Detecting Mastication - A Wearable Approach

Abdelkareem Bedri
College of Computing
Georgia Tech
abedri6@gatech.edu

Apoorva Verlekar
College of Computing
Georgia Tech
ayp3@gatech.edu

Edison Thomaz
College of Computing
Georgia Tech
ethomaz@gatech.edu

Valerie Avva
College of Computing
Georgia Tech
vavva3@gatech.edu

Thad Starner
College of Computing
Georgia Tech
thad@gatech.edu

ABSTRACT

We explore using the Outer Ear Interface (OEI) to recognize eating activities. OEI contains a 3D gyroscope and a set of proximity sensors encapsulated in an off-the-shelf earpiece to monitor jaw movement by measuring ear canal deformation. In a laboratory setting with 20 participants, OEI could distinguish eating from other activities, such as walking, talking, and silently reading, with over 90% accuracy (user independent). In a second study, six subjects wore the system for 6 hours each while performing their normal daily activities. OEI correctly classified five minute segments of time as eating or non-eating with 93% accuracy (user dependent).

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation (e.g. HCI)**]: User Interfaces

General Terms

Human Factors, Experimentation, Design

Keywords

ear; eating; food intake monitoring; wearable computer

1. INTRODUCTION

In trying to map the relationship between diet and disease [14] researchers have relied on self-report methods to answer questions about what, how and when people eat. Food frequency questionnaires and 24-hour recalls are the most typical instruments employed in this context [5, 14]; they are widely used by health experts and serve as the foundation for nutritional surveillance systems [3].

ICMI '15, November 09-13, 2015, Seattle, WA, USA

DOI: http://dx.doi.org/10.1145/2818346.2820767
.

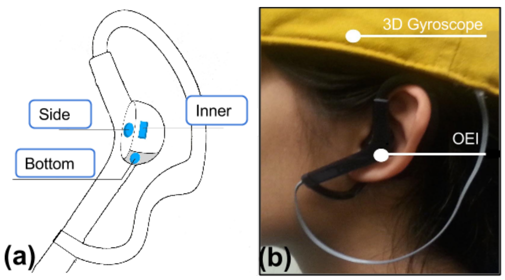

Figure 1: OEI system: a) The proximity sensors on the earpiece. b) User wearing OEI

However, self-report survey tools suffer from biases and are dependent on one's ability to recollect events and activities [8, 9, 10]. This problem has motivated research towards systems that can automatically and objectively track eating both in laboratory and real-world settings [1, 12, 13].

One option is to capture images or video of the user's day continuously and then search for episodes of eating. However, the review of the data would be onerous. A wearable system that could detect eating could help health researchers (or the wearer) find the appropriate episodes in video quickly. Other journaling options include triggering capture only during suspected eating events or querying the user as to what they are eating. Beyond dietary monitoring, an eating detection system might also prompt the user to take medication or perform a necessary procedure (e.g., insulin injection) at appropriate times.

We explore the use of the Outer Ear Interface (OEI) to detect mastication automatically, distinguishing eating activities from non-eating ones. OEI is packaged in an off-the-shelf consumer earbud that mounts outside the ear canal. The system monitors jaw movement by using proximity sensors to measure the deformation it causes in the ear canal and uses a gyro to prevent errors due to body motion.

Contributions of this paper include

- The design of a system that distinguishes eating activities from non-eating activities.
- An evaluation of the system in a laboratory setting where activity sessions were divided into 5-minute segments and labelled as eating or not-eating.
- An evaluation in real-world settings encompassing six one-hour long sessions with six participants.

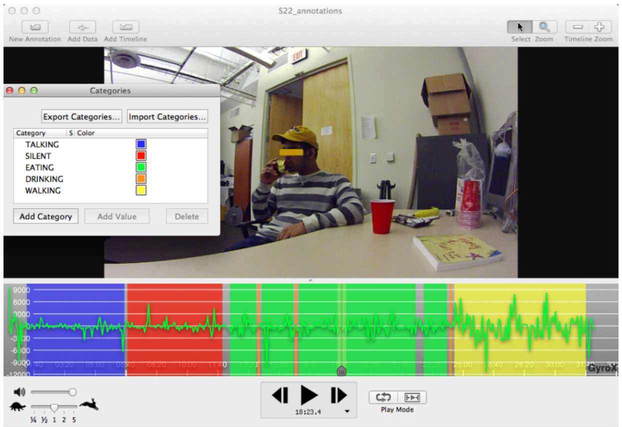

Figure 2: Annotating video using ChronoViz.

2. RELATED WORK

Over the last 20 years, wearable and ubiquitous computing researchers have proposed a variety of methods for automating food tracking and recognition. Acoustic and inertial sensing have been the most commonly used modalities.

Several food intake monitoring systems have been designed for neck placement, such as Bodyscope and Bodybeat [11, 15]. Kalantarian et al. demonstrated distinguishing solids and liquids with an F-measure of 85% and 86% with a sensor located in the neck region. Fontana et al. detected food intake with 89.8% precision, 89.9% recall, and 89.8% accuracy [6] fusing data from a jaw motion sensor, a hand gesture sensor and an accelerometer.

Laboratory studies by Amft et al. [2] with four participants suggests that the placement of a microphone inside the ear canal can distinguish between speech and eating foods (except for very soft foods like rice) at around 99% accuracy and can distinguish between four types of foods with 80% to 100% accuracy. However, wearing a microphone in the ear canal can be obtrusive to the user.

In a previous paper on OEI [4], we report frame level accuracy for detecting eating, talking, walking and silence in a lab setting. Here, we expand upon that analysis, focusing on eating, and move toward real-world practicality of the method with an in-the-wild experiment.

3. SYSTEM DESCRIPTION

The current version of the OEI system (Figure 1) contains three infrared proximity sensors placed in an off-the-shelf sport earpiece and a 3D gyroscope placed in a hat. The sport earpiece has an adjustable loop that goes around the pinna to stabilize the unit, and the three proximity sensors are placed orthogonally with respect to each other to cover a wider area of the ear canal. Due to the wide variety of ear shapes, fitting the sensors properly is difficult; having the three proximity sensors provides redundancy. When a user walks, the earpiece may slightly shift in an out along with the body motion. This movement causes the proximity sensors to generate signals similar to those obtained during mastication. The gyroscope data helps combat this problem (see Section 7). The sensors are sampled at 100 Hz, and the data is logged on an SD card.

4. DATA COLLECTION AND LABELING

Our first experiment was conducted in a laboratory setting. We invited participants to wear the system for about half an hour and asked them to perform a fixed set of activities, including eating, drinking, walking, reading aloud and being silent. The second experiment was an in-the-wild study where we asked participants to wear the system for six hours and perform their normal daily activities. We had different participants in both sets of experiments.

4.1 Laboratory Experiment

In the laboratory study, we had a total of 23 participants with ages varying from 18 to 41, a mean of 24 and a standard deviation of 4.83. We had to discard data from three participants due to errors in data collection.

We designed the study to be easily replicated to ensure uniformity across subjects. Each subject began with 5 minutes of reading aloud from a book, followed by 5 minutes of silently reading or browsing the Internet. Next were 10 to 15 minutes of eating and drinking. Participants were encouraged to behave as they would in normal eating environments. However, the range of food types they could eat was restricted to M&Ms, apples, and bananas. They could also drink from a glass of water. Afterwards, an experimenter walked for five minutes with the participant to a crowded street and down a flight of stairs to test the system's robustness to the user's movement. On average, about 45% of the recorded time was spent in eating activities. Sessions were video recorded for later annotation of ground truth except for when the user left the lab for walking.

We used ChronoViz [7] for annotating the video footage obtained in the laboratory study (Figure 2). The annotation was done by two authors and then verified by a third. Only a few boundary errors were found in the verification step. To synchronize the video footage with the data recorded by OEI on ChronoViz, we used a simple head gesture (move head sidewise from left to right 4 or 5 times) which was distinct in the video as well as in the gyroscope data.

4.2 Wild Experiment

We conducted an in-the-wild study to test the system in more realistic conditions. Participants in this study used the system for six hours each during their daily activities. They were asked to keep a record of any activities that exceeded 15 minutes and to indicate whether they were eating-related activities. This duration can be sufficient for real time applications in the health field, such as reminding a patient to take medication with food. For food intake activities, participants had to specify when start and end to within one minute. We used these labels as ground truth. We obtained data from a total of 6 participants (4 females). One of the authors visually checked regions marked by the participants as eating activity to verify that the gyro and proximity data looked appropriate. Another author reviewed these annotations. A few boundary errors were found and corrected. Of the total time recorded by the participants, 11.2% of the time was spent in eating activities.

5. FEATURE SELECTION

We selected five features to use from the sensor data. The first feature is obtained by dimensionality reduction using Principal Component Analysis (PCA). We performed PCA on the three proximity sensors and selected the first component. Proximity data were smoothed (with an averaging window size of 5 samples), and we took the first derivative of the values to remove signal offset before the dimensionality reduction step. PCA helps recover the jaw movement from the OEI without being affected by the heterogeneity of the user's ear shape.

The second feature reflects the energy in the jaw movement signal, which can be a major difference between speech and mastication gestures. Energy is computed using the RMS value of a sliding window of 50 samples across the values of the first feature. The last three features are the raw gyroscope data from axis X, Y and Z. These features can help detect body motion like walking, which may result in shaking or shifting the OEI sensor causing it to produce signals similar to the ones caused by jaw movement.

In a previous work [4] we used an energy feature computed by averaging the energy in the band between 1.2 Hz to 4.6 Hz. Generating this feature requires short time FFT calculation, which is computationally expensive when compared with RMS. To assess the validity of using the RMS feature we ran a frame level evaluation (frame size 5 seconds) using user independent HMM models built from the same data set [4] and obtained a slightly higher accuracy (83.7 % using FFT feature to 84.3 % with RMS feature).

6. EATING EVENT RECOGNITION

Our goal is to evaluate OEI's capability to recognize eating events when they happen. In food logging, users or health researchers split the time period they are monitoring into smaller time segments, and report in which segments food intake occurs. We follow a similar approach here.

6.1 Training and Testing

Hidden Markov models (HMMs) were trained of eating and null classes using frames of 5 seconds with 50% overlap, extracted from labeled intervals. To recognize eating segments in the recorded data, we performed a sequence of steps. First, we extracted 5 second frames from the entire recorded period using a sliding window that advanced by one second. After computing the features on each frame we test their accuracy against the trained models and classify results as an $Eating$ or Non_eating frame. A binary vector is then generated from recognition results to be compared with ground truth labels.

Before the final comparison the binary recognition vector is filtered using two stages. Using a predefined window size of 15 seconds, the first stage looks for periods where continuous eating is recognized. Unless a period is larger or equal to the window size it will be considered as Non_eating. The second stage smooths the results of the first stage using an averaging window of 150 seconds.

Comparison between ground truth and predicted results are made by segmenting the recorded period into same size non-overlapping intervals. Intervals are marked as eating if any eating activity was labeled or predicted in that interval. We use Precision $= TP/(TP + FP)$, Recall $= TP/(TP + FN)$ and Accuracy (percentage of correctly labelled segments) $= (TP + TN)/(Number\_of\_segments)$ as metrics.

7. LABORATORY EXPERIMENT RESULTS

Using the data collected from 20 participants for a period of half an hour in the lab environment, we trained HMM models for eating as the main class and talking, walking and silence as three null classes. Although drinking events were annotated, they were not used in training steps due to insufficient data. We empirically selected a 10-state, left-right topology for talking and silence classes, and an eight-state left-right topology with additional transitions between $(4 \to 1)$ and $(7 \to 5)$ for eating and walking classes.

Table 1 shows the mean and the standard deviation of the precision, recall and accuracy percentages by evaluating the system in intervals of 1 minute, 3 minutes and 5 minutes, tested on leave-one-user-out, user independent models.

As expected, the gyro was important in distinguishing eating and other classes. Evaluating the eating classifier in a window of 15 seconds without the incorporating inertia features resulted in a significant drop in accuracy (86.7% $\to$ 70.5%), precision (86.3% $\to$ 63.2%)and recall (76.5% $\to$ 50.4%). 67% of false positives and 80.1% of false negatives were due to walking activities.

Table 1: User independent laboratory results

Window Size	Precision %	Recall %	Correct %
1 minute	88.2(± 12.1)	86.1 (± 20.7)	88.5(± 10.4)
3 minutes	89.6(± 15.2)	90.2 (± 16.2)	87.8 (± 12.9)
5 minutes	92.6(± 13.5)	93.3 (± 14.5)	90.4 (± 12.9)

Table 2: User dependent in-the-wild results

ID	False+/hour	Precision %	Recall %	Correct %
S1	0.7	58.3	100	93.1
S2	0.4	57.1	80	94.6
S3	0.4	66.7	100	95.9
S4	0.4	76.9	100	96.1
S5	1.6	52.2	92.2	83.6
S6	0.6	70	100	93.8
Avr	0.68(±0.5)	63.5(±9.2)	95.4 (±8)	92.9 (±4.7)

8. IN-THE-WILD RESULTS

We trained two HMM models, one each for $Eating$ and Non_eating. A left-right topology of eight states with no skip states was used for both models. We used leave-one-user-out cross validation training for user independent testing. For user dependent testing, we randomly reserved 33% of each user's session as an independent test set and used the rest for training.

In Tables 2 and 3 we list user dependent and user independent test results. We provide the average false positive per hour rate as an additional metric for evaluation. The

Table 3: User independent in-the-wild results

ID	False+/hr	Precision %	Recall %	Correct %
S1	1.1	42.9	85.7	87.5
S2	3	16	80	70.3
S3	2.7	24	100	75.3
S4	1.4	47.4	90	85.7
S5	0.6	73.3	84.6	91.8
S6	1.6	46.7	100	83.3
Ave	1.7 (±0.9)	41.7(±20.2)	90.1(±8.3)	82.3 (±8)

time interval window for both tables is five minutes (72 total segments per participant).

9. DISCUSSION

In order to discover food allergies and understand eating behaviors and disorders, health researchers and clinicians conduct laboratory experiments and diary studies. To help encourage patient compliance with reminders based on eating detection, the OEI project strives to create a semi-automatic food journaling device that is unobtrusive and appears similar to current consumer electronics devices on the market. Given the current OEI hardware, one can imagine the proximity sensors and gyro integrated into a popular Bluetooth headset, like the LG Tone Pro or Jawbone. In this sense, OEI is less obtrusive than the devices described by Fontana et al.[6], Rahman et al. [11, 15], and Yatani and Truong[15]. When a potential eating session is detected, a connected smartphone app might ask the user to confirm that they are eating and prompt the wearer to journal the food either by speaking it or typing it on their phone. For such applications, precision and recall for eating detection can significantly improve by adapting the classifier to a specific patient. This goal can be achieved by customizing preexisting user-independent models with correctly classified samples from the current user. For in-lab studies, one could imagine integrating the OEI with a camera pendant, like the Narrative Clip or Looxcie, that constantly images the user's hands and automatically marks segments with a high probability of being an eating event.

In both proposed uses, the recognition system must have a high recall rate and not have so many false positives to annoy the wearer or overwhelm the health researcher. In that regard, the results here are encouraging. The user independent, leave-one-out laboratory study shows recall, precision, and accuracy rates above 90% (a baseline system that would mark everything as *Non_eating* would be 55% accurate).

The in-the-wild results are also encouraging, with a recall rate of over 90%. While the accuracy was poorer than baseline (82% versus 89% for a system that always returns *Non_eating*), the result is understandable given the bias towards the null classes. Assuming the current precision results, a health researcher using a camera pendant would only need to review 1/4 of the continuously captured imagery, and the wearer of an eating journal reminder system would be interrupted falsely 1.7 times per hour. These numbers are higher than desired but may reflect a lack of training data. To investigate, we created a user dependent (and session dependent) model by using two thirds of the data frames extracted from each in-the-wild session to train models for each user and tested the model on the remainder of the 6 hours of data obtained from the same user in the same session. False positives per hour were reduced to 0.68, and the precision improved to 63.5%, suggesting that more data would indeed help improve rates. Looking at the false positives, we noticed a considerable amount of talking (e.g. having a phone call, on a meeting and cooking with a friend), which suggests that creating a null class for talking and adding a microphone might further improve the results.

10. CONCLUSION AND FUTURE WORK

We evaluated the OEI system for detecting eating events both in-the-wild and in the laboratory. We focused development towards making a non-intrusive, everyday-wear system which could help patients remember to take medicine with food or help clinicians sift through "lifelogs" to find eating events. The results are promising and suggest that larger amounts of in-the-wild training and additional sensors may help improve system performance. Going further, perhaps combining OEI's proximity sensors with Amft's microphone-based type-of-food detector [2] might overcome previous limitations with soft foods and lead to a system that identifies more food types.

11. REFERENCES

[1] O. Amft. A wearable earpad sensor for chewing monitoring. *Sensors, 2010 IEEE*, pages 222–227, 2010.

[2] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster. Analysis of chewing sounds for dietary monitoring. In *UbiComp*. Springer-Verlag, Sept. 2005.

[3] E. Archer, G. A. Hand, and S. N. Blair. Validity of U.S. Nutritional Surveillance: National Health and Nutrition Examination Survey Caloric Energy Intake Data, 1971–2010. *PLoS ONE*, 8(10):76632, Oct. 2013.

[4] A. Bedri, A. Verlekar, E. Thomaz, V. Avva, and T. Starner. A wearable system for detecting eating activities with proximity sensors in the outer ear. In *ISWC*, pages 91–92. ACM, 2015.

[5] Z. Z. Falomir, M. M. Arregui, F. F. Madueño, D. D. Corella, and O. O. Coltell. Automation of Food Questionnaires in Medical Studies: A state-of-the-art review and future prospects. *Computers in Biology and Medicine*, 42(10):964–974, Oct. 2012.

[6] J. M. Fontana, M. Farooq, and E. Sazonov. Automatic Ingestion Monitor: A Novel Wearable Device for Monitoring of Ingestive Behavior. *IEEE Transactions on Biomedical Engineering*, 61(6):1772–1779, 2014.

[7] A. Fouse, N. Weibel, E. Hutchins, and J. D. Hollan. Chronoviz: a system for supporting navigation of time-coded data. In *CHI*, pages 299–304. ACM, 2011.

[8] D. R. Jacobs. Challenges in research in nutritional epidemiology. *Nutritional Health*, pages 29–42, 2012.

[9] K. B. Michels. A renaissance for measurement error. *International journal of epidemiology*, 30(3):421–422, June 2001.

[10] K. B. Michels. Nutritional epidemiology–past, present, future. *International journal of epidemiology*, 32(4):486–488, Aug. 2003.

[11] 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. *MobiSys*, pages 2–13, 2014.

[12] 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*, 12(5):1340–1348, Apr. 2012.

[13] E. Thomaz, C. Zhang, I. Essa, and G. D. Abowd. Inferring Meal Eating Activities in Real World Settings from Ambient Sounds. In *IUI*, pages 427–431, New York, USA, 2015. ACM Press.

[14] W. Willett. *Nutritional Epidemiology*. Oxford University Press, Oct. 2012.

[15] K. Yatani and K. N. Truong. BodyScope: a wearable acoustic sensor for activity recognition. *UbiComp*, pages 341–350, 2012.