Eating Episode Detection with Jawbone-Mounted Inertial Sensing

Keum San Chun[1], Hyoyoung Jeong[2], Rebecca Adaimi[1] and Edison Thomaz[1]

***Abstract*— Recent work in Automated Dietary Monitoring (ADM) has shown promising results in eating detection by tracking jawbone movements with a proximity sensor mounted on a necklace. A significant challenge with this approach, however, is that motion artifacts introduced by natural body movements cause the necklace to move freely and the sensor to become misaligned. In this paper, we propose a different but related approach: we developed a small wireless inertial sensing platform and perform eating detection by mounting the sensor directly on the underside of the jawbone. We implemented a data analysis pipeline to recognize eating episodes from the inertial sensor data, and evaluated our approach in two different conditions: in the laboratory and in naturalistic settings. We demonstrated that in the lab (n=9), the system can detect eating with 91.7% precision and 91.3% recall using the leave-one-participant-out cross-validation (LOPO-CV) performance metric. In naturalistic settings, we obtained an average precision of 92.3% and a recall of 89.0% (n=14). These results represent a significant improvement (>10% in F1 score) over state-of-the-art necklace-based approaches. Additionally, this work presents a wearable device that is more inconspicuous and thus more likely to be adopted in clinical applications.**

I. INTRODUCTION

Obesity rates continue to rise around the world. The latest estimates suggest that close to 40% of U.S. adults are obese. This decades-long trend has been fueling the sharp increase of obesity-related diseases such as type-2 diabetes, heart disease, stroke, asthma, and cancer. The impact of this pandemic has been devastating: tens of millions of people die each year, and annual medical costs in the range of 150 billion dollars in the U.S. alone [1]. Also linked with dietary intake, eating disorders have emerged as serious illnesses with long-lasting consequences. For example, anorexia and bulimia nervosa have been linked with osteoporosis and medical issues tied to the gastrointestinal, cardiovascular, and endocrine systems [2], [3], [4]. In light of these diet-related conditions, health researchers and technologists have long looked for *practical* methods to objectively track dietary intake; these methods could be part of interventions to assist people in weight loss, or serve as a screening mechanism to identify individuals (e.g., high-school students) who may be at risk for eating disorders.

With advances in mobile and embedded computing technologies, the vision of Automated Dietary Monitoring (ADM) with mobile and wearable devices has become increasingly more viable. Critically, ADM can bring objectivity to the food logging process. Today, food tracking is based on self-reports such as paper- or mobile phone-based food diaries. These approaches have many shortcomings; they require people to enter detailed information about foods consumed, which is burdensome at best and inaccurate at worst. Health researchers and nutritional epidemiologists have been relying on food frequency questionnaires and 24-hour food for decades. However, these validated survey instruments have similar weaknesses: the data they collect can be unreliable and prone to recall biases [5].

[1]Keum San Chun, Rebecca Adaimi and Edison Thomaz are with the Cockrell School of Engineering, The University of Texas at Austin, Austin, TX, USA gmountk@utexas.edu

[2]Hyoyoung Jeong is with McCormick School of Engineering, Northwestern University, Evanston, IL, USA hyoyoungj@northwestern.edu

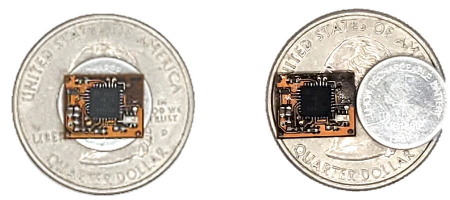

Fig. 1. The figure shows our Bluetooth-enabled wireless sensing device placed over a quarter for comparison. The button cell battery is underneath the board on the left, and next to it on the right. The device samples acceleration data at a rate of 20 Hz.

A significant body of work has been dedicated to ADM, and many approaches leveraging acoustic and inertial sensing have been proposed for eating detection. For instance, there have been attempts to detect eating moments by listening to chewing and swallowing sounds using microphones [6], [7]. Also wrist-mounted trackers have been used to detect food intake gestures [8].

In recent years, researchers have experimented with ADM methods whose sensors are attached to a necklace sitting at the base of the neck [9], [10]. Using proximity sensing, these approaches infer eating activities by continuously measuring the distance between the sensor and the jawbone. In studies conducted in free-living environments, precision and recall measurements over 70% were obtained. Despite these encouraging results, this approach to eating detection was found to have important shortcomings. Chun *et al.* reported that as people performed daily activities such as walking, changes in body position caused the device to move slightly from side to side, causing the sensor to misalign with the tip of the jawbone as its distance point of reference, which resulted in inaccurate measurements and performance loss. Another difficulty involves the use of light-based proximity sensors; they are highly susceptible to ambient light [9].

In this study, we propose a small-scale, wireless wearable

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

device, shown in Figure 1, that permits continuous and direct jawbone motion tracking with an inertial sensor. Motivated by prior work showing the successful use of mastication as a proxy for eating activities, we attach our device to the underside of the jawbone with an adhesive bandage (See Figure 2) [11], [9]. Our studies in lab and naturalistic settings show that our method represents an improvement of over 10% (F1 score) over state-of-the-art necklace-based eating detection approaches.

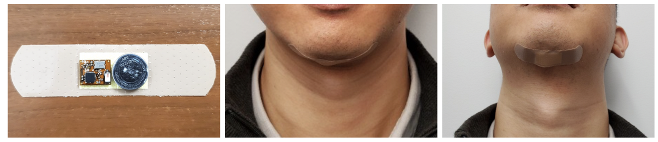

Fig. 2. The wearable was placed on a bandage along with a coin cell battery. This bandage can be attached on the underside of the jawbone to track its movement.

II. Mastication as Proxy for Eating

Mastication is an essential part of the digestive process that helps the breakdown of food and its mixing with saliva containing enzymes. As a distinctive behavioral marker of eating, chewing provides useful information about eating, eating patterns, and possibly other contextual cues, e.g., the texture of food being consumed [12]. We hypothesize that tracking jawbone movements with a sensor directly attached to the jawbone is more robust to motion artifacts, resulting in superior eating detection performance in naturalistic environments.

In order to detect eating activity from chewing, we employ the hierarchical method proposed by Chun *et al.* [9]. In this approach, eating episodes are decomposed into observable sub-actions starting with *chewing* followed by *chewing bout*, as described below:

- **Chewing**: Continuous chewing movement of the jaw lasting 4 seconds or longer. From empirical observations, jawbone movements shorter than 4 seconds are unlikely to correspond to chewing activity and is often due to talking.
- **Chewing Bout**: During an eating episode, chewing is not continuous from the beginning of a meal until the end; people might take momentary breaks for a variety of reasons such as to wipe their mouth with a napkin, take a sip, cough or sneeze, or respond to questions. In our method, a chewing bout corresponds to the combination of consecutive chewing and such non-eating activities that take place within 20 seconds or shorter of each other. Like with chewing, we came to the 20-second duration threshold empirically.
- **Eating Episode**: a group of *chewing bouts* within 2 minutes of each other. Eating episode is what people normally call a breakfast, lunch, or dinner. Eating episode contains not only *chewing bouts* but also longer non-eating activities such as a brief conversation with friends and watching TV, which often arise during eating.

III. System Design

The wearable device was designed to be small (8.6mm-by-10.5mm), low-power and operate wirelessly. As shown in Figure 2), it can be easily concealed underneath an adhesive bandage. The system consists of a wireless MCU, an inertial measurement unit (IMU), and a single rechargeable lithium ion battery. The wireless MCU enables Bluetooth communication with an Android device, and the IMU allows acceleration data collection at a sampling rate of 20Hz. Since the device is intended to be used in contact with the skin, it was designed around a flexible printed circuit board (0.13mm) to maximize comfort.

IV. Data Collection

The system was evaluated in two IRB-approved studies: a lab study and an *in-the-wild* study. The lab study was conducted to test the system in a controlled environment where each participant was closely monitored and ground truth could be reliably obtained. The wild study was conducted to evaluate the system in truly naturalistic settings.

A convenience sample of 10 participants (4 males and 6 females, all college students) were recruited for the lab study through an ad posted on a college campus online bulletin board. The lab study participants were aged between 20 and 33 years old (24.8 ± 5.03). For the *in-the-wild* study, a different set of 15 participants (7 males and 8 females, also college students) were recruited. The ages of these participants ranged between 19 and 22 years old (20.3 ± 0.98).

A. Laboratory Study

Participants performed a combination of eating and non-eating activities in the lab study. Non-eating activities were included to train the recognition model with negative examples. These included watching TV, having a conversation, performing a light-intensity workout, drawing, taking notes with pen and paper, web-browsing, and tooth-brushing (3 minutes for each activity), and walking (5 minutes). The food served for the eating activities included a combination of soup, nuts, ramen noodles, burrito bowls, yogurt, fruit and ice cream. These specific foods items were chosen to represent a variety of food types in terms of viscosity, texture and ease of chewing and swallowing. For all eating activities, the subject was free to stop eating whenever he or she wanted. In total, the lab study lasted for 45 minutes on average. We should note that although we recruited 10 participants for the study, data for only 9 participants could be used due to a data collection error.

B. In-the-Wild Study

The free-living study lasted for 6 hours and did not impose any requirement neither in terms of activities nor location. The only request was that participants consumed at least one meal within the 6-hour session. Participants reported spending time in a wide variety of ways in the *in-the-wild* study, such as performing computer tasks and talking on the phone. These activities and meals consumed were verified as part of our annotation process. Out of the 15 participants recruited for the *in-the-wild* study, 1 participant did not consume a meal during the 6-hour session. Therefore, we discarded the data of this participant from our analysis.

C. Data Annotation

In the lab study, participants were video recorded and researchers later annotated all chewing activities by closely reviewing jawbone movements in the video. Inter-rater reliability using Fleiss' kappa was found to be 0.88 [13]. For the *in-the-wild* study, participants wore a front-facing wearable camera that shot 10-second video clips every minute. These clips depicted participants' activities and context in unconstrained settings, and provided a reliable measure of ground truth for eating episodes.

V. Data Analysis

We applied a *three-phase* pipeline for inferring eating episodes from the inertial sensing data captured by the device (See Figure 3). Phases I, II and III were applied to both the lab and to the *in-the-wild* study data sets.

A. Phase I: Pre-Processing and Frame Extraction

The jawbone-mounted wearable device captured acceleration data along the x, y, and z axes at 20Hz. The data was first normalized using the z-score measure [14], and then frames were extracted using a 4-second sliding window with 50% overlap.

B. Phase II: Feature Extraction and Classification

We extracted a combination of statistical and spectrum-based features for each frame, as shown in Table I, and used it to train a random forests classifier to recognize chewing. We chose this classification algorithm because ensemble learning methods such as random forests have been successfully employed in ADM in prior work [8].

C. Phase III: Clustering

This phase consisted of two sequential grouping steps, which we named Clustering-1 and Clustering-2:

- Clustering-1: This step processed the stream of *chewing* predictions to infer *chewing bouts* using the DBSCAN unsupervised learning algorithm [15]. The DBSCAN algorithm was chosen because it does not require the number of clusters to be determined *a priori*.
- Clustering-2: Here, *eating episodes* were inferred from *chewing bouts* using conditional merging. Adjacent *chewing bouts* were merged based on the temporal distance between them to produce labels representing *eating episodes*. Specifically, if two adjacent *chewing bouts* were separated by no more than 120 seconds, the two *chewing bouts* were merged together as an *eating episode*.

TABLE I
SUMMARY OF FEATURES

Feature #	Description
1-2	max. of y- and z-axis acc.
3-4	min. of y- and z-axis acc.
5-6	skewness of y- and z-axis acc.
7	skewness of the FFT of z-axis
8-9	kurtosis of the FFT of x- and z-axis
10-12	energy of x-, y-, z-axis acc. in f >1.25 Hz band
13-15	max. FFT of x-, y-, z-axis acc. in f >1.25 Hz band
16	covariance of y-axis and z-axis acc.
17	max. FFT of z-axis acc. in 3.75 Hz >f >1.25 Hz band
18	energy of y-axis acc. in f >3.75 band
19-21	energy of x-, y-, z-axis acc. in 2.50 Hz >f >1.25 Hz band
22-23	energy of x-, z-axis acc. in 3.75 Hz >f >2.50 Hz band
24	energy of z-axis acc. in 5.00 Hz >f >3.75 Hz band

TABLE II
SUMMARY OF RESULTS

Study	Detection Unit	Precision	Recall	F1 Score
Lab Study	Chewing	0.802	0.474	0.595
	Chewing Bout	0.850	0.649	0.736
	Eating Episode	0.917	0.913	0.914
In-the-Wild Study	Eating Episode	0.923	0.890	0.906

VI. Results and Discussion

A. Laboratory Study

Using Leave-One-Participant-Out Cross Validation (LOPO-CV), we evaluated the performance of a random forest classifier (50 estimators) trained on the lab data at three hierarchical units: *chewing*, *chewing bout*, and *eating episode*. At the *chewing*, LOPO-CV resulted in 80.2% precision and 47.4% recall. At *chewing bout*, a precision of 85.0% and a recall of 64.9% were obtained. And the evaluation at *eating episode* resulted in a precision of 91.7% and a recall of 91.3%. A summary of the LOPO-CV results is given in Table II.

B. In-the-Wild Study

We trained a classifier using the lab study data and applied the classifier to the data of each participant collected in the wild. Due to the lack of annotation at the level of *chewing* for the *in-the-wild* study, the evaluation was only performed at the level of *eating episodes*, where an average precision of 92.3% and an average recall of 89.0% were obtained. (See Table II).

C. Parameters

In Phase I of our analysis pipeline, features were extracted with a sliding window of 4 seconds of sensor data. We chose this frame duration heuristically; we varied the frame window size from 2 to 8 seconds at 1-second increments and applied LOPO-CV to our data sets, comparing classification performance using the F1 measure. We observed that best performance was achieved with the window size set to 4 seconds in both the lab and *in-the-wild* study data.

Similarly, we needed to choose a maximum temporal distance (MTD) for the *conditional merging* step in Phase III

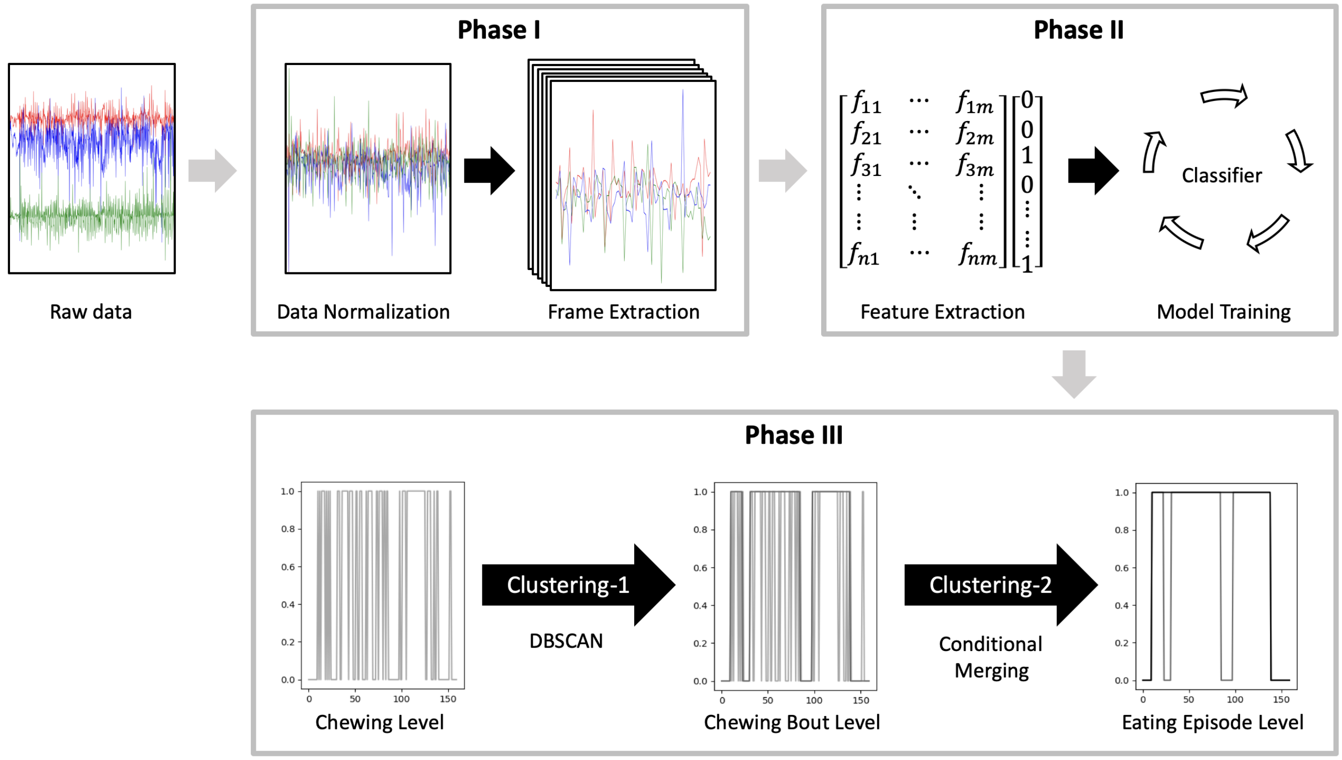

Fig. 3. Red, blue, and green colors indicate acceleration data along the x-axis, y-axis and z-axis, respectively. In phase I, the raw data were normalized using z-scoring and segmented into frames. In phase II, features were extracted from each frame and used for training a classifier. In phase III, predicted chewing frames were clustered (clustering-1) to infer chewing bout. Then, chewing bouts were merged (clustering-2) to predict eating episode.

(Clustering-2). We compared F1 performance while sweeping the MTD parameter from 20 to 400 seconds in 20 second increments. From this analysis, we found out that a MTD of 120 seconds led to the best eating episode classification performance.

VII. Conclusion

In this paper, we put forth an approach for detecting eating episodes based on a small, discrete, and wireless inertial sensing device attached to the underside of the jawbone. In a study conducted in naturalistic environments with 14 participants, this method showed overall precision of 92.3% and recall of 89.0%, an improvement of over 10% (F1 score) over prior work that also leverages mastication as a proxy for eating detection. We believe our method represents a promising contribution towards the development of a practical ADM solution that can be used in everyday life for clinical and non-clinical applications.

References

[1] D. P. Guh, W. Zhang, N. Bansback, Z. Amarsi, C. L. Birmingham, and A. H. Anis, "The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis," *BMC public health*, vol. 9, no. 1, p. 88, 2009.

[2] P. S. Mehler and M. Krantz, "Anorexia nervosa medical issues," *Journal of Women's Health*, vol. 12, no. 4, pp. 331–340, 2003.

[3] D. A. Klein and B. T. Walsh, "Eating disorders," *International Review of Psychiatry*, vol. 15, no. 3, pp. 205–216, 2003.

[4] S. Crow, "Medical complications of eating disorders," *Eating disorders review, part*, vol. 1, pp. 127–36, 2005.

[5] J.-S. Shim, K. Oh, and H. C. Kim, "Dietary assessment methods in epidemiologic studies," *Epidemiology and health*, vol. 36, 2014.

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

[7] 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*, pp. 2–13, 2014.

[8] 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*. ACM, 2015, pp. 1029–1040.

[9] K. S. Chun, S. Bhattacharya, and E. Thomaz, "Detecting eating episodes by tracking jawbone movements with a non-contact wearable sensor," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 4, 2018.

[10] S. Zhang, D. Nguyen, Z. King, J. Pradeep, and N. Alshurafa, "Habits necklace: A neck-worn sensor that captures eating related behavior and more," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. ACM, 2018, pp. 484–487.

[11] S. Zhang, Y. Zhao, D. T. Nguyen, R. Xu, S. Sen, J. Hester, and N. Alshurafa, "Necksense: A multi-sensor necklace for detecting eating activities in free-living conditions," *arXiv preprint arXiv:1911.07179*, 2019.

[12] L. Mioche and M. Peyron, "Bite force displayed during assessment of hardness in various texture contexts," *Archives of Oral Biology*, vol. 40, no. 5, pp. 415–423, 1995.

[13] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa." *Psychological Bulletin*, vol. 72, no. 5, p. 323, 1969.

[14] L. Al Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *Journal of Computer Science*, vol. 2, no. 9, pp. 735–739, 2006.

[15] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.