extracted from each band to form the fourth group of features. They are the MFC, the location (i.e. the index) of the MFC, and the spectral energy. The spectral energy is defined as the sum of squared spectrum components in each band.

In total, 23 features are extracted. To eliminate the scaling effects among different features, all the features are normalized using the z-score normalization algorithm [20].

The eating activity detection is formulated as a two-class classification problem. The positive class corresponds to eating activity, while the negative class corresponds to other daily activities, such as speaking/reading, sitting, standing, walking, drinking, coughing, etc. Five commonly used classification algorithms are compared: decision tree (DT), nearest neighbor (NN), multi-layer perceptron (MLP), support vector machine (SVM) and weighted support vector machine (WSVM).

DT algorithm builds a pattern classifier from a labeled training dataset using a divide-and-conquer approach. It recursively selects the attribute that is used to partition the training dataset into subsets until each leaf node in the tree has a uniform class membership [21]. NN algorithm is an instance-based learning method. It only stores the training samples but does not generate a specific classification model. During classification, the distances between the test sample and all training samples are calculated. The test sample is assigned the same class label as its nearest neighbor. MLP algorithm is a feedforward artificial neural network model. It maps a set of inputs onto a set of appropriate outputs. It uses a supervised technique called backpropagation to train the network and obtain the parameters [22]. SVM algorithm is based on the foundation of statistical learning theory. It gains promising empirical performance in the fields of nonlinear and high dimensional pattern recognition [21, 23].

The above four algorithms have different rationales and model structures. Comparison of their recognition results should demonstrate performance of the proposed method on eating activity detection in a comprehensive and unbiased way. WSVM can deal with the uneven class size problem of SVM by assigning larger weights to classes with fewer samples [24]. Therefore, it is also included for comparison.

B. Chews Counting

The Z axis acceleration data is used to count the number of chews in one sliding window. The chews counting module contains the following five steps:

Step 1: Median filtering. We first use a 7th-order one-dimensional median filter [25] to remove the sampling noise in the acceleration data. The median filter runs through the sliding window sample by sample, and replaces each sample with the median of neighboring samples [26].

Step 2: Non-chewing window elimination. In the eating activity detection module, one window is identified as eating activity or non-eating activity. An eating window may contain not only chewing, but also food intake and swallowing. In section II.B, we indicate that the Z axis directs outward and is vertical to the tangent plane of the skull at the contact position. Therefore, when the user bows down his head and feeds food into his mouth, the Z axis directs downwards in some degree. Accordingly, there is a positive decomposition of gravity on the Z axis. The gravity decomposition generates a large convex peak of the Z axis acceleration data, as shown in Fig. 5. Normally, the magnitude of the convex peak is much larger than the acceleration variation during chewing. For simplicity, we calculate the difference between the maximum acceleration and the minimum acceleration in one window. If the difference is larger than a predefined threshold, $MagDiff$, this window is considered as a non-chewing window and eliminated.

Step 3: DFT based frequency analysis. For one chewing window, we need to count the number of temporalis bulges. The most straightforward method is to count the peaks of Z axis acceleration data. However, as shown by the red line in Fig. 6, the Z axis acceleration data is very noisy even after median filtering. There are lots of false peaks caused by the vibration of the skull during chewing. One observation is that the chewing frequency is consistent with the primary periodicity of the acceleration data. We use the MFC to reconstruct the acceleration data only with primary periodicity. The reconstructed data is shown as the blue line in Fig. 6. Therefore, for chews counting, we propose to utilize DFT to transform the acceleration data from time domain into frequency domain.

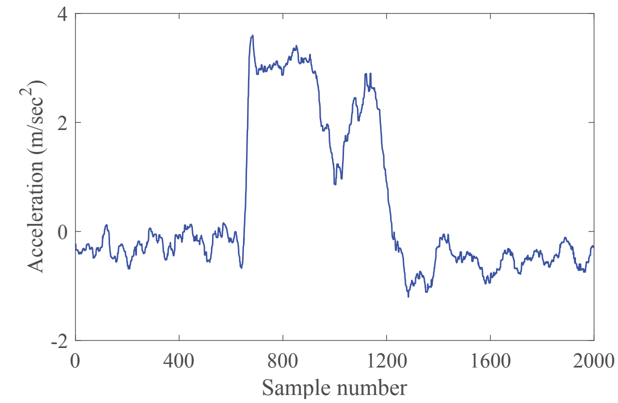

Fig. 5. Convex peak during food intake

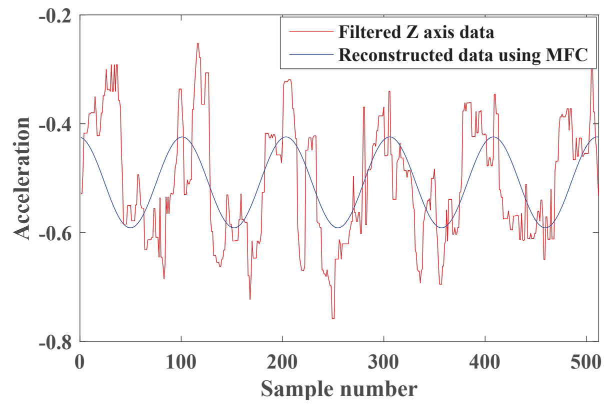

Fig. 6. Noisy acceleration data and reconstructed data using MFC

Step 4: MFC extraction in chewing frequency range. After DFT based frequency analysis, we extract the MFC in the chewing frequency range, i.e. $[0.5, 2.5]$ Hz.

Step 5: MFC based chews counting. We take the frequency corresponding to MFC, $freq_{MFC}$, as the approximate chewing frequency. Then we estimate the number of chews in one window by multiplying $freq_{MFC}$ with the time length of the window. The time length can be obtained through dividing the window length by the sampling rate.

IV. Experiment and Evaluation

In this section, we introduce the experimental evaluation on eating activity detection and chews counting.

A. Data Collection and Ground Truth

We recruit four volunteers to collect the experimental dataset [27] of seven daily activities, including eating, reading/speaking, standing, sitting, walking, drinking and coughing. For eating, reading/speaking, standing, sitting and walking activities, one volunteer performs each of them for 6 to 9 minutes. For drinking and coughing activities, one volunteer performs each of them about 30 seconds. For reading/speaking and walking, the volunteers are asked to perform in three different speeds (slow, moderate and fast), and each speed for 2 to 3 minutes. The acceleration data of three axes and the EMG data are sampled simultaneously. In total, about 150 minutes of data is collected.

These activities are manually labeled during data collection. To serve as ground truth of chews counting, the EMG data is manually identified and counted to obtain the total number of chews for each volunteer.

B. Evaluation of Eating Activity Detection

Three tests are conducted to evaluate the eating detection performance of the proposed method. 1) Cross validation test (CVT). CVT combines all subjects' samples to form the dataset. It uses the cross validation method to evaluate the general eating detection accuracy on multiple subjects; 2) Self test (ST). ST only uses the samples of the subject himself/herself to form the dataset. For the ST evaluation of each subject, the same cross validation method as above is used; 3) Leave-one-subject-out test (LOSOT). LOSOT uses the samples of all subjects except one to form the dataset and train the classification model accordingly. Then the model is tested using the samples of the excluded subject. LOSOT shows how generic the detection model is for unknown subjects.

Weka toolkit [21] is used for classifier training and testing. For DT, the J48 algorithm is used. For SVM and WSVM, the LibSVM wrapper for Weka [28] is used. We adopt the default parameters for all classifiers in the following experiments. For cross validation, the fold number is set to 10. Because the samples of negative class are about three times that of positive class, the weights of WSVM are set to 3 for positive class and 1 for negative class.

Four evaluation metrics are used to quantify the classification performance. They are *accuracy*, *precision*, *recall* and *F-score*, which are defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$F\text{-}score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{9}$$

where *TP* denotes true positive, *TN* denotes true negative, *FP* denotes false positive, and *FN* denotes false negative.

1) Cross validation test (CVT): Fig. 7 shows the *accuracy*, *precision*, *recall* and *F-score* of CVT for different window lengths: 256, 512 and 1024. From Fig. 7 we see that: (1) For all these five classification models, their *accuracy* is larger than 90%, and their *F-score* is larger than 80%. Table I and Table II give the *accuracy* and *F-score* of the five classifiers.

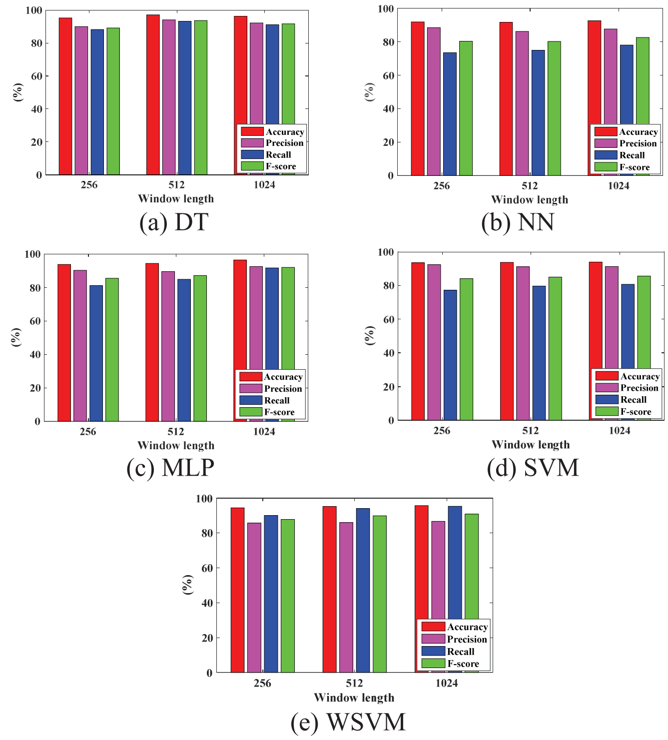

Fig. 7. CVT results of five classifiers

The average *accuracy* of the five classifiers are 93.8% (window length of 256), 94.4% (window length of 512) and 95.0% (window length of 1024), respectively. The average *F-score* of the five classifiers are 85.4% (window length of 256), 87.2% (window length of 512) and 88.5% (window length of 1024), respectively. This indicates that the proposed eating activity detection method is able to accurately distinguish eating activity from the six other daily activities; (2) The *accuracy* outperforms *precision*, *recall* and *F-score* in all these cases. Through comparing equations (6), (7) and (8), we know that the *accuracy* metric not only takes *TP* and *FP* into consideration, but also *TN* and *FN*. High *accuracy* indicates that *TN* is much larger than *FN* and *FP*. That is to say, all the models can identify most negative samples; (3) For DT, NN, MLP and SVM, their *precision* is better than *recall*. This is because that, according to equations (7) and (8), *FP* is smaller than *FN*. This implies that all these four models misclassify more positive samples as negative class than negative samples as positive class. In other words, these models are biased to negative class. Through assigning a larger weight to the positive class, WSVM reverses the bias and obtains higher *recall* than *precision*; (4) As to the five classifiers, DT, MLP and WSVM outperform NN and SVM. Specifically, DT performs best for window length of 256 and 512. MLP performs best for window length of 1024; (5) The classification performance improves with the increase of window length, but the improvement is only two to three percent. As longer window length causes larger time delay, in all the following experiments, the window length is set to 512 to balance the accuracy and delay.

TABLE I
THE *accuracy* OF FIVE CLASSIFIERS IN CVT

Win.	DT	NN	MLP	SVM	WSVM	Average
256	95.2%	91.9%	93.8%	93.5%	94.4%	**93.8%**
512	97.1%	91.7%	94.4%	93.7%	95.2%	**94.4%**
1024	96.2%	92.6%	96.5%	93.9%	95.7%	**95.0%**

TABLE II
THE *F-score* OF FIVE CLASSIFIERS IN CVT

Win.	DT	NN	MLP	SVM	WSVM	Average
256	89.1%	80.3%	85.5%	84.1%	87.8%	**85.4%**
512	93.6%	80.2%	87.2%	85.0%	89.8%	**87.2%**
1024	91.6%	82.6%	92.1%	85.6%	90.8%	**88.5%**

2) Self test (ST): Fig. 8 depicts the ST results of these four users. For user 1 and user 2, the ST performances of the five classifiers on almost all these four metrics are highly accurate. Both the *accuracy* and *F-score* are higher than those of CVT. For user 3, the *F-score* of SVM is a little low. For user 4, while the *accuracy* of all these classifiers is above 90%, the *F-score* of NN and SVM is lower than 80%. However, DT and MLP still perform quite good for user 4. The *accuracy* of DT and MLP is larger than 95%, and the *F-score* of DT and MLP is above 90%. We believe that the performance difference between different classifiers is mainly because that we adopt the default parameters for all classifiers. The performance of these classifiers could be improved after parameter optimization.

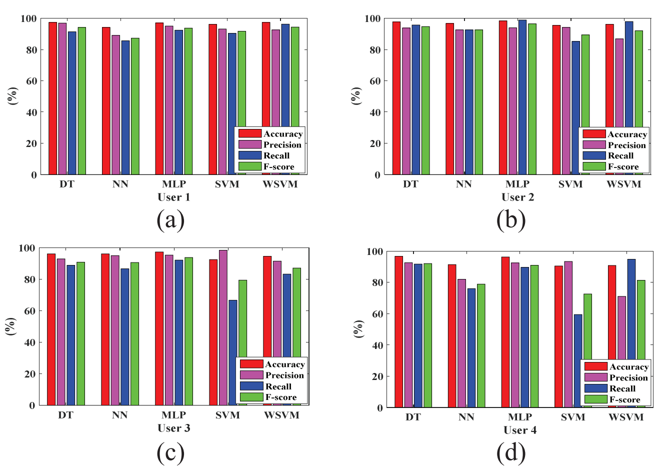

Fig. 8. ST results of four users

3) Leave-one-subject-out test (LOSOT): Fig. 9 presents the LOSOT results of four users. Comparatively, the performance of LOSOT falls below that of CVT and ST. This is reasonable as the data of testing user is not included in the training dataset. Table III shows the average *accuracy* and *F-score* of the five classifiers for these four users. For user 1, user 2 and user 3, the average *accuracy* is between 89.8% and 93.4%, and the average *F-score* is between 76.6% and 85.1%. These are still good. For user 4, the average *accuracy* is 81.6%, but the average *F-score* is only 55.4%. Why is the detection performance of user 4 lower than that of the other three users? We believe that it is because of sensor position bias, which causes larger variance of the sampled acceleration data for user 4.

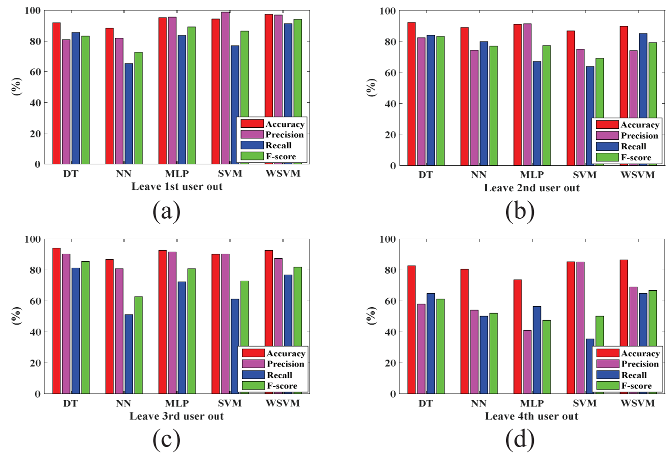

Fig. 9. LOSOT results of four users

TABLE III
THE AVERAGE *accuracy* AND *F-score* IN LOSOT

	User 1	User 2	User 3	User 4
Avg. *accuracy*	93.4%	89.8%	91.1%	81.6%
Avg. *F-score*	85.1%	77.1%	76.6%	55.4%

C. Evaluation of Chews Counting

We evaluate the chews counting accuracy for each user in the following experiments. For non-chewing window elimination, the difference threshold, $MagDiff$, is set to 3. To describe the chews counting accuracy, the following detection error rate is used:

$$Error\ rate = \frac{|Detection - Ground\ truth|}{Ground\ truth} \times 100\% \quad (10)$$

Table IV describes the ground truth in terms of the number of chews, chewing time and chewing frequency of these four users. Compared with the chewing frequency range reported in reference [19], the chewing frequencies in Table IV are a little low. This is because we do not exclude the time spent in biting and swallowing.

Table V depicts the chews counting results for four users. The error rates are 9.9%, 21.8%, 4.0% and 13.2%. Though the error rate for user 2 is a little high, the average error rate of four users is about 12.2%.

In Step 2 of chews counting module, we drop the whole window if the acceleration magnitude difference is larger than the predefined threshold. The dropped window may

TABLE IV
GROUND TRUTH OF CHEWING OF FOUR USERS

	User 1	User 2	User 3	User 4
Number of chews	473	596	323	380
Chewing time (Sec.)	532	481	492	461
Chewing frequency	0.9 Hz	1.2 Hz	0.7 Hz	0.8 Hz

TABLE V
CHEWS COUNTING RESULTS OF FOUR USERS

	User 1	User 2	User 3	User 4
Chews counting	520	466	310	330
Error rate	9.9%	21.8%	4.0%	13.2%

contain a few chews. Therefore, in most cases, the chews counting results are underestimated, as we can see from the results of user 2, 3 and 4. The best solution is to design a segmentation algorithm to extract whole chewing segments for chews counting. We leave this as our future work.

For the threshold of acceleration magnitude difference, we use a fixed value for all the users. Considering the user difference and sensor location variance, a user-dependent and online adaptive threshold should be better.

In summary, the above experiments demonstrate that the proposed method is able to accurately detect users' eating activity and count the number of chews.

V. RELATED WORK

A self-report diary [29] is often used in dietary monitoring. Though people may roughly record the food type and amount under request, they often tend to miscalculate and underreport the food consumed [30]. Furthermore, few people are willing to estimate the eating speed or count the number of chews during eating. Recently, some progress has been made on automatic dietary monitoring, including audio based methods, video based methods and motion based methods.

An audio based method uses a wearable microphone to detect the sound during eating. Reference [6] uses a modified Bluetooth headset with embedded microphone to collect sounds in a user's throat area. Time domain features, frequency domain features and cepstral features are extracted from the recorded sounds to train the classification model. The F-measure accuracy reaches 79.5% and 71.5% for laboratory study (12 activities) and small-scale in-the-wild study (4 activities), respectively. Reference [5] places the microphone inside the ear canal to differentiate chewing from speech and silence. At the same time, this method can also differentiate several types of food, such as potato chips, apple, mixed lettuce and pasta, based on the difference of their chewing sounds. Even though the above related works demonstrate the validness of the audio based method, the sensor deployment is invasive. A user may not be willing to wear a microphone near the throat or in the ear during eating. Besides, audio recording raises potential privacy concerns.

An Active Appearance Model (AAM) is utilized in [7] to track a subject's face and detect chewing activity from surveillance video. This is based on the observation that variations in AAM parameters have distinct periodicity during chewing. The experimental results demonstrate a cross-validated percentage agreement of 93.0%. The video based method needs no on-body sensor, and the video sequence can be acquired without any user intervention. However, this method brings many privacy concerns. Its accuracy is also affected by environmental lighting changes and face occlusion.

In recent years, the popularity of wearable devices (smart glasses, smartwatch, smart wristband, etc.) has made motion based dietary monitoring possible. Reference [31] combines accelerometers from a smartwatch and Google glass to recognize a user's eating activity. Reference [32] only uses a glasses mounted accelerometer to distinguish chewing from non-chewing activities. Similarly, reference [33] uses motion sensors on Google glass to detect head movement and infer eating activity. Reference [10] integrates an EMG sensor and vibration sensor into 3D-printed eyeglasses for detecting chewing and identifying food categories. Reference [8] uses a watch-like device, which is embedded with a micro-electro-mechanical gyroscope, to track wrist motions and detect food intake. Moreover, reference [34] uses a smartwatch to detect eating activity. Reference [35] designs a sensor-embedded digital fork, Sensing Fork, to sense a child's eating behavior. Furthermore, a mobile game named Hungry Panda is developed to encourage the child to eat diverse foods during mealtime. However, the above methods are mainly for eating or chewing activities detection. Chews counting is not included. Li et al [36] propose a novel method that embeds a small accelerometer inside artificial teeth to capture unique motion characters during chewing, drinking, speaking and coughing. This method is too invasive to be widely accepted. In addition, safety issues and frequent battery changes also need to be addressed. In [27], we published a poster paper to present our preliminary idea, by which this technology had not been fully developed yet. In that poster publication, we used a single axis accelerometer to detect eating activity and count chews. However, this method requires much larger sliding windows. Besides, further experiments show that its eating detection accuracy is very low for the leave-one-subject-out test. We believe that it is because one axis acceleration data is not enough to distinguish motion characters between eating and other activities, and the extracted features are not representative enough.

In addition to audio, video and motion based methods, there are some other ways for dietary monitoring. For example, references [11] and [37] deploy a piezoelectric sensor below the ear to capture the movement of the lower jaw and detect chewing rate. Reference [38] places a piezoelectric sensor on the temporalis to detect the chewing bouts. However, the piezoelectric sensor needs to be attached on the skin tightly. This is invasive and unfriendly for users.

VI. DISCUSSION AND FUTURE WORK

We chose multiple pieces of watermelon as the food in all evaluation experiments. Watermelon is one of the softest foods. Thus, the user chews it with little effort. According to commonsense, the more strength the mastication muscles use, the greater the muscle bulge is. Therefore, if eating watermelon can be accurately recognized by the proposed method, it is reasonable to expect that eating harder food could also be accurately recognized. Besides, it may be possible for the proposed method to recognize the food types through detecting the chewing strength. We leave this for future work.

For the ground truth in terms of the number of chews, the EMG data is used as the reference to identify and count each chew. The EMG signal during food intake is ignored. At the same time, as shown in the green boxes in Fig. 10, occasionally the EMG signal may be less obvious and hard to identify. Thus, the ground truth may not be perfectly obtained, but the error is very small according to our observation.

For the feature extraction, when the window length is 256, there is only one frequency component in the band of $(0, 0.5)$ Hz. Thus, the MFC location feature in this band always equals 1, which is useless for classification. We delete this feature in the CVT experiment for window length of 256.

The proposed eating detection method is complementary to other methods for recognizing more complex dietary activities. We also leave this for future work.

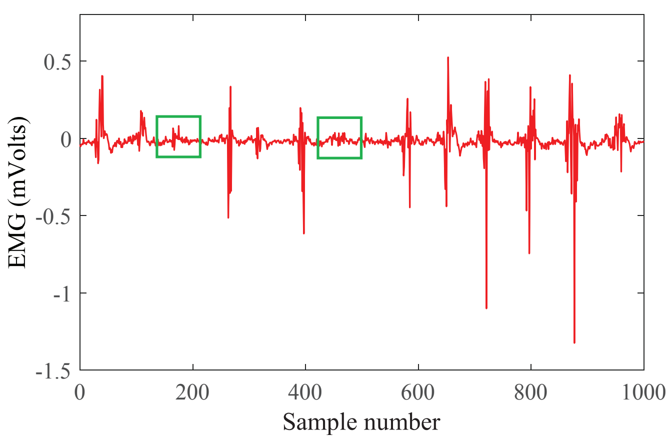

Fig. 10. Less obvious EMG signal

VII. Conclusion

In this paper, we propose a novel eating activity detection and chews counting method. It is done through identifying the mastication muscle contractions using a triaxial accelerometer attached on the temporalis. The accelerometer is embedded in a headband, and only the acceleration data is recorded. Therefore, the proposed method is noninvasive and privacy-preserving. Experiments are conducted with multiple human subjects. The results demonstrate that the proposed method accurately distinguishes eating activity from other daily activities using only 5 seconds of acceleration data. Moreover, the average error rate of chews counting for four users is 12.2%.

Acknowledgment

Special thanks to all volunteers in our user studies and all the anonymous reviewers. This work was supported by NSF CNS-1253506 (CAREER) and NSF CNS-1618300.

References

[1] [Online]. Available: http://www.who.int/mediacentre/factsheets/fs310/en/

[2] C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal, "Prevalence of childhood and adult obesity in the united states, 2011-2012," *JAMA*, vol. 311, no. 8, pp. 806–814, 2014.

[3] [Online]. Available: http://www.dailymail.co.uk/health/article-2141077/Eating-meals-quickly-raise-risk-diabetes-half-times.html

[4] K. L. Schneider, E. Panza, B. M. Appelhans, M. C. Whited, J. L. Oleski, and S. L. Pagoto, "The emotional eating scale. can a self-report measure predict observed emotional eating?" *Appetite*, vol. 58, no. 2, pp. 563–566, 2012.

[5] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster, "Analysis of chewing sounds for dietary monitoring," in *Proceedings of the 7th International Conference on Ubiquitous Computing*, ser. UbiComp '05. Springer-Verlag, 2005, pp. 56–72.

[6] K. Yatani and K. N. Truong, "Bodyscope: A wearable acoustic sensor for activity recognition," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. ACM, 2012, pp. 341–350.

[7] S. Cadavid, M. Abdel-Mottaleb, and A. Helal, "Exploiting visual quasi-periodicity for real-time chewing event detection using active appearance models and support vector machines," *Personal and Ubiquitous Computing*, vol. 16, no. 6, pp. 729–739, 2012.

[8] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Applied psychophysiology and biofeedback*, vol. 37, pp. 205–15, 04 2012.

[9] S. Zhang, M. H. Ang, W. Xiao, and C. K. Tham, "Detection of activities by wireless sensors for daily life surveillance: Eating and drinking," *Sensors (Basel, Switzerland)*, vol. 9, no. 3, pp. 1499–1517, 2009.

[10] R. Zhang and O. Amft, "Regular-look eyeglasses can monitor chewing," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, ser. UbiComp '16. ACM, 2016, pp. 389–392.

[11] E. S. Sazonov and J. M. Fontana, "A sensor system for automatic detection of food intake through non-invasive monitoring of chewing," *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1340–1348, 2012.

[12] [Online]. Available: http://www.gustrength.com/muscles:temporalis-location-action-and-trigger-points

[13] [Online]. Available: http://en.wikipedia.org/wiki/Muscles_of_mastication

[14] [Online]. Available: http://www.gustrength.com/muscles:masseter-muscle-actions-and-trigger-points

[15] [Online]. Available: https://en.wikipedia.org/wiki/Lateral_pterygoid_muscle

[16] Shimmer sensor platform. [Online]. Available: www.shimmersensing.com
[17] X. Wang, "High accuracy distributed target detection and classification in sensor networks based on mobile agent framework," Ph.D. dissertation, Knoxville, TN, USA, 2004.
[18] [Online]. Available: https://www.mathworks.com/help/matlab/ref/fft.html
[19] J. Po, J. Kieser, L. Gallo, A. Tesenyi, P. Herbison, and M. Farella, "Time-frequency analysis of chewing activity in the natural environment," *Journal of dental research*, vol. 90, no. 10, pp. 1206–10, 08 2011.
[20] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.
[21] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*, 4th ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016.
[22] [Online]. Available: http://en.wikipedia.org/wiki/Multilayer_perceptron
[23] J. Yang, S. Wang, N. Chen, X. Chen, and P. Shi, "Wearable accelerometer based extendable activity recognition system," in *2010 IEEE International Conference on Robotics and Automation*, May 2010, pp. 3641–3647.
[24] X. YANG, Q. SONG, and Y. WANG, "A weighted support vector machine for data classification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 21, no. 05, pp. 961–976, 2007.
[25] [Online]. Available: https://www.mathworks.com/help/signal/ref/medfilt1.html
[26] [Online]. Available: https://en.wikipedia.org/wiki/Median_filter
[27] S. Wang, G. Zhou, L. Hu, Z. Chen, and Y. Chen, "Care: Chewing activity recognition using noninvasive single axis accelerometer," in *UbiComp/ISWC'15 Adjunct (Poster)*. ACM, 2015, pp. 109–112.
[28] Y. EL-Manzalawy and V. Honavar, "Wlsvm: Integrating libsvm into weka environment," 2005. [Online]. Available: http://www.cs.iastate.edu/yasser/wlsvm
[29] V. A. Vance, S. J. Woodruff, L. J. McCargar, J. Husted, and R. M. Hanning, "Self-reported dietary energy intake of normal weight, overweight and obese adolescents," *Public Health Nutr.*, vol. 12, no. 2, pp. 222–7, 2009.
[30] D. A. Schoeller, "How accurate is self-reported dietary energy intake?" *Nutrition Reviews*, vol. 48, no. 10, pp. 373–379, 1990.
[31] X. Ye, G. Chen, and Y. Cao, "Automatic eating detection using head-mount and wrist-worn accelerometers," in *2015 17th International Conference on E-health Networking, Application Services (HealthCom)*, Oct 2015, pp. 578–581.
[32] G. Mertes, H. Hallez, T. Croonenborghs, and B. Vanrumste, "Detection of chewing motion using a glasses mounted accelerometer towards monitoring of food intake events in the elderly," in *International Conference on Biomedical and Health Informatics*, 10 2015, pp. 1–5.
[33] S. A. Rahman, C. Merck, Y. Huang, and S. Kleinberg, "Unintrusive eating recognition using google glass," in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '15, 2015, pp. 108–111.
[34] E. Thomaz, I. Essa, and G. D. Abowd, "A practical approach for recognizing eating moments with wrist-mounted inertial sensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '15. ACM, 2015, pp. 1029–1040.
[35] A. Kadomura, C.-Y. Li, K. Tsukada, H.-H. Chu, and I. Siio, "Persuasive technology to improve eating behavior using a sensor-embedded fork," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14. ACM, 2014, pp. 319–329.
[36] C.-Y. Li, Y.-C. Chen, W.-J. Chen, P. Huang, and H.-h. Chu, "Sensor-embedded teeth for oral activity recognition," in *Proceedings of the 2013 International Symposium on Wearable Computers*, ser. ISWC '13. ACM, 2013, pp. 41–44.
[37] M. Farooq and E. Sazonov, "Automatic measurement of chew count and chewing rate during food intake," *Electronics*, vol. 5, no. 4, 2016.
[38] ——, "Segmentation and characterization of chewing bouts by monitoring temporalis muscle using smart glasses with piezoelectric sensor," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1495–1503, 2017.