Toward a Wearable Sensor for Eating Detection

Shengjie Bi†, Tao Wang†, Ellen Davenport†, Ronald Peterson†,
Ryan Halter†, Jacob Sorber*, and David Kotz†
†Dartmouth College and *Clemson University

ABSTRACT

Researchers strive to understand eating behavior as a means to develop diets and interventions that can help people achieve and maintain a healthy weight, recover from eating disorders, or manage their diet and nutrition for personal wellness. A major challenge for eating-behavior research is to understand when, where, what, and how people eat. In this paper, we evaluate sensors and algorithms designed to detect eating activities, more specifically, *when* people eat. We compare two popular methods for eating recognition (based on acoustic and electromyography (EMG) sensors) individually and combined. We built a data-acquisition system using two off-the-shelf sensors and conducted a study with 20 participants. Our preliminary results show that the system we implemented can detect eating with an accuracy exceeding 90.9% while the crunchiness level of food varies. We are developing a wearable system that can capture, process, and classify sensor data to detect eating in real-time.

1. INTRODUCTION

Intake monitoring plays an important role in preventing and treating many diseases including obesity and diabetes. In contrast to many commercial sensing devices that measure physical activity (*caloric output*) such as Fitbit, similar devices to track eating (*caloric intake*) have lagged behind. Accurate eating recognition is the basis for automatic dietary monitoring and can help trigger other kinds of sensing or inquiries. For instance, a wearable camera could be triggered when the eating recognition system detects eating; a digital food journal, which includes times and durations of eating and pictures of food, can be generated and sent to nutritionists for analysis.

Despite substantial research on technology for automatic eating recognition [1, 2, 5, 7, 12, 6], the most common method is still manual record-keeping. It has not yet been possible to accurately and automatically detect eating outside the lab; thus our interest is to make a wearable system robust enough for free-living scenarios. To move from laboratory environments into the wild, there are many challenges in implementing such a system. First, in out-of-lab settings, a variety of environmental noise and subject activities could be misclassified as eating. For instance, coughing may trigger sensor readings similar to those encountered when eating. Second, the material properties of food (e.g., hardness) vary significantly, which require different chewing forces and result in different measurable signals. For soft food like yogurt, the chewing process only generates low-amplitude signals and is challenging to detect, while harder foods like carrots generate high-amplitude signals and are relatively easier to detect. Lastly, it is challenging to build a system that is energy-efficient, unobtrusive and comfortable to wear for an entire day.

The main contributions of this paper are 1) a comparison between two sensing modalities (acoustic and EMG) in terms of performance and usability for free-living scenarios, and 2) demonstrations of the potential for implementing this system as a robust wearable for long-term use in free-living scenarios.

WearSys'17, June 19, 2017, Niagara Falls, NY, USA

ISBN 978-1-4503-4959-8/17/06…$15.00

DOI: http://dx.doi.org/10.1145/3089351.3089355

2. BACKGROUND

We first define the term *eating* used in this paper as "an activity involving the consumption of food and consisting of chewing and swallowing." Detecting the occurrences and durations of eating is the foundation of all other objectives for automatic dietary monitoring, such as food classification and calorie content estimation. Below is a brief overview of common existing methods, which we categorize into three main types: acoustic, EMG, and other.

2.1 Acoustic approach

There are two main types of acoustic sensors: microphones designed for recording sound from the air, and contact microphones designed for recording sound conducted through a solid [8]. Amft et al. evaluated the air-conducted sound intensity of chewing and speech when a microphone is placed at different locations on or near the body [1]. They found that the best microphone position is inside the ear canal, directed towards the eardrum. A microphone placed at this location can capture body-generated sound at a higher magnitude than ambient noise. Papapanagiotou et al. proposed a system that integrates a microphone and a PPG sensor in an ear hook, connected via wire to a data logger equipped with an accelerometer [7]. They evaluated their system in a semi-free-living scenario on 14 subjects and achieved an F1 score of 0.761 for eating detection.

Contact microphones have an advantage over conventional microphones because they only capture vibrations directly from the body surface and naturally avoid ambient noise. Rahman et al. designed a wearable sensing system consisting of a customized contact microphone, an ARM microcontroller, and an Android smartphone [8]. They achieved an average recall of 71.2% for a nine-class classification of different body sounds in laboratory conditions.

2.2 EMG approach

Electromyography (EMG) may be effective in detecting eating, because EMG sensors can capture the movement of muscles used for eating — if the sensor is placed on the correct locations. Zhang et al. fused three electrodes into a 3D-printed eyeglass frame to capture muscle signals during chewing [12]. They achieved a precision and recall of 80% for chewing detection in laboratory conditions and compared signal performance for various electrode placements, sizes, and types. They further developed Bite Glasses, which integrated EMG electrodes and a vibration sensor into the eyeglass frame to identify chewing and texture of food [11].

2.3 Other approaches

Other methods involve inertial sensors, piezoelectric sensors and proximity sensors. The inertial approach focuses on extracting motion patterns during the eating process, especially wrist motion. Shen et al. used a customized wrist-worn device to record the process of eating a single meal for 271 participants [9]. Piezoelectric sensors are capable of producing a voltage at their terminals in response to mechanical stress [10]. By placing a piezoelectric film against the throat, Kalantarian et al. developed a necklace that can detect changes in mechanical stress on necks during the swallowing process [5]. Another novel method uses three proximity sensors in an earpiece to monitor jaw movement by measuring ear-canal deformation during chewing [2].

3. APPROACH

Our goal is to develop a wearable device that can last a waking day and recognize eating in free-living scenarios. Researchers have explored several body locations for eating detection, which include inside the ear canal [1, 2, 7], against the throat [5, 8], and on the wrist [9]. To ensure user comfort for long periods of time and not impede hearing during daily activities, placing sensors inside the ear canal may not be acceptable. The throat is physically close to the location of swallowing, but placing sensors against the throat may be considered too obtrusive by users. Wrist-worn devices tend to be unobtrusive and acceptable to the public, but wrist motion is relatively limited for eating detection and we expect it to be difficult to achieve high accuracy, especially in free-living scenarios.

We chose to place sensors behind the ear; this location is physically close to where chewing happens, giving us access to sound, motion, and electromyographic activities related to eating. A device placed behind the ear does not impede hearing and could be minimized in size to be physically unobtrusive (as in modern hearing aids).

3.1 Bench-top apparatus

We evaluated two off-the-shelf sensors for a behind-the-ear device: a contact microphone (CM-01B, Measurement Specialties) and an EMG sensor (AT-04-001, MyoWare Muscle Sensor). Both sensors are connected to a data acquisition device (DAQ) (USB-1608G, Measurement Computing) with a 20 kHz sampling rate and a 24-bit resolution, while the data collected is processed and analyzed on a laptop.

As shown in Figure 1, the location we used for acoustic sensing is the tip of mastoid bone, a relatively hard surface behind the ear. We fixed the contact microphone under a headband during data collection to maintain stable contact with the body. For the EMG sensor, we used three Ag/AgCl electrodes with gel (24mm in diameter), placed as shown in Figure 2. The ground electrode can be placed anywhere on the body as long as it is relatively far away from the other two electrodes. For convenience, we placed the ground electrode on the back of participants' necks. Figure 3 shows an experiment setup where both sensors are attached to a participant.

Figure 1: Contact microphone

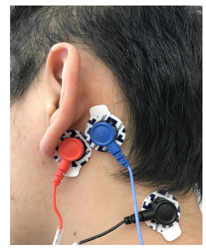

Figure 2: EMG electrodes

Figure 3: Experiment setup

3.2 Wearable apparatus

In addition, we developed a wearable device (Figure 4). Based on the results in Section 5.2, we chose to incorporate only a microphone in our wearable device. We fused the contact microphone, microcontroller (ATSAMD21, SparkFun), SD card and 400 mAh battery into a headband. For this preliminary prototype, we expect the battery life to be at least 8 hours. We plan to test it in out-of-lab, day-long, free-living scenarios.

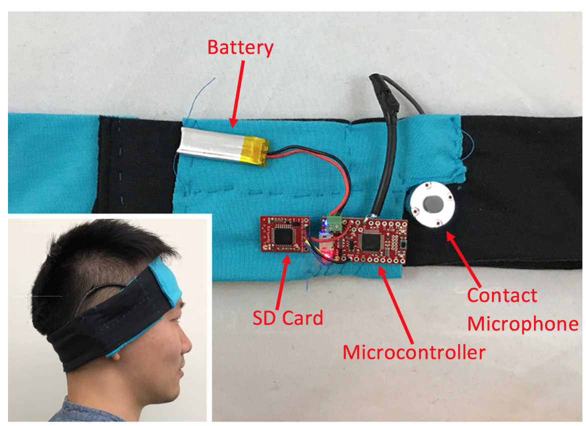

Figure 4: Wearable apparatus

4. METHOD

Our experiments involved multiple stages including data collection on the bench-top apparatus, feature extraction, feature selection and classification on a laptop.

4.1 Data Collection

With the approval of our Institutional Review Board (IRB), we collected data from 20 participants (8 females, 12 males; aged 21-30). For the first 10 participants, we collected data using both contact microphone and EMG sensors. Based on the experiments with the first 10 participants (Section 5.2), we concluded that the EMG sensor was infeasible for free-living scenarios and provided only limited improvement to the accuracy of eating detection. We thus collected data from the second 10 participants using only the contact microphone. All the activities listed in Table 1 were performed, in sequence, by each participant. The total duration of both positive cases (*Eating*) and negative cases (*Non-eating*) are each 12 minutes. All participants ate the same six types of food, shown in Figure 5, among which three (protein bars, baby carrots, crackers) are crunchy while the other three (canned fruits, instant foods, yogurts) are soft. While recording each activity, participants were asked to refrain from performing any other activity and to minimize the gaps between each mouthful. All data recorded during each activity was labeled as the activity.

Figure 5: Six types of food used for data collection

Activity	Description	Duration
Eating	Eat a protein bar	2 minutes
Eating	Eat several baby carrots	2 minutes
Eating	Eat several crackers	2 minutes
Eating	Eat canned fruit	2 minutes
Eating	Eat instant food	2 minutes
Eating	Eat yogurt	2 minutes
Talking	Read an article aloud	5 minutes
Silence	Relax and avoid chewing	5 minutes
Coughing	Cough	24 seconds
Laughing	Laugh	24 seconds
Sniffling	Sniffle	24 seconds
Deep Breathing	Deep breath	24 seconds
Drinking	Drink water	24 seconds

Table 1: The list of activities performed by each participant for data collection

4.2 Feature Extraction and Selection

As sampling rate is one of the most important factors driving power consumption for wearable sensors, we hoped to use a relatively low sampling rate. After testing a range of sampling rates from 250 Hz to 4000 Hz, we chose 500 Hz for eating detection in our system. As a result, all raw data was first downsampled from 20 kHz to 500 Hz before feature extraction. Since the frequency of non-speech body sounds is generally higher than 20 Hz [8], we used a high-pass filter to minimize the frequency components lower than 20 Hz. The filtered signals were segmented into time windows with uniform length and 50% overlap. In this work, we experimented with window sizes ranging from 1 second to 5 seconds and the results are shown in Figure 7. For each time window, we used the open-source Python package *tsfresh* to extract a common set of 206 features per sensor from both time and frequency domains.

The two sensors provide a total of 412 features for evaluation. To improve computational efficiency, we selected relevant features based on feature significance scores and the Benjamini-Yekutieli procedure [3]. Each feature is individually and independently evaluated with respect to its significance for predicting the target under investigation and a p-value is generated to quantify its significance. Then, the Benjamini-Yekutieli procedure evaluates the p-value of all features to determine which ones to keep.

4.3 Classification

We designed a two-stage classification model. In the first stage, to filter out most of the time windows labelled as *silence* using simple thresholding, we calculate the average variance of all time windows labelled as *silence* by ground truth, and find all time windows with lower variance in the entire data set and mark them as "evident silence periods". After separating training and testing data, we train our classifier on the training set excluding the "evident silence periods". Similarly, during testing, we arbitrarily mark the time windows in the testing set that are "evident silence periods" as *Non-eating*. To reduce energy consumption, when we implement a compact, low-energy device, the first stage classification can be done in hardware so that the device does not need to process data during the "evident silence periods". In the second stage, we choose a Logistic Regression classifier to perform a 2-class classification to classify *Eating* and *Non-eating*. We chose Logistic Regression as it is lightweight enough to be implemented in a resource-limited wearable. In both the training and testing data sets, *Eating* is one class and all other seven activities are treated as another class, *Non-eating*.

5. EVALUATION

We evaluated our methods when sensor, window size, bit resolution and number of features vary. We also conducted an uncontrolled-food experiment.

5.1 Evaluation metrics

To evaluate the accuracy of our classifier, we compared its output for each time window against the ground-truth label for that time window. In other words, each time window is an independent test case that results in one of four outcomes:

True positive (TP): Both the classifier and ground truth indicate *Eating*.

False positive (FP): The classifier indicates *Eating* and ground truth indicates *Non-eating*.

True negative (TN): Both the classifier and ground truth indicate *Non-eating*.

False negative (FN): The classifier indicates *Non-eating* and ground truth indicates *Eating*.

We then evaluate our method with three metrics:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

The accuracy score is balanced as we configured our data to be 50% positive cases (*Eating*) and 50% negative cases (*Non-eating*).

We used Leave-One-Person-Out (LOPO) cross-validation to evaluate our classifier's performance. A LOPO model is relatively unbiased because the classifier is asked to detect eating for a new person whom it has not seen before. The model iterates over all possible combinations of the training and testing data set. For each iteration, the data set is divided into two subsets: the testing set (data from one participant) and the training sets (data from all other participants). The classifier is trained on the training sets and outputs three metrics (accuracy, precision, and recall) on the testing set for each iteration. As summary metrics, we calculated the mean and standard deviation of these three scores across all iterations.

5.2 Sensor Comparison

Figure 6 shows the results of eating detection with contact microphone and EMG, independently and combined, for the first 10 participants. During our experiments, we found that repeatable and effective placement of the electrodes used for collecting EMG signals was a challenging task and participants found this task to be unpleasant. Moreover, Figure 6 shows that EMG and contact microphone improve accuracy by 3.2% (with a p-value of 0.005) relative to use of the contact microphone alone. Although statistically significant, this difference is not great enough to be worthwhile given the extra cost, effort and size that would be incurred. EMG also appears to yield the worst performance on its own. We thus decided that it is infeasible to integrate EMG sensors into a wearable suitable for free-living scenarios. In the final 10 participants, we collected data using only the contact microphone and used data from the contact microphone alone for evaluation in Sections 5.3 and 5.4.

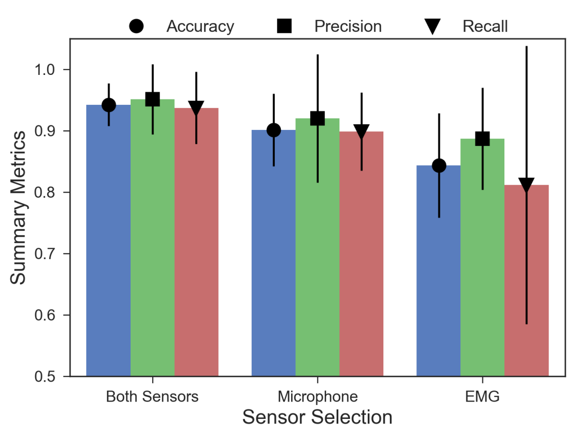

Figure 6: Summary metrics when using contact microphone and EMG, independently and combined (error bars represent standard deviation).

5.3 Parameter evaluation

We explored the effect of different window sizes on accuracy of eating detection in our system by testing a range of window sizes from 1 second to 5 seconds. Based on the accuracy results shown in Figure 7, we chose a 3-second window size for all later evaluations as it yielded the best accuracy.

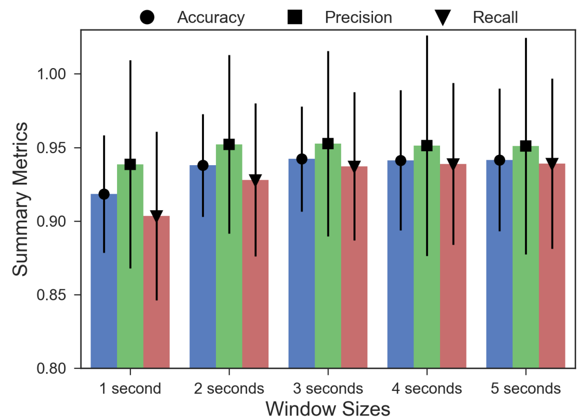

Figure 7: Summary metrics when window size ranges from 1 second to 5 seconds (error bars represent standard deviation)

Moreover, we evaluated whether the bit resolution of analog-to-digital converters (ADC) affects the classification performance. We rounded our raw data (in decimal form) to the third decimal place before feature extraction to simulate 10-bit resolution ADC in a 1V voltage range. As shown in Table 2, lowering the bit resolution did not have a substantial effect on the accuracy of eating detection, so we used a 10-bit resolution for later evaluation.

Resolution	Accuracy	Precision	Recall
24-bit	**0.942** **± 0.036**	**0.953** **± 0.063**	**0.937** **± 0.050**
10-bit	0.935 ± 0.043	0.943 ± 0.075	0.934 ± 0.052

Table 2: Results when bit resolution was 24-bit and 10-bit (mean value ± standard deviation)

Finally, considering the limited computational resources of wearable platforms, we further selected a smaller number of features using the Recursive Feature Elimination (RFE) algorithm with a Lasso kernel. Figure 8 shows the results when the number of top features ranged from 1 to 70.

In general, an increased number of features can benefit accuracy but the improvement is limited (the largest difference of accuracy was less than 8%). To achieve a relatively high accuracy and avoid overfitting due to insufficient features, we chose the top 8 features for later evaluation (Table 3). When we only used the top 8 features for classification, the accuracy, precision, and recall metrics were 90.9%, 91.9%, and 91.1% respectively.

5.4 Uncontrolled-food evaluation

To further evaluate our system on food that was not used for training and under a more realistic condition, we designed an uncontrolled-food experiment. First, using the acoustic

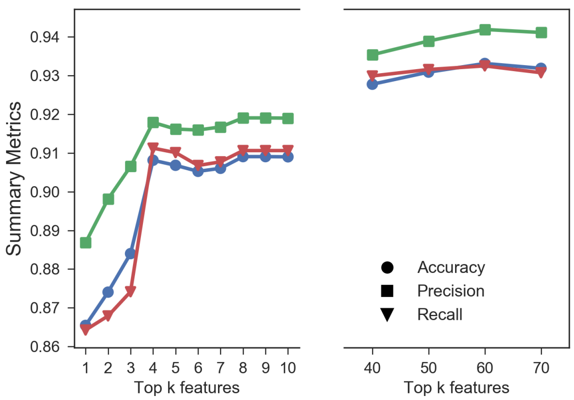

Figure 8: Results when number of features ranged from 1 to 70

Feature type	Description	Number
Kurtosis	Kurtosis	1
Mean	Number of values higher than mean	1
Sum	Sum over the absolute values of changes	1
Peak	Number of peaks at different width scales	4
Friedrich coefficients	Coefficients of polynomial h(x) fitted to the deterministic dynamic of Langevin model [4]	1

Table 3: Top 8 features

data collected from all 20 participants, we trained a classifier with the top 8 features (Table 3) extracted using the same methods as described in Section 4.2. Then, using the bench-top apparatus described in Section 3.1, we asked one participant to conduct a sequence of activities and used the pre-trained classifier to classify these activities in real time with the same classification methods as described in Section 4.3. The food was brought in by the participant (Figure 9) and not like the food used in the training data. To conveniently annotate activities for the ground truth, we asked the participant to perform a series of activities lasting 30 or 15 seconds each following an arbitrarily predetermined routine. The total time length of each type of activity performed in the routine is shown in Table 4.

Figure 9: Food brought in by the participant

Activity	Number of periods	Total time length
Eating	10	300 seconds
Talking	4	120 seconds
Silence	6	105 seconds
Coughing	1	15 seconds
Laughing	1	15 seconds
Sniffling	1	15 seconds
Deep Breathing	1	15 seconds
Drinking	1	15 seconds

Table 4: Activities performed in uncontrolled-food evaluation

The accuracy, precision, and recall metrics for this experiment were 91.5%, 95.1%, and 87.4%. These results show that our system can work properly when participants eat food that was not used for training the classifier. This experiment, however, was only conducted on one participant. In the future, we plan to test on more participants in free-living scenarios to evaluate the performance of our system.

6. DISCUSSION AND FUTURE WORK

In the LOPO cross-validation, we evaluated our system under relatively strict conditions. First, half of the food we used in the experiments was soft and challenging to detect. After training the classifier with six types of food with different hardness levels, we expect our system to be able to detect food with a variety of hardness. Second, the window size we chose for eating detection was 3 seconds, relatively short compared to a meal or an ordinary mouthful. In fact, we aim for our system to detect different types of eating including meals and momentary snacks during long periods, which could be a useful feature for nutrition studies. Third, subjects conducted activities in the experiments for a longer duration than may occur in daily life. In a 12-minute recording session for negative cases, we recorded 5 minutes of silence, 5 minutes of talking and 24 seconds of activities like coughing, sniffling and so forth. We expect the durations of these negative cases, which are relatively hard to distinguish from eating, to be shorter while silence lasts much longer in real life.

One of the limitations to our experiments is that our system relies heavily on chewing detection. If a participant performed an activity with a significant amount of chewing but no swallowing (e.g., chewing gum), our system may output false positives; activities with swallowing but no chewing (e.g., drinking) will not be detected as eating although they may be of interest to some dietary studies. More explorations in swallowing recognition can be a good research direction.

In the future, we plan to construct a behind-the-ear wearable device that can encapsulate the microphone, battery, and data-processing hardware. With this device, we aim to conduct an out-of-lab, day-long, free-living experiment to test our approach for eating detection. Such an experiment will allow us to explore the effect of ambient noise, subject motion, or other subject activities. We plan to explore methods for noise reduction, feature definition, feature selection, and classification, and new metrics for inferring various types of eating activities.

7. CONCLUSION

In this paper, we propose a wearable system for eating detection in free-living scenarios. We developed a bench-top apparatus and collected data of 8 activities from 20 participants. In LOPO cross-validation experiments, we achieved accuracy over 90.9% with 500 Hz sampling rate, 10-bit resolution, 3-second window size and 8 features for eating detection of 6 types of food with different crunchiness level (3 crunchy and 3 soft). Based on the promising results reported in the paper, we plan to further improve our wearable apparatus and evaluate its performance in free-living scenarios.

8. ACKNOWLEDGEMENTS

We thank all the group members of Auracle project (Xing-Dong Yang, Kelly Caine, Kofi Odame, Joe Skinner, Kevin Storer, Jun Gong) for their feedback, and Rui Liu and Gunnar Pope for their help. We also appreciate the advice about eating-behavior research from Tauhidur Rahman, Tanzeem Choudhury, John Batsis, William Kelley, Lisa Marsch, Tam Vu, Adam Hoover, Eric Muth and Joseph Paydarfar.

This research results from a research program at the Institute for Security, Technology, and Society at Dartmouth College, supported by the National Science Foundation under award numbers CNS-1565269 and CNS-1565268. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

9. REFERENCES

[1] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster. Analysis of chewing sounds for dietary monitoring. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2005.

[2] A. Bedri, A. Verlekar, E. Thomaz, V. Avva, and T. Starner. Detecting mastication: A wearable approach. In *Proceedings of the ACM on International Conference on Multimodal Interaction*, 2015.

[3] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. In *Annals of Statistics*, 2001.

[4] R. Friedrich, S. Siegert, J. Peinke, S. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister. Extracting model equations from experimental data. *Physics Letters A*, 2000.

[5] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh. A wearable nutrition monitoring system. In *Proceedings of the International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2014.

[6] M. Mirtchouk, C. Merck, and S. Kleinberg. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2016.

[7] V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos. A novel chewing detection system based on PPG, audio and accelerometry. *IEEE Journal of Biomedical and Health Informatics*, 2016.

[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 *Proceedings of the Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2014.

[9] Y. Shen, J. Salley, E. Muth, and A. Hoover. Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables. *IEEE Journal of Biomedical and Health Informatics*, 2016.

[10] T. Vu, F. Lin, N. Alshurafa, and W. Xu. Wearable food intake monitoring technologies: A comprehensive review. *Computers*, 6(1), 2017.

[11] R. Zhang and O. Amft. Bite glasses: Measuring chewing using EMG and bone vibration in smart eyeglasses. In *Proceedings of the ACM International Symposium on Wearable Computers*, 2016.

[12] R. Zhang, S. Bernhart, and O. Amft. Diet eyeglasses: Recognising food chewing using EMG and smart eyeglasses. In *Proceedings of IEEE International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2016.