# Automatic Eating Detection Using Head-Mount and Wrist-Worn Accelerometers


Xu Ye
Department of Computer Science
University of Massachusetts Lowell
Lowell, Massachusetts 01854
ye.xu.nj@gmail.com

Guanling Chen
Department of Computer Science
University of Massachusetts Lowell
Lowell, Massachusetts 01854
glchen@cs.uml.edu

Yu Cao
Department of Computer Science
University of Massachusetts Lowell
Lowell, Massachusetts 01854
ycao@cs.uml.edu



***Abstract*—Automatic Eating Detection (AED) provides an important tool to help users regulate their dietary behavior for many health applications, such as weight management. In this paper we propose an AED solution using a head-mount and a wrist-worn accelerometers that are commonly available in commercial wearable devices. Experimental results, using Google Glass and Pebble Watch, validated that the proposed approach is highly effective to detect head motion from chewing and to detect hand-to-mouth (HtM) gestures when eating, resulting in 89.5% to 95.1% detection accuracy. Further we combined the features from both devices to achieve 97% cross-person eating detection accuracy and the average error when predicting duration of eating meals was only 105 seconds.**


## I. Introduction

The World Health Organization reported that overweight and obesity are one of the leading risks for global deaths. At least 2.8 million adults die each year as a result of being overweight or obese. In addition, a considerable proportion of the diabetes, the heart disease and certain cancer are attributable to overweight and obesity. Monitoring eating behavior thus is important for weight management as it can improve self-awareness of one's food consumption in everyday life. Many existing dietary monitoring tools, however, often rely on self-reports that are burdensome and error prone.

In this paper, we propose Automatic Eating Detection (AED) using motion patterns derived from accelerometers, which are energy-efficient and widely-available in the off-the-shelf wearable devices. Our approach detects Hand-to-Mouth (HtM) gestures (moving food to mouth for a bite) using a wrist-worn accelerometer; and detects unique head motions during chewing using a head-mount accelerometer. The proposed algorithms are validated using a Google Glass and a Pebble Watch, though we expect other head-mount or wrist-worn wearable devices (e.g. smart earphones or wristbands) would also work well.

The contributions of this paper include: 1) a classification algorithm to detect chewing motion using a head-mount accelerometer; 2) a classification algorithm to detect HtM gestures using a wrist-worn accelerometer; 3) combining features from both devices to improve detection accuracy and to predict the length of eating session; and 4) experimental evaluation with 10 subjects and a variety of food types to validate these algorithms using off-the-shelf devices. By avoiding using more advanced sensors (e.g. gyroscope, strain sensor, and microphone), the proposed solution is more scalable and user friendly than other AED approaches, such as instrumented environment (e.g augmented tabletop or ambient camera) or special dinnerware (e.g. augmented trays or forks).

## II. Automatic Eating Detection

Eating activity can be thought as consecutive mixture of chewing sessions and hand to mouth (HtM) gestures. Thus we aim to solve the AED problem by detecting two related events: (1) detecting chewing sessions with a head-mount accelerometer (e.g. Google Glass); and (2) detecting HtM gestures with a wrist-worn accelerometer (e.g. Pebble Watch). We first evaluate these two approaches separately and then combine the features from both devices to improve detection accuracy and to predict eating session duration.

### A. Data Collection

We recruited 10 participants (7 male and 3 female) for this study. These volunteers, from 18 to 52 years old (average 28.7), have average body mass index $27.39kg/m2(SD\pm4.38)$. The participants did not present any medical condition that would hinder food intake or performing normal activities.

A Google Glass and a Pebble Watch were provided to each participant during the study period. Both Pebble and Glass have a pre-installed app we developed to collect accelerometer data. An Android smartphone with a data assembling app was also provided to participants who were asked to keep the phone with them to ensure proper data transmission. The participants worn Pebble on their dominated hand and Glass during the study period. Since accelerometer are common in almost all wearable devices and relatively energy-efficient, we chose to use accelerometer alone for eating detection. The built-in state-of-art accelerometers in Pebble and Glass have low noise, high accuracy, good stability, and capability of measuring low frequency signals. The acceleration data on Pebble and Glass were continuously sampled at 50Hz and transmitted to the phone through Bluetooth.

The participants joined two data-collection sessions, each consisted of four parts: 1) jogging for 5 minutes at their own pace, resting for 5 minutes, and then jogging for another 5 minutes; 2) walking for 10 minutes at their own pace; 3) reading aloud for 5 minutes, resting for 5 minutes, and then reading aloud for another 5 minutes; 4) eating a meal by selecting one item from each of the following food groups: a) pizza, fries, and pasta; b) pudding, cereal with milk; c) burger, sandwich, steak, whole wheat bread; and d) apple, carrot, salad, potato chips. Participants were sitting solely in front of an observer without interactions with other people during eating sessions.

We chose these activities (e.g. jogging, walking, reading) to test how well our algorithms can differentiate head and wrist motions during eating from other typical daily life activities. We also chose four food groups that may generate different chewing motions due to different food textures [1]: soft solid and easy to chew (group a), semi-solid (group b), soft solid and hard to chew (group c), and crunchy hard (group d).

When the participants were eating their selected meals, an observer recorded the start time and end time of each eating session that includes typical HtM gesture, bite, chewing and swallowing sub-activities.

### B. Eating Detection with Google Glass

*1) Segmentation:* We define a chewing session as the period starting after a bite of food to the happening of the first swallow following the bite. For the accelerometer data collected with Glass during eating, the researcher recorded the time of bites and swallows as golden standards of labeling the ground truth of chewing sessions. This is because by observation, people usually chew right after a food bite until a swallow. Thus, bite, chew and swallow are repetitive sequential actions during an eating session that could provide unique features leading to high eating detection accuracy.

The labelled chewing session data and other activities' data was divided into non-overlapping epochs of 4s duration. The 4s threshold selected for the epoch was found to present the best tradeoff between the frequency of chewing and time resolution of food intake monitoring. By observation, a complete chewing session was rarely less than 4 seconds and previous studies [2] revealed that the frequency of chewing differs less between individuals and has well-defined narrow frequency range between 1-2.5 Hz. Therefore, 4s epoch can have 4 to 10 complete chewing motions that are sufficient for detection purpose. Our dataset consists of 3325 chewing epochs, 2950 reading epochs, 2962 walking epochs and 2766 jogging epochs. We have less jogging epochs since one participant did not finish the 10-min jogging.

*2) Feature Extraction:* The accelerometer is located on the optics pod of the Glass device, which users rotate to align the device with their sight. As the sensor outputs the acceleration according to optics pod's frame of reference, there is a misalignment between the user's and optics pod's frame of reference even though Glass is at a relatively fixed position of people's head. We cannot assign the user's chewing motion to one of the sensor axes since we cannot measure the angle of the optics pod directly. In order to produce useful data out of the three axes of the accelerometer and reduce the dimensionality of feature vector, we computed the magnitude of the accelerometer vector. Before processing the data, we first filtered out irrelevant data from the magnitude signal using a Butterworth low-pass of order 20, with a cutoff frequency of 5Hz, since chewing motion occurs in the 1-2.5 Hz frequency range [2]. The Butterworth filter is one of the popular recursive linear filters as it smoothes the time series by eliminating signals above the cutoff frequency while keeping the signals in the pass-band undistorted.

TABLE I
FEATURES EXTRACTED FROM THE GLASS EPOCHS

| # Description | # Description |
|---|---|
| 1 Minimum value | 11 Energy of entire frequency spectrum |
| 2 Maximum value | 12 Energy spectrum in chew range* |
| 3 Mean value | 13 Entropy of spectrum in chew range |
| 4 Root Mean Square | 14 Peak frequency in chew range |
| 5 Median value | 15 Energy spectrum in other range |
| 6 Number of zero crossing | 16 Entropy of spectrum in other range |
| 7 Mean time between ZC | 17 Peak frequency other range |
| 8 Number of peaks | |
| 9 Mean distance of peaks | |
| 10 Mean amplitude of peaks | |

*Chewing range:1-2.5 Hz

We then created a feature vector from magnitude signal of each epoch containing both time-and frequency-domain features. Table I lists all features extracted from individual epochs.

*3) Classification:* We experimented three popular classification algorithms: Naive Bayesian, Support Vector Machine (SVM) and Decision Tree C4.5 using the Weka machine-learning toolkit. We trained and tested classifiers in both in-person and cross-person cross-validation classification. For in-person classification, we built and tested models over subject's own data using 10-fold cross validation. For cross-person classification, we built models based only on other people, i.e. for each of the subjects, we trained a classifier using data from the other subjects, testing it on that subject and we looped the process for all subjects. Results of cross-person classification could be the most unbiased estimate of our classifier's performance if the classifier can be generalized to a new person it has never seen. The performance results from in-person classification can be thought as the ceiling performance of the system. Our results show that SVM outperformed the other classifiers, and Figure 1 shows that the average SVM in-person classification accuracy of 95.1% and cross-person accuracy of 89.5%. Note that cross-person classification for Subject 4 is the worst (<80%), because he has a unique eating habit not observed from other people (he likes to listen to pop songs and shakes his head during eating). However, the in-person classification for this subject is still above 90%.

### C. Eating Detection with Pebble Watch

*1) Segmentation:* During eating a HtM gesture can be broken down to three parts: 1) ascending period corresponding to using hand to send food to mouth 2) biting period in which people bite the food 3) descending period during which people put hand to a rest position or grab food for next bite. Thus

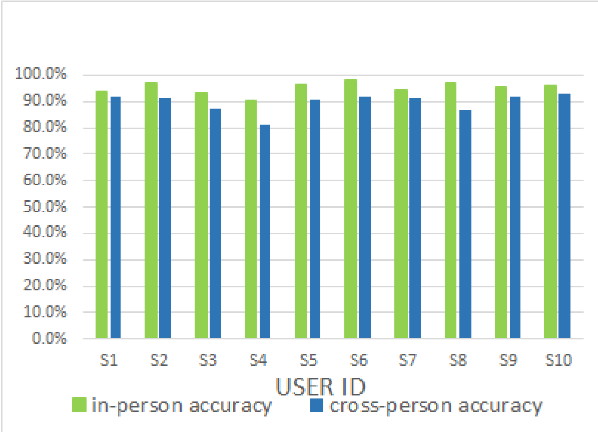


Fig. 1. SVM classification performance on detecting chewing sessions

a bite must happen in the middle of an HtM gesture, and we labelled ground truth of HtM gestures according to the recorded biting time. We extracted a 4-second HtM epoch from the accelerometer data by including 2 seconds before the bite and 2 seconds after the bite. Since HtM happens quickly, the 4s HtM epoch should have sufficient information for detection purpose.

*2) Feature Extraction:* In order to recognize the HtM gesture, we need to extract features that can effectively discriminate different types of gestures. By analyzing the collected accelerometer data at sampling rate 50 Hz on Pebble, we found that 30-55% of the accelerometer data during eating is gravitational data (e.g. no acceleration on any axis), i.e. over 30% even in the HtM episodes.

We know the accelerometer measures the linear accelerations on each axis, while the components of gravitational field vector g exist all the time. If an accelerometer reading is gravitational, it has no linear accelerations or the linear accelerations are negligible compared with gravitational acceleration. Hence, the accelerometer outputs are only the components of g on three axes, which can be used to compute the accelerometer rotary angles ("orientation"). Assuming that the initial orientation of Pebble is lying flat with the earth's gravitational field aligned with z-axis. The output of accelerometer G is:

$$G = \begin{pmatrix} G_x \\ G_y \\ G_z \end{pmatrix} = Rg = R \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta\sin\phi \\ \cos\theta\cos\phi \end{pmatrix} \quad (1)$$

Where $R$ is the rotation matrix, $\phi$ is roll angles on $X$ axis and $\theta$ is pitch angles on $Y$ axis. Since the orientation angles are dependent on the order in which the rotations are applied, we can adopt the commonly used aerospace sequence of roll then pitch and finally a yaw rotation. We can compute $\phi$ and $\theta$ from (1):

$$\phi = \arctan\left(\frac{G_y}{G_z}\right) \quad (2)$$

$$\theta = \arctan\left(\frac{-G_x}{\sqrt{G_y^2+G_z^2}}\right) \quad (3)$$

TABLE II

FEATURES EXTRACTED FROM THE WATCH EPOCHS

| # Description | # Description |
|---|---|
| 1 Number of pitch $P$, $120 < P < 170$ | 13 Maximum value of magnitude |
| 2 Number of roll $R$, $-90 < R < 25$ | 14 Minimum Value of magnitude |
| 3 SD of pitch $P$, $120 < P < 170$ | 15 Mean value of magnitude |
| 4 SD of roll $R$, $-90 < R < 25$ | 16 SD of magnitude |
| 5 Mean value of pitch $P$, $120 < P < 170$ | 17 RMS of magnitude |
| 6 Mean value of roll $R$, $-90 < R < 25$ | 18 Median value of magnitude |
| 7 SD of all pitch $P$ | 19 Number of peaks |
| 8 SD of all roll $R$ | 20 Mean distance of peaks |
| 9 Maximum value of all pitch $P$ | 21 Mean amplitude of peaks |
| 10 Maximum value of all roll $R$ | |
| 11 Min value of all pitch $P$ | |
| 12 Min value of all roll $R$ | |

During eating, the hand may pause briefly when biting food. At that moment the linear acceleration is negligible and the gravitational data can be used to calculate wrist orientation. This is an important observation for detecting HtM gestures since the wrist orientation is typically outward and upward when biting food. Table II includes both the orientation features (1-12) and magnitude features (13-21) for HtM classification. The orientation feature selection was based on two observations: 1) wrist is always higher than elbow when biting, which can be represented by pitch on Y axis ranging from 90 degree to 180 degree 2) inside of hand is always facing to body while biting, which can be represented by roll on X axis ranging from -90 degree to 0 degree. In order to find the best threshold for orentation features, we tuned the pitch in the range of [90, 180] by pace of 5 degree, and the roll in the range of [-90, 0] by pace of 5 degree. Experiments showed that pitch features in range [120, 170] and roll features in range [-90, 25] provide the best outcome.

*3) Classification:* Across the 20 meals, the subjects had a range of 28 to 52 HtM gestures for each meal (38 HtMs on average). SVM was again used to classify the 4s epoch accelerometer data from Pebble Watch. The mean in-person precision was 94.48% with recall of 95.02%. The mean cross-person precision was 92.8% with recall of 90.3%. If we remove orientation features and the classification accuracy based on only magnitude features dropped to 72.5% for precision and 59.5% for recall. *This shows the proposed orientation feature is critical for HtM detection using a single accelerometer.*

### *D. Eating Detection with Combination of Google Glass and Pebble Watch*

Eating typically follows a relatively predictable pattern, involving both hand and mouth motions. In this section, we explore the combination of Glass and Pebble accelerometer data to improve eating detection accuracy and to predict how long the user has been eating.

First, as we collected accelerometer data from the Pebble Watch and the Google Glass simutanously, we experimented to combine the data from both devices to improve the eating detection. In order to compare the performance, we used the same segmentation size and the same feature sets as used in

Glass and Watch detecion. We labelled all HtM and chewing sessions as eating sessions, and others as non-eating sessions. SVM still outperformed other algorithms in the experiments with average cross-person accuracy of 97%, which is better than the cross-person accuracy of using Watch or Glass alone for eating detection. This confirms our hypothesis that combining both head and hand motion features can lead to better classification performance.

Second, we observed that an eating cycle typically lasts for about 15 seconds and repeats 20 to 30 times over an eating session. In each eating cycle, users first experience a transition from HtM to bite, to chew and then swallow. Obviously users may not follow this strict sequence and may skip certain eating stages. For example, chew could return to bite or swallow could jump to chew directly. The dependence between two successive eating stages, however, still exists. This inherent temporal sequence, intuitively, could thus be leveraged to predict the start and the end of eating session by combining motion data from Glass and Pebble. In this study, we tried HMM model on the data by simply using the sequence of combined classification results of HtM and chewing detections as the input of HMM to estimate the eating session duration. The duration was calculated from the difference of the start time and end time. In our experiment, we trained a HMM model for the motion transition of eating and recognize an input sequence as eating motion if $P(O|\lambda)$ is above a fixed threshold (80% in our study) otherwise a non-eating motion. The beginning of the first two consecutively recognized eating epochs was recorded as the eating start time, and the end of the last two consecutively recognized eating epochs was recorded as the eating finish time. The average eating duration error was $\pm 105$ seconds. For a typical 25 minutes meal eating activity, this represents about 7% of the entire session.

## III. Related Work

Based on the observation that a roll of the wrist often occur in most eating situations, Bite Counter uses a wrist-worn gyroscope to detect wrist roll motion pattern for eating with thresholds empirically determined for roll velocity and time intervals [3]. Similarly we focus on a single head-mount or wrist-worn accelerometer that is energy efficient and widely available in today's wearable devices. Our algorithms can recognize head motions caused by chewing and the hand to mouth gestures when eating. Unlike Bite Counter that uses gyroscope to measure the angular velocity of axis along the wrist to detect bites, which could reach a highest recall of 94% but with high false positive rate of about 20%, our method can achieve both good precision and recall using only accelerometer data by considering novel orientation features.

More complex wearable sensing systems were also built to recognize fine-grained ingestive behaviors. Amft et al. used 4 inertial sensors placed on the lower/upper arms to detect eating and drinking arm gestures [4], and later proposed to use earpads to recognize chewing events [5] and to use an EMG/microphone integrated collar device to recognize swallowing events [6]. Acoustic features have also been proposed to detect food intake events, using wearable microphones [7], [8], as chewing and swallowing generate unique sound patterns [9]. These research prototypes so far focus on detection accuracy but lack the commericial availability and usability evaluation (e.g. comfort level and energy consumption to wear in a free-living setting).

## IV. Conclusion

The proposed automatic eating detection is effective in detecting chewing motion using a head-mount accelerometer and in detecting hand-to-mouth gestures using a wrist-worn accelerometer during eating activities. By combining features from both devices, we can improve detection accuray and predict the duration of eating session. Thus we believe this approach is practical since accelerometer is energy-efficient and widely available in commercial wearable devices. For pragmatic considerations, wearing a smart glass may be awkward though it can be acceptable to populations engaged in a short-term weight-loss intervention program. In addition, our algorithm only requires accelerometer data and light-weight feature calculations. Thus it is conceivable for this approach to work well with other wearable devices, such as earphones or wristbands. We plan to build a real system and dietary monitoring applications using the proposed method.

## Acknowledgment

This work is supported partly by the National Science Foundation (NSF) under Grant No. 1040725 and No. 1016823.

## References

[1] K. D. Foster, A. Woda, and M. A. Peyron, "Effect of texture of plastic and elastic model foods on the parameters of mastication," *J. Neurophysiol.*, vol. 95, no. 6, pp. 3469–3479, 2006.

[2] M. A. Peyron, O. Blanc, J. P. Lund, and A. Woda, "Influence of age on adaptability of human mastication," *J. Neurophysiol.*, vol. 92, no. 2, pp. 773–779, 2004.

[3] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "A new method for measuring meal intake in humans via automated wrist motion tracking," in *Applied Psychophysiology and Biofeedback*, 2012, pp. 37:205–215.

[4] O. Amft, H. Junker, and G. Troster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *Proc. Ninth IEEE International Symposium on Wearable Computers*, 2005.

[5] O. Amft, "A wearable earpad sensor for chewing monitoring," in *Sensors,IEEE*, 2010.

[6] O. Amft and G. Troster, "Methods for detection and classification of normal swallowing from muscle activation and sound," in *Pervasive Health Conference and Workshops*, 2006.

[7] K. Yatani and K. N. Truong, "Bodyscope: A wearable acoustic sensor for activity recognition," in *UbiComp*, 2012, pp. 341–350.

[8] T. Rahman, A. T. Adams, M. Zhang, E. Cherry, B. Zhou, H. Peng, and T. Choudhury, "Bodybeat: A mobile system for sensing non-speech body sounds," in *MobiSys*, 2014.

[9] S. Passler and W. Fischer, "Acoustical method for objective food intake monitoring using a wearable sensor system," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, 2011.