Eat Sense V2: Dietary Pattern Detection Using Wearable Sensors

Eric Huang [*] Pushkar Godbole Devpriya Dave

Hao Jiang

ABSTRACT
An automatic eating detection device would prove invaluable in fighting the occurence of obesity and obesity related health issues. This work investigates the use of non-invasive head-based and wrist-based wearable devices for the purpose of detecting eating activity. Our approach utilizes audio and motion data at the head and the wrist. In a trial study with 10 participants performing eating and non-eating activities, our method demonstrated an average precision and recall of 86.12% and 81.36% using 10-fold cross-validation. These results are promising and suggest further research in this domain may generate beneficial results outside of the laboratory.

General Terms
Wearable Computing

Keywords
Activity recognition, Google Glass, Eating detection

1. INTRODUCTION
The occurrence of obesity is growing and is a topic of great concern worldwide. According to WHO, about 2.8 million people die annually worldwide as a result of being overweight or obese, the primary cause being improper dietary habits [9]. With the advent of wearable sensors such as the Google Glass and Pebble SmartWatch, unobtrusive continuous monitoring of behavioral features such as eating patterns has become possible. Here we propose Eat Sense, a wearable computing method to recognize and analyze the eating habits of the wearer. The primary objective of this project is to implement an algorithm to detect when the wearer eats, using the sensor modalities available to Google Glass and Pebble SmartWatch. Such a technology would enable us to detect and classify eating patterns such as eating times, durations, food types etc. Additionally, availability of such data for multiple users could make possible the creation of an eating-habits database to identify demographic information about the dietary habits based on age, sex, geographical location, climate etc. Furthermore, this would allow doctors to monitor an overweight patient's food intake without forcing the patient to log meals themselves. The applications of such information in health care would be limitless.

[*]CS 6601, Fall 2014
Georgia Institute of Technology

2. RELATED WORK
Currently there are no widely used, hands-free tools for monitoring food intake. Before the advent of wearable sensing, Beidler et al. [2] developed a web based application called Personal Nutrition Assistant. This type of system requires the user to input their food intake details after every meal. More recently, the MyFitnessPal mobile app people to manually log dietary habits. However, people may forget or dislike to do this extra task. Our system could alleviate the impetus on people to self-report caloric intakes after each meal. Related work using sensor placements and modalities similar to our set-up show encouraging results for eating activity detection. Our work differs from the prior in that we study eating detection using non-invasive sensors and a much wider range of sensor sources.

The BioGlass project being developed by Hernandez et al. [6], uses the gyroscope and accelerometer in the Google Glass to measure the heart and respiratory rates of the wearer. This demonstrates the high fidelity of Google Glass for detecting miniscule, periodic motions; we believe this fidelity will prove useful in chewing detection. A paper by Dong et al. uses a wrist motion sensor to detect the biting frequency of the wearer under the premise that the wrist of a person undergoes a peculiar rolling motion when he bites on to food [3]. Particularly relevant to our work is the paper by Rahman et al.[8] that makes use of a custom made head-mounted piezo-microphone to detect non-speech body sounds and extract features such as eating, drinking, breathing, laughing using spectral classification. However in this work, the microphone has to be pressed against the neck making it highly inconvenient for regular use. The first work on chewing sound analysis by Drake et al.[4] shows that chewing different foods generates distinct characteristic sounds. Since then, the chewing action and sound has been widely studied in food science. The work by Amft et al.[1] suggests that the placement of microphone near the outer ear canal yields highest level of food intake sounds and the lowest signal-to-noise ratios. However, wearing an ear-mounted microphone can be obtrusive to the end-user. Building on the work by Afmt et al., Pabler et al.[7], have developed an HMM based classifier to detect eating using chewing sounds captured by a head-mounted microphone achieving near 80% precision and recall in a laboratory setting.

In our work, we intend to augment our previous head and wrist-motion based eating classification with head and wrist-

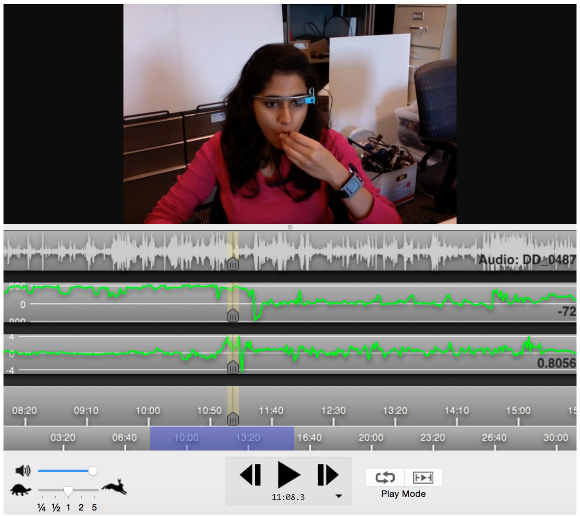

Figure 1: Example "eating activity" in ChronoViz.

mounted microphone data. Since the relevance of all sensor modalities have been independently established, we expect to achieve better precisions and recalls by combining them together.

3. EXPERIMENT

3.1 Equipment Setup

Each participant wore a Google Glass which recorded 3-axis accelerometer and gyroscope data at 50 hz. All participants wore the Pebble smartwatch on the outside of the right wrist and were asked to prefer the right hand for eating. We recorded 3-axis Pebble accelerometer data at 25 hz. We recorded single-channel (mono), 16 kHz audio from the head and wrist microphones. The head microphone was a Olympus VN-4100PC Digital Voice Recorder. The wrist microphone was a Sony ICDUX70 Digital Voice Recorder. Both recorders used Olympus ME 15 clip-on microphones. The head microphone was clipped around Google Glass's touch pad near the right temple, pointing downwards towards the mouth. The wrist microphone was clipped on the Pebble smartwatch, pointing towards the hand. We chose this set of sensor placements and modalities because they were unobtrusive and likely to be popular in the future. We used external microphones for speed-of-development reasons; however, the Google Glass has a microphone in the same location and future smart watches (like the Apple watch) will have quality microphones.

3.2 Experimental Setup

For the data collection, each participant donned the sensors described above and was asked to perform a sequence of four activities. The participants were not told anything about nature of the data collection beyond that it was for activity detection. The activities alternated between non-eating and eating. Furthermore, they were chosen to represent real-world scenarios. For the first activity, the participant watched a video while sitting and then answered a series of conversational questions about the video. The second activity was a repeat of the first activity, except that the participant was given a bag of popcorn and allowed to eat it. In the third activity, the participant was asked to walk around and make a phone call or play with refrigerator magnets. For the fourth activity, the participant was asked to eat a 6-inch Subway sandwich while maintaining a normal conversation. On average, each activity lasted around 6 to 8 minutes. We collected the above activity sequences for ten college-aged participants. We conducted our experiments in laboratory conditions (as opposed to in real dining locations). For ground-truth annotation purposes, we also video recorded all data collection sessions, making sure to start the video before starting the sensor.

3.3 Preprocessing & Annotation

We annotated our experimental studies using ChronoViz [5], see Figure 1. Using this tool, we marked every food intake moment (hand-to-mouth) as an "eating activity". We also manually synchronized each participant's sensor data to their video recording. We used ground-truth video to estimate the sensor start time and then fine-tuned the alignment by matching sounds or motions in the video to microphone sounds or hills and dips in the accelerometer and gyroscope data, respectively. As the clocks on our sensors were not synchronized with each other or the video, we sometimes observed a timing drift of approximately one seconds over the course of the twenty to thirty minute data collection sessions. In effect, a signal highly synced at the beginning of the video would be one second out-of-sync at the end. We avoided this issue by using frames of 7 second widths (2 + 5 on each side) in our classification.

4. IMPLEMENTATION

The design of our eating activity recognition algorithm can be classified majorly into three steps, those of activity detection, feature extraction and finally activity classification.

4.1 Frame Extraction

The annotation data from the experimental phase is used to demarcate the start of each eating activity. Considering the average chewing time of an individual, an eating frame is defined starting two seconds before the eating annotation and lasting five seconds beyond the annotation. The offset data across all recording devices is then used to align the frames and extract the raw data to be fed to the feature extractor.

4.2 Feature Extraction

This is the most significant phase of the eat sense algorithm. In this phase, the raw frame data from all devices is condensed into a finite set of meaningful features to be fed to the activity classifier. The features can primarily be classified into three categories:

4.2.1 Head Motion Features

The two key attributes to be captured from the head motion of an eater are the head orientation and oscillation. To capture these aspects, the following 12 features are extracted from the Google Glass accelerometer and gyroscope: Mean: x, y, z (3 features), Standard deviation: x, y, z (3 features), Two peak frequencies: x, y, z (6 features).

4.2.2 Wrist Motion Features

Since an eating activity always begins with a peculiar a hand-towards-mouth action, the rotation and translation of the subject's wrist reveal vital information necessary to capture the start of a potential eating activity. To that end, the following 9 features are extracted from the Pebble SmartWatch accelerometer: Peak: x, y, z (3 features), Mean: x, y, z (3 features), Standard deviation: x, y, z (3 features).

4.2.3 Audio Features

In addition to the towards-mouth hand motion and oscillatory head motion, the raw audio data from the head and wrist worn microphones was used to augment the algorithm with audio features. The Yaafe audio feature extractor has been used for computing the audio features. Based on works by Amft et al.[1] and Rahman et al.[8] the following 8 audio properties of the frame sounds are used for each of the head and wrist microphones: Zero Crossing Rate (1 feature), Amplitude Modulation (8 features), Spectral Roll-off (1 feature), Loudness per frequency-band (24 features), Energy (1 feature), Envelope Shape Statistics (4 features), Spectral Shape Statistics (4 features), Temporal Shape Statistics (4 features). The combination of these three feature types yields an array of 115 features per sample that are fed to the Activity classifier to detect eating.

4.3 Activity Classification

In our previous work, Support Vector Machines (SVMs) were used for activity classification, due to their memory efficiency and effectiveness in classifying high dimensional data. However SVMs were observed to perform very poorly in light of the new (more realistic) data, particularly with the introduction of audio features. On the other hand, Binary Decision Trees (BDT), as used by Amft et al., performed extremely well in eating detection as may be seen in the subsequent Evaluation. Hence, BDTs have been used for the frame classification. The python based scikits-learn package has been used for implementing the BDTs.

5. EVALUATION

A 10-fold cross-validation (random selection without replacement) was used to evaluate the performance of the Decision Tree based binary classifier. To avoid formation of skewed datasets, the testing and training datasets were constructed with equal number of eating and non-eating samples for each trial. The following table shows the performance of the classifier on different modalities of features (not shown: SVMs achieved 0% accuracy when using all features):

Data	**Precision**	**Recall**
Glass	75.40%	73.75%
Pebble	80.73%	77.80%
Head Mic.	65.29%	66.94%
Wrist Mic.	65.04%	66.58%
Glass, Pebble	82.72%	81.77%
Glass, Pebble, Head Mic.	86.12%	81.36%
Glass, Pebble, Head and Wrist Mic.	83.44%	82.56%

Table 1: Eating classification results.

6. DISCUSSION

The above results establish two key inferences with respect to eating classification using wearable sensors. Firstly, Binary Decision Tree based classifiers are able to perform consistently well even in distinguishing complex and similar tasks (such as talking on phone vs eating). It is highly likely that many of the used audio features are merely adding to noise in classification data. But in absence of deeper knowledge regarding audio processing, all of the potentially relevant audio features have been used for this study. We anticipate that the primary reason for the failure of SVMs in this context, is their inability to incorporate large number of features (115) by weighing them based on their significance. In future work, one may achieve better results by down-selecting the features using Principal Component Analysis.

Secondly, the results show that although the different sensor modalities perform relatively poorly when used individually, their cummulative effect is larger than the parts. Particularly, the combination of Glass accelerometer, Pebble accelerometer and head microphone performs the best in this context. With the potential of wearable sensors in eating detection established, the work can be furthered in the direction of real-time execution and behavioral clustering for multiple users, with an aim to build a demographic eating database. Such a database would add a completely new dimension to the new paradigm of data-centric health care.

7. REFERENCES

[1] 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*, UbiComp'05, pages 56–72, Berlin, Heidelberg, 2005. Springer-Verlag.

[2] J. Beidler, A. Insogna, N. Cappobianco, Y. Bi, and M. Borja. The PNA Project. In *Proceedings of the Sixth Annual CCSC Northeastern Conference on The Journal of Computing in Small Colleges*, CCSC '01, pages 276–284, USA, 2001. Consortium for Computing Sciences in Colleges.

[3] Y. Dong, A. Hoover, and E. Muth. A device for detecting and counting bites of food taken by a person during eating. In *Bioinformatics and Biomedicine*, pages 265–268, 2009.

[4] B. K. Drake. Food crushing sounds: An introductory study. *Journal of Food Science*, 28(2):233–241, 1963.

[5] A. Fouse. Chronoviz. http://chronoviz.com/index.html.

[6] J. Hernandez, Y. Li, J. M. Rehg, and R. W. Picard. Bioglass: Physiological parameter estimation using a head-mounted wearable device. In *International Conference on Wireless Mobile Communication and Healthcare (MobiHealth)*, Nov. 2014.

[7] S. Päßler and W.-J. Fischer. Food intake monitoring: automated chew event detection in chewing sounds. *IEEE Journal of Biomedical and Health Informatics*, 18(1):278–89, Jan. 2014.

[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 *International Conference on Mobile Systems, Applications, and Services*, MobiSys '14, New York, NY, USA, 2014. ACM.

[9] W.H.O. World health organization: Obesity: Situation and trends. http://www.who.int/gho/ncd/risk_factors/obesity_text.