FitByte: Automatic Diet Monitoring in Unconstrained Situations Using Multimodal Sensing on Eyeglasses

Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel
Carnegie Mellon University
Pittsburgh, PA, USA
bedri, dianali, rushil, kbhuwalk, mayankgoel@cmu.edu

Figure 1: FitByte was trained and validated using data collected in five unconstrained situations: (from *left* to *right*) in a lunch meeting, watching TV, grabbing and consuming a quick snack from a cafe, exercising in a gym, and hiking outdoors.

ABSTRACT
In an attempt to help users reach their health goals and practitioners understand the relationship between diet and disease, researchers have proposed many wearable systems to automatically monitor food consumption. When a person consumes food, he/she brings the food close to their mouth, take a sip or bite and chew, and then swallow. Most diet monitoring approaches focus on one of these aspects of food intake, but this narrow reliance requires high precision and often fails in noisy and unconstrained situations common in a person's daily life. In this paper, we introduce FitByte, a multi-modal sensing approach on a pair of eyeglasses that tracks all phases of food intake. FitByte contains a set of inertial and optical sensors that allow it to reliably detect food intake events in noisy environments. It also has an on-board camera that opportunistically captures visuals of the food as the user consumes it. We evaluated the system in two studies with decreasing environmental constraints with 23 participants. On average, FitByte achieved 89% F1-score in detecting eating and drinking episodes.

Author Keywords
Eating detection, drinking detection, diet monitoring, health sensing, activity recognition, wearable computing, earables, and ubiquitous computing.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04.
http://dx.doi.org/10.1145/3313831.3376869

CCS Concepts
•Human-centered computing → Ubiquitous and mobile computing design and evaluation methods; •Applied computing → Consumer health;

INTRODUCTION
To understand the relationship between diet and health, nutritionists and users often rely on user-maintained food journals. Monitoring diet involves three main questions: (1) *whether* the user is eating [13, 15]; (2) *what* is the user eating [20, 36] ; (3) *how much* is the user eating [10, 32]. Maintaining such detailed information can be intimidating and users often miss recording their foods on time or forget what they ate earlier in the day [16, 18]. As a solution, many researchers have developed wearables to automatically monitor diet. These proposed systems usually work well in relatively-constrained environments, but there are three main unsolved challenges: (1) usable performance in the user's environment has been elusive; (2) many food types are hard to detect (such as liquids and soft foods); and (3) the wearable sensors and the machine learning models do not generalize across users.

A primary reason for these challenges is the gap between the noise in training and test data used to train a machine learning model. A model trained on clean data collected in a lab does not generalize adequately for data from a noisy environment. Although a number of recent efforts have tried to address this challenge, it is still hard to collect training data in completely unconstrained settings. Moreover, most diet monitoring wearables detect one of the actions that occur during food intake; *e.g.,* hand movement while bringing the food to the mouth [13, 29], biting and chewing [5, 36], or swallowing [25, 33]. These approaches often fail when there is no clear, repetitive hand movement (*e.g.,* user using a straw), or when

the user consumes liquids or soft foods (*e.g.,* yogurt, ice cream, soup) [4, 19]. Also, regular hand-to-mouth/face gestures are not always associated with food intake (*e.g.,* smoking, face scratching, adjusting glasses). Thomaz *et al.* found that wide variation in how users perform the hand-to-mouth gesture makes it difficult to build generalizable models [29]. Thus, it is attractive to detect all actions associated with consuming food but performing such a task from a single wearable has not been attempted yet.

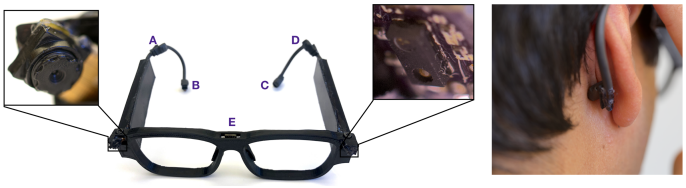

Figure 2: FitByte hardware. The device has one camera, one proximity sensor, and six IMUs. One IMU each at A, B, D, and E. At C, there are two IMUs: one gyroscope and one 4 kHz accelerometer to measure body vibrations behind the ear. The left temple houses the battery and the right one has the microcontroller. The IMUs are attached to a flexible fulcrum (*right*) to ensure snug fit and good connection with heads of different sizes. The temple tips are also flexible so the user can twist them to ensure good fit.

In this paper, we introduce *FitByte*, a pair of eyeglasses that tracks the wearer's food consumption using multi-modal sensing to capture all food consumption actions. FitByte (Figure 2) detects: (1) chewing by monitoring jaw motion using four gyroscopes around the wearer's ears; (2) swallowing by listening to vibrations in the throat using a high-speed accelerometer; (3) hand-to-mouth gestures using a proximity sensor; and (4) visuals of the consumed food using a downward-pointing camera. The camera points downwards to capture only the area around the user's mouth (Figure 8); thus maintaining the privacy of the wearer and people around them. The built-in camera also provided the groundtruth information about the user's activities for one of the two studies performed to model and evaluate FitByte. To develop FitByte's machine learning and sensor selection algorithm, we put 18 participants in noisy conditions (such as hiking, exercising, lunch meetings) as they consumed foods and drinks of their choice. These situations allowed us to collect training and validation data while the user was walking, talking, eating, drinking sporadically, and naturally performing other activities in noisy environments. Modeling using such noisy data allows the algorithm to generalize across conditions and perform well in free-living conditions. Our experiments show that FitByte identifies eating episodes with 94.1% recall and 91.4% precision in all five situations.

To test the system further, we developed a real-time implementation of our learned model to turn sensors on or off depending on the model's inferences. The most power-hungry sensor on FitByte is the camera. The camera is also privacy-invasive. Thus, we turned the camera on only when the model detected that the user was eating or drinking. We evaluated this real-time implementation with five participants over 91 hours. Each participant wore FitByte for 12 hours each day for up to two days. Overall, across the two studies, FitByte was able to detect 61 out of 69 meals or snacks, and falsely detected only 7 eating episodes. In future, we plan to show FitByte's inferences and captured visuals on the user's phone. At the end of the day, the users will be able to browse through the inferences and recall what they ate. Our results show that the users, on average, will get less than one false positive per day. Given FitByte will include a visual for the inferred meal, the users will be able to filter out false inferences quickly. To evaluate the clarity of the visuals captured, we recruited two volunteers who correctly identified the food type in 57 out of 62 meals/snacks. Finally, we conducted a preliminary assessment of FitByte's perceived privacy and social acceptability aspects through semi-structured interviews with study participants.

The main contributions of this work are:

1. The design and implementation of sensor-equipped eyeglasses that monitor all actions of food intake from a single wearable.
2. A data processing pipeline to identify food consumption moments and automatically record food visuals to aid in identifying the food type.
3. A real-time implementation of the algorithm that allows an untethered wearable to monitor diet and capture food visuals using the built-in battery.
4. A preliminary investigation of FitByte's social acceptability and privacy concerns.
5. An annotated dataset of multi-sensor data collected in the user studies to aid in reproducibility and enable expansion of current work.

RELATED WORK

Prior work in automatic diet monitoring (ADM) has focused on detecting atomic actions that a user makes to eat or drink, such as detecting hand to mouth movement, chewing, and swallowing. Researchers have tried to identify these actions by monitoring activities of the wrist, jaw, and throat, as well as detecting chewing and swallowing sounds using different sensing modalities [23, 27].

Detecting Jaw Motions

Several sensing approaches have been employed to detect jaw movement. GlassSense [12] monitors jaw activity from the temple using two load cells embedded in the hinge of custom eyeglasses to detect eating episodes. Similarly, Farooq and Sazonov [14] used a piezoelectric strain sensor placed on the temporalis muscle to detecting chewing bouts. Bedri *et al.* [5, 6] used three infrared proximity sensors embedded in an off-the-shelf earpiece. The sensors detect the ear canal deformation due to movement of the lower jaw bone tip. Chun *et al.* [11] used an infrared proximity sensor placed on a necklace and positioned it pointing upward to detect jaw motion. Rahman *et al.* used the inertial sensor placed in Google Glass to collect a data set of human activities in a controlled setting from 38 participants [24]. EarBit is another system that used inertial sensors to detect jaw motion due to chewing [4]. Bi *et al.* [7] put EMG gel electrodes and a contact microphone behind participants' ear. Zhang and Amft built custom 3D printed eyeglasses with EMG sensors [34, 35]. The EMG

dry electrode is placed on the eyeglasses temples to capture the Temporalis muscle movement. The system achieved 95% F1- score for detecting chewing instances in unconstrained environments.

All these approaches to detect jaw motion work and many of the recent work modeled the data from semi-constrained environments. However, because these approaches focus on detecting only jaw motion, it is hard to detect liquids and soft solids such as yogurts and ice-creams. For that, there is a need to add other sensing modalities.

Detecting Chewing and Swallowing Sounds

To detect liquids and solids, the most promising approach is to listen to throat sounds using a sensor in the ear or on the neck. Some commercial products like breastfeeding monitors[1] also use a similar approach. In lab studies, past work has shown that a microphone placed inside the ear could distinguish eating from other activities [3]. A number of efforts have also studied the neck as the sensor location to listen to chewing and swallowing sounds [8, 21, 25, 33].

One of the primary challenges with these approaches is achieving usable performance outside the lab as microphones (even surface-coupled) are extremely susceptible to environmental noise and motion artifact.

Detecting Hand Movements

Another well-studied approach to detect food consumption is by observing hand movements using inertial sensors. Amft *et al.* [1] instrumented two participants with four XSens-MT-9B motion sensors placed on the upper and lower arm of each hand, who performed several activities in a controlled setting. Dong *et al.* [13] instrumented participants with inertial sensors on wrist for long periods of time (between 8.5 and 12 hours). Thomaz *et al.* [29] also had participants wear an inertial sensor on wrist and asked them to engage in several eating and non-eating activities.

Tracking hand motion for dietary monitoring is promising because of ubiquity of wrist-worn motion sensors. However, large variance between users and the similarity of the hand-to-mouth gesture to other regular hand movements makes it challenging in unconstrained environments.

Identifying Food Type

To identify food type, some approaches rely on differences in the pattern of motion or sounds produced when eating or drinking that specific food [9, 19, 25]. Such approaches often fail in unconstrained situations as correlation between food type and body gestures may not be sufficient to account for general variability in the environment. Other approaches utilize a camera to capture images of the food during an eating/drinking event to identify its type. Prior attempts that have used a camera suffered from a few issues including the inability to identify moments of interest to trigger the camera, inability to consistently capture good view of the food, managing the camera's power consumption, and privacy concerns [17, 28, 30].

[1] https://mymomsense.com

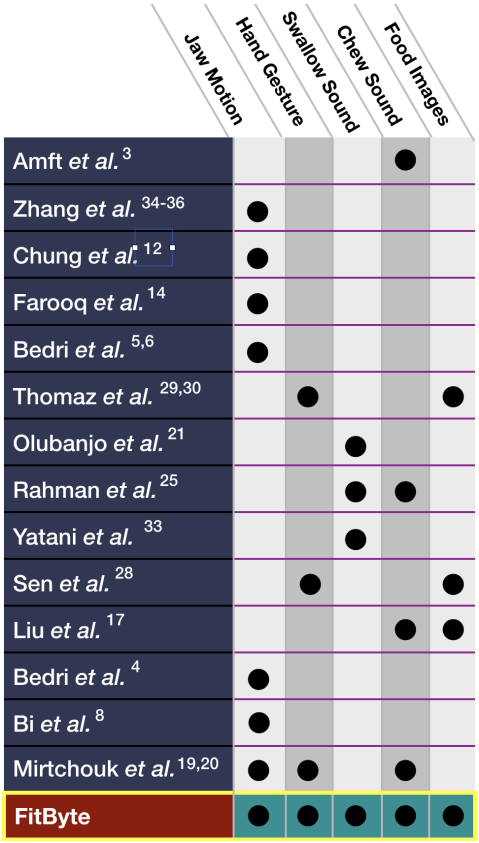

Figure 3: An overview of physical phenomena sensed by past research efforts. FitByte builds on past work and aims to sense all of these physical phenomena. This table highlights representative examples from the literature and it does not provide a complete survey.

What is Missing?

As Figure 3 shows, most approaches summarized so far focus on sensing one particular physical phenomenon that captures some aspect of food intake. However, to counter the noise of real-world, it is attractive to utilize the redundancy of different sensing modalities. By capturing multiple physical phenomena during food intake, diet monitoring systems can better detect eating and drinking instance in unconstrained environments. For example, Mirtchouk *et al.* [19, 20] used Google Glass, two smartwatches, and a headset to capture jaw motion and hand gestures using inertial sensors, and recorded chewing sounds using an in-ear microphone. However, wearing multiple devices was uncomfortable and socially-unacceptable.

HARDWARE

FitByte attempts to address the challenges listed at the end of last section by finding sensing proxies for each sensing approach such that it can be placed on a pair of eyeglasses. For example, instead of detecting hand motions via a wrist-worn motion sensor, FitByte uses a proximity sensor to sense when the hand comes close to the mouth. Overall, FitByte detects jaw motion, hand gestures, swallowing and chewing sounds, and opportunistically records food visuals to aid the user in recalling their foods and drinks (last row of Figure 3).

In this section, we describe the utility of different hardware components of FitByte.

Form Factor

To ensure good compliance, it is important to use a commonplace and comfortable form factor. 76% of the adult population in the U.S. wears some form of vision correction; with more than 50% using eyeglasses[2]. This number is poised to increase further as smart eyeglasses become more popular and useful. Moreover, eyeglasses provide a perfect platform to sense multiple phenomenon simultaneously.

Sensors

Existing diet monitoring approaches have mostly focused on detecting one food intake action [22]. We believe, given noisy situations encountered by most sensors, it is important to maximize the number of sensed phenomena and add some redundancy to sensing.

Proximity Sensor

Hand-to-mouth gestures are quite indicative of food consumption. Past work has investigated the use of wristworn IMUs to model the shape of motion of the user's hand as they consume different foods [29]. Unlike past works that use wristworn motion sensors, FitByte uses an infrared proximity sensor (VCLN-4040) with a range of 20 cm (sampled at 50 Hz) at the left edge of the frame facing the mouth region. From this location, the sensor only detect when the hand comes close to the mouth region (Figure 4). Given this sensor is very power-efficient, we also use it as a switch to turn more power-hungry sensors in FitByte's real-time implementation.

Gyroscopes

A number of past research efforts have shown that mastication can be detected by observing movement of facial muscles [4,5, 11]. To track chewing, we placed four gyroscopes (MPU9250; sampling at 50 Hz) on the arms of the eyeglasses to monitor the movement of the temporalis muscle and the jaw bone from both sides (Figure 2: A-D). Although one gyroscope might be enough to measure this movement, we placed four sensors to evaluate the best location for the sensor and utility of combining information from multiple sensors. In addition, we added a fifth gyroscope in the nose bridge (Figure 2: E) to help in canceling any large body motions (such as head turning or walking) captured by other gyroscopes.

High-Speed Accelerometer

To monitor swallowing and chewing sounds, we use an accelerometer (MPU9250, sampling at 4 kHz). Instead of placing the sensor directly on the throat, we placed the sensor as close to the throat while still being on the eyeglasses. We attached the sensor to the tip of the right temple (C in Figure 2); which positions it underneath the ear and close to the lower jaw and throat. At this location, it can capture vibrations propagated due to swallowing (vertical arrows in Figure 4). As evident in the figure, the sensor also captures vibrations due to chewing and talking. We will model the accelerometer data to filter out the noise from talking in the next section.

[2]https://www.thevisioncouncil.org/sites/default/files/Q415-Topline-Overview-Presentation-Stats-with-Notes-FINAL.PDF

Camera

To help capture visuals of the consumed food, we use a miniaturized camera (Adafruit Mini Spy camera (480p video and 1280×720 photo)[3]. We placed the camera at the top-right corner of the frame to capture activities around the mouth region (Figure 8). This position stops the camera from capturing the user's entire face or scene in front of them. In addition, we removed the microphone from the cameras.

Microcontroller and Power

FitByte uses a Teensy 3.6 board. The Teensy and the camera module are placed in the right arm of the eyeglasses (Figure 2). To power the setup, we used two 150 mAh LiPo batteries and the SparkFun LiPo Charger Basic (Micro-USB) placed in the eyeglasses' left arm.

Fitting

To ensure a universal fit, we iterated over different designs and evaluated them with five new participants at each iteration. For the final design, instead of 3D printing the whole chassis, the temple tips are made out of 10 gauge solid copper wire covered with heat shrink. This ensures that the users can twist and turn the temple tips to their size and ensure good contact. The wire is also flexible enough that it flexes as the user's jaw moves. To measure throat vibrations, it is important to have the high-speed accelerometer in contact of skin. Thus, we added 3D printed flexible fulcrums to hold the sensors snug (Figure 2 - *(Right)*). None of the participants in various pilot studies or the formal data collection found FitByte uncomfortable. However, participants who were not used to wearing eyeglasses felt minor fatigue at the end of some of the sessions.

ALGORITHM

In this section, we explain our signal processing and machine learning approach to detect when food is consumed (Figure 5). We also assess what sensors are most useful for an accurate detection and develop a real-time implementation that relies on a subset of sensors. Once it is inferred that the user is eating/drinking, FitByte opportunistically records food visuals.

[3]https://www.adafruit.com/product/3202

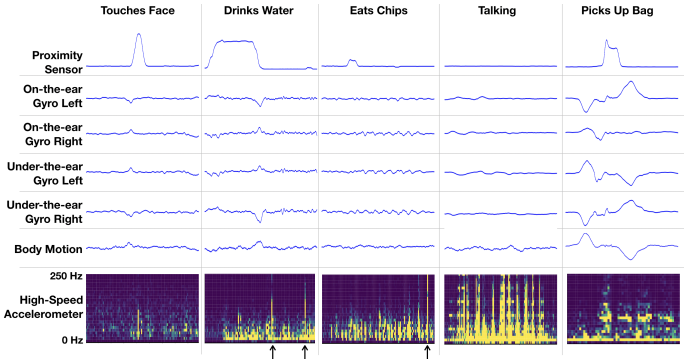

Figure 4: Signals from FitByte's sensors as the user performs different activities. The point in times marked by the vertical arrows at the bottom indicate swallows.

Initially, FitByte records data from all 5 gyroscopes (50 Hz), high-speed accelerometer (4 kHz), and the proximity sensor (50 Hz). We then condition and filter the sensor data, and extract relevant features. A machine learning model then recognizes eating and drinking events and distinguishes them from other everyday activities such as movement, talking, and no-activity.

Signal Conditioning and Feature Extraction

First, all data is smoothed with a 5-second moving average window to remove any high-frequency noise. Second, we compute the first derivative of the gyroscope signals to remove any drift. Then, we segment the conditioned signals for each sensor into 5-seconds windows sliding by 1 second.

Features for the Gyroscopes

FitByte uses gyroscopes near the ear to monitor jaw motion, the gyroscope data is repetitive for FitByte too (Figure 4 "Eats Chips"). Bedri *et al.* [4] used a similar gyroscope mounted near the ear for diet monitoring. They developed features to estimate the periodicity and shape of the repetitive motion of a masticating jaw. Thus, we use the same features as Bedri *et al.* for all four gyroscopes (*i.e.,* 13 features $\times$ 3 axes $\times$ 4 gyroscopes = 156 features).

Features for the Proximity Sensor

For the proximity sensor, we calculate mean, variance, entropy, absolute median, number of peaks above an empirically-defined threshold, and variance of duration between peaks.

Features for the High-Speed Accelerometer

For the accelerometer, FitByte extracts features from the spectrogram (|FFT| = 40 bins) after quantizing it into 18 bins. Figure 4 shows the spectrograms for the accelerometer only up to 250 Hz. Most of the information related to dietary activities were concentrated in this lower frequency band. Thus, we dedicated four equal size bins for the region under 100 Hz. The region between 100 Hz and 600 Hz was divided into 10 50 Hz bins, and 600 Hz to 2 kHz was divided into 4 bins. We used the same 5 second window to compute feature from all 18 bins. We specifically calculate mean, variance, entropy, 95% and 5% percentile, number of peaks, and variance between the peaks. These features mainly focus on measuring the energy and the degree of variation in each bin.

Detecting Food Consumption

FitByte's 5 second long feature extraction window moves with a step size of 1 second. Thus, we classify every second into 5 activities: *eating*, *drinking*, *walking*, *talking*, and *silence (or no activity)*. We trained a Random Forest classifier (Scikit-learn implementation, default parameters, 100 trees). To ensure user independence, we validated our models using leave-one-user-out-cross validation and did not use any data from the same participant.

FitByte's primary task is to detect food consumption episodes. This recognition is performed in three stages:

Frame-level Recognition:

Here we detect whether the user is consuming food at a 1 second resolution. Achieving reasonable precision and recall at such high resolution is not directly useful for the wearer, but it lays the foundation for other more usable results.

Intake-level Recognition

At this stage, we convert the high-resolution inferences into an intake-level decision, *i.e.,* whether the user took a bite (informed either by the hand-to-mouth gesture sensed by the proximity sensor or biting sensed by gyroscopes) and then continued to chew (for at least 3 seconds) or swallowed or gulped. Although FitByte does not estimate the amount of food consumed, researchers have found that estimating food amount would depend on accurately detecting each intake gesture [2, 20]

FitByte makes intake-level recognition by averaging the confidence values of frame-level inferences with a 5-second window and setting a threshold at 0.5 overall confidence. We then drop detected intakes that were less than 3 seconds long. In our evaluation we use the coverage and the delay metrics. The **coverage** can be defined as the percentage of the event's duration that was correctly recognized. The **delay** is the time between the beginning of the event and the time the system starts to recognize it.

Episode-level Recognition

From the user's perspective, to maintain their food journal, they mainly need to note each meal or snack or drink. We call these events "episodes." We assume that two consecutive food episodes will be separated by at least 5 minutes. We compute the duration of the episode by merging any detected intakes that are within 5 minutes from each other.

Identifying Food Type

FitByte does not directly detect food types. It aids the users in recalling their foods and drinks by showing them opportunistically recorded visuals of foods. We use the information from other sensors to detect an opportune moment to capture the visuals from the camera. This reduces the user's information load. For each food consumption episode, we identify the moment when FitByte is most confident of its inference. We initially experimented with simply taking a still photograph at the right moment. However, it is often hard to ensure that the image is not blurry or occluded. Thus, starting at the moment of high confidence in inference, we extract a 30 second video clip from the camera. These videos can be shown to the user after the food consumption episode or at the end of the day to recall the actual food. The same footage can also be labeled by crowd workers or a machine learning model to further automate the overall process. In our current evaluation, we simulated the crowd workers scenario by employing 2 independent research volunteers to label food types from the extracted video clips. The crowd workers had the option of looking at a thumbnail of the extracted clip to label it or watch the video in case they were not sure.

DATA COLLECTION AND SYSTEM EVALUATION

We conducted the data collection and evaluation of FitByte in two separate studies. In the first study, we collected a dataset from a set of short common everyday activities to build models for eating and drinking detection and evaluated the performance of sensor combinations. In the second study,

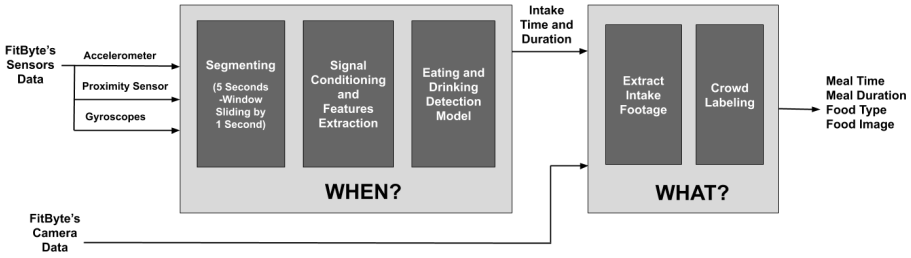

Figure 5: FitByte's machine learning pipeline

we assessed the ecological validity of FitByte by testing the developed models on a new 91 hours dataset collected in the unconstrained free-living environment. We also did a preliminary investigation on the perceived privacy and social acceptability aspect of the system.

Scripted Semi-Constrained Study

Evaluating a diet monitoring system in unconstrained situations is often done by running long sessions that extend from a few hours to a whole day. This is done to ensure that the participant encounters enough noisy situations and eats at will, at their pace. Building robust eating detection models require fine-grain annotations of all activities during these long sessions (mostly done by recording video footage of the session). This approach requires laborious labeling effort and is usually limited by the battery life of the recording device [4, 11]. Thus, instead of asking the user to wear FitByte for extended periods, we put them in noisy situations and got concentrated usage of the device.

Study Design

In this study, the participants performed five different activities (one in each session): a lunch meeting, grabbing and consuming snacks from a nearby cafe, exercising, hiking, and watching TV (Figure 1). These situations were chosen to ensure the participants get to talk, walk, encounter noisy situations, and eat food of their choice, at their pace, in a real-world setting.

When participants came in, they wore FitByte, and the researcher helped them adjust the temple tips for fitting, comfort, and snugness. Each session lasted for 15 to 30 minutes. After setting up the device, participants had the freedom to perform the session alone (except for the lunch meeting) or in the company of one of their friends or colleagues.

We did not restrict any of the activities to a specific place. The snack break consisted of walking to a cafe or a nearby store, buying and consuming a drink and/or a snack. The participants watched TV in a home environment, where they had the choice of snacks and beverages during the session. They exercised in an on-campus gym, or at their house. Lastly, only hiking *required* the participant to walk throughout the entire session, and it was conducted in either a park or the CMU campus lawns.

For each activity, we collected ten sessions from 5 males and 5 females participants (18 to 36 years old). Not all participants were able to perform all five activities due to time constraints. No external cameras were used to record participants' actions in this study. The only camera used is FitByte's built-in camera, and it was set to run on video mode throughout the session. We assessed the footage from this camera during annotation to identify the participant's activities.

Annotations

To annotate the dataset, we used Elan 5.2[4]. Two researchers labeled the dataset and a third researcher reviewed the annotations. Using the videos and audio obtained from the on-board camera, we labeled all activities in a session at a 1 second resolution. The activities were annotated as either eating, drinking, talking, motion/walking, or silence. We segment bites and chews into separate intakes by assuming that any chewing, or swallowing separated by more than 5 seconds belongs to different intakes. For eating, the intake ends when the participants stop chewing. For drinking, the intake ends after 1 second of the user bringing their hand down or as soon as the participant starts talking.

Free-living Environment Unconstrained Study

In this study, we aim to evaluate the performance of FitByte for an extended period of time in the real world without any constrains on the participant's behavior.

Study Design

We asked participants to wear FitByte continuously for 12 hours a day for as many days they can. Due to the small battery, the onboard camera can only record videos for a limited duration. Thus, for ground truth, we used an external camera similar to the onboard one and attached it to the participant's shirt. The camera faced upward to capture the participant's face. We powered this external camera with an external battery kept in the participant's pocket (4000 mAh).

At the end of the study, we asked the participants about their perception of social acceptability and privacy implications of the device in a semi-structured interview. To ensure FitByte can run for more than 12 hours using an onboard battery, we implemented the real-time version of the machine learning algorithm. We developed this algorithm based on data collected in the first study.

[4]Elan. https://tla.mpi.nl/tools/tla-tools/elan/download/

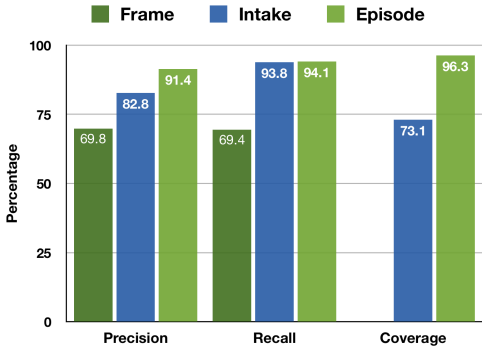

Figure 6: Eating Detection Results for the semi-constrained study. Frame-level results will not have coverage because those inferences are already made at a 1 second resolution.

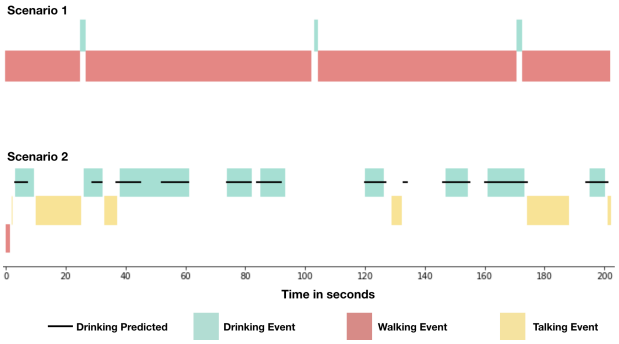

Figure 7: Shows timeline of two scenarios from the user study. *(Top)* FitByte fails to detect drinking activity when the user occasionally sips liquid while walking. *(Bottom)* However, FitByte succeeds in detecting drinking episodes when the user drinks for longer and drinking is not completely occluded by other activities.

To evaluate the real-time version, we recruited 5 participants (1 female), age between 21-30 years, all university students. Three participants wore the device for two days and two for one day. All session recordings lasted for 12 hours except P5. With P5, the prototype malfunctioned and we had to end the study after 7 hours. In total, we collected 91 hours of free-living data.

Participants started the study at different times in the morning (between 8 am and 11 am) and took it off 8 or 12 hours later. The dataset contains a very diverse set of activities across different participants, which included cooking, driving, working in a chemical lab, working in an office, laying down, taking public transports, grocery shopping, exercising in a gym and many more.

Annotations

The annotation process was similar to the short term study. Since collected data is used as a test dataset, annotations were only made for eating and drinking instances and every other activities were considered part of the null class. The external mini camera footage was used as ground truth for participants' activities. All annotations were done by one member of the research team and reviewed by a second member.

RESULTS

In this section we present FitByte's performance with regards to detecting eating and drinking at the frame, intake, and episode level. We will also discuss the results for an evaluation we ran with 2 volunteers to recognize the food type from video segments automatically generated by FitByte.

Eating Detection

We conducted a user-independent evaluation for detecting eating instances with all sensors on FitByte. At a **frame-by-frame level** (*i.e.,* every 1 second) the system achieved 83.1% accuracy. When aggregating and filtering the results to **intake level**, the system obtained 93.8% recall, 82.8% precision (considering intake detection is a binary task, recall and accuracy are same). The average coverage for these intake events (*i.e.,* the intake duration detected by the model) is 73.1% and the mean delay in detecting the beginning of the event is 2.5 seconds(the mean intake duration in the annotated videos was 56.3 seconds). The intake-level inferences are useful to quantify the amount of food consumed. At the **episode level**, the system was able to detect 32 out of the 34 eating events in the data set and only 3 falsely recognized episodes. The overall mean coverage for detected activities was 96.3% and the average delay was 6.5 seconds (the average duration for eating events was 304.3 seconds). Figure 6 provides a summary of the results.

Drinking Detection

For identifying drinking episodes, the system obtained 64.5% recall and 56.7% precision at the intake level. On investigating the reason for significantly low performance as compared to eating, we found that drinking in unconstrained situation happened in three different ways – either sporadic, short sips of liquid, mixed with other noisy activities (especially while hiking), or more continuous drinking events where the user took more than small sips with some sporadic noisy activity (*i.e.* series of short sips,or along sustained series of gulps) For example, having a coffee while reading a book at a cafe. Figure 7 shows an actual scenario from our data collection for the two cases. While FitByte fails at detecting situations like Scenario 1 in Figure 7, it very accurately identifies events similar to Scenario 2 (7 episodes in the dataset) where the duration between sips does not exceed 30 seconds.

Considering our goal here is to assist the users in maintaining their food journal, we also considered combining the eating and drinking results to assess the ability of detecting *food consumption* events. Even here, FitByte would still capture a mixed eating and sporadic drinking event as food consumption and would provide a footage of the episode that would contain both activities. In this case, our food consumption episode classifier obtains 97.5% recall and 92.8% precision.

Identifying Food Type

We triggered the camera using FitByte's IMUs and proximity sensors to capture food visuals (Figure 8). To assess the efficacy of our automatic trigger for the camera, we recruited two volunteers to identify food type from video snippets generated by FitByte. From each episode that was classified as eating or drinking, we generated 2 video snippets and showed them

to the volunteers. The volunteers viewed the first 10 seconds of the video and identified food type. If they were not sure about the food type, they were presented with 2 options; either to continue watching the video for up to 30 seconds, or move on to the next video. Each volunteer assessed 20 randomly-sampled sessions. For all 40 trials, the volunteers were able to correctly identify the food type for 37 trials. We found that all misclassified videos had extremely low lighting or significant occlusions by the hand. In general, the results indicated that FitByte can be used to effectively recall meals and snacks at the end of the day by quickly scrubbing through the captured videos. Sample videos can be seen here: Video 1; Video 2; and Video 3[5].

Sensor Selection

FitByte uses multiple sensors for diet monitoring. While these sensors help in accurately identifying eating moments and food types, they are probably also an overkill. We decided to have all the sensors on the initial prototype to provide the necessary redundancy for analysis. Thus, we investigated how different sensors contribute in the end. Instead of investigating the contribution of individual features in the machine learning model, we developed different models with a subset of sensors. We did not change any hyper-parameters or tried to tune them as the goal was not to formally benchmark each sensor. Figures 9 shows the comparison of the performance of different sensing modality and the combination of all sensors. It is evident that the 4 kHz accelerometer was the best performing sensor, and the proximity sensor was the worst. However, none of the three sensors can beat the performance of combining all their data together. When reviewing cases where individual modalities fail, our findings corroborated with past research (*i.e.*, the modality that detects chewing (gyroscopes) fails in detecting drinks and the proximity senor produces false positives from undesired hand-to-mouth gestures). Although the proximity sensor performs worst in comparison to other sensors in isolation, when used *with* other sensors (Figures 10), this sensor is important and an important first line of defense. It acts as a low-power trigger for other costlier sensors. We can see evidence of this claim in the improved performance for sensor combinations that include the proximity sensor. The combination of accelerometer, gyrocope behind the ear, and proximity sensor gives the highest accuracy among all other combinations (Figure 10). This shows that by using one IMU (accelerometer+gyroscope) and a proximity sensor we can capture food consumption moment with an accuracy close to combinations of all sensors.

[5]If a video link does not work, please contact the first author at: bedri@cmu.edu

Figure 8: Samples from FitByte's on-board camera for food intake moments.

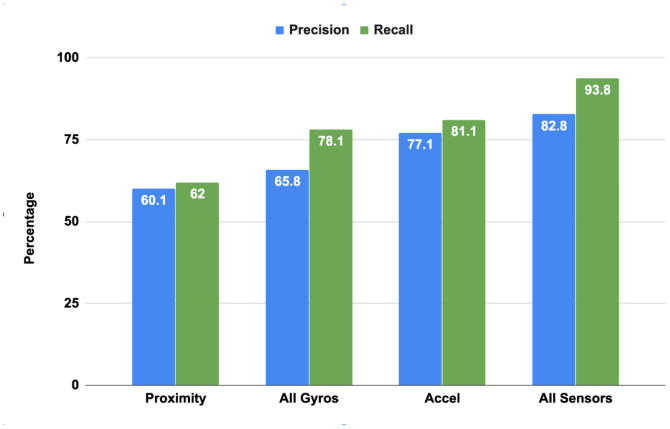

Figure 9: Performance of different sensing modalities compared to the performance of all sensors in the semi-constrained Study

Real-time Implementation

Informed by the outcomes of the first study, we made modifications to FitByte to improve its battery life and make it practical for real-world applications. The modifications include changes to the hardware design and introducing a policy for sensor activation. These changes enabled the system to run for a day on the onboard battery without a recharge. To reduce FitByte's power consumption, we made the system so that it only uses a single temple gyroscope (bottom-right), nose-bridge gyroscope, accelerometer, proximity sensor, and camera.

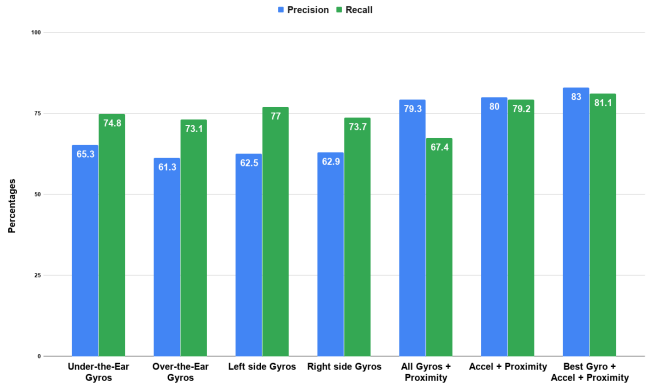

Figure 10: Performance of different subsets of sensors in the semi-constrained Study

Moreover, we noticed that most of the activities contained signal in the 0-1 kHz band. Therefore, we decided to sample the accelerometer at 2 kHz, and we also reduced the sampling rate for other sensors to half because the sensed activities (*i.e.,* chewing, walking, hand to mouth gestures) occur at less than 10 Hz frequency. We verified the validity of this approach by training and testing our food intake models on the downsampled version of the data set and the performance was largely unaffected. We also optimized the processor power consumption by enabling the deep sleep functionality and setting the processor clock to 16 MHz. All these steps helped in significantly reducing the overall power consumption.

With all these modifications, the overall measured current of the system, excluding the camera is 28.4 mA at 3.7 V during regular operation. When triggering the camera, the camera consumes 110 mA and the drawn current by the rest of the system jumps to 100 mA because the processor uses two I/O pins to control activation and recording of the camera. To make sure that the users can quickly browse through the videos, we restrict the captured video duration to a maximum of 2 minutes. If we assume the maximum number of triggers per hour (30 times) the camera will be active for 390 seconds per hour, which means the camera will draw 11.9 mA/h and rest of the system will draw approximately 36.1 mA/h. Thus, in the worst-case scenario, the system requires an 864 mAh battery to last for 18 hours. The FitByte prototype used in the first study had 300 mAh battery. To increase the charge capacity of the device, we removed the internal battery charging board and added 600 mAh in battery capacity. The final prototype had 900 mAh charge capacity without significantly changing the physical dimensions (3 mm increase in the arm width).

Unconstrained Evaluation in a Free-living Environment

To assess the performance of FitByte we evaluated the accuracy using the trained model from the scripted semi-constrained evaluation.

Using the same filtering parameters for in the short term evaluation, the system was able to detect 22 out of 28 episodes with 89% average coverage. The missed episodes were short (less than 10 seconds), and two of them are drinking episodes. The system had 4 false positives corresponding to silence and talking activities. On the intake level the system achieved 84.7% precision, 75.4% recall and 68.2% coverage and on the frame level it achieved 65.3% precision and 60.7% recall.

From the detected events, we extracted the associated short video footage captured by the Fitbyte camera and showed them to a crowd worker to identify the food type. On average the system triggered 122 times per session. We marked videos that were recorded during the event or close to it (5 seconds before or after an event) as videos of interest. We asked one crowd worker to visualize and identify all food types seen in the video. The selected videos ranged between 20 to 5 per session. From the 22 recognized food intake episodes, the crowd worker was able to identify the food type in 20 events correctly. Two events were not easily recognized because bad lighting conditions. Here are samples of the captured videos Video 1, Video 2 and this is a sample of a video with low lighting condition Video 3.

Privacy and Social Acceptability

As part of our evaluation, we did a preliminary assessment of FitByte's perceived privacy and social acceptability. We conducted semi-structured interviews with the five participants in the long-term study after they wore the device for a day or two in public. In general, participants thought the use of the eyeglasses form factor helped in making the device socially acceptable. People around them were either curious to know what does this special looking glasses do, or they were indifferent about it, but none of the participants reported any perceived feelings of discomfort from wearing the device in public. One participant mentioned wearing the device in a cafe and he said "I was surprised no one was looking at me. I ordered my coffee and the cashier did not ask me about it" Another participant mentioned "When I was walking around on campus people stopped and asked me what are the special glasses for? I think they probably noticed it's 3D printed and has no lenses on it". All participants said that they would wear a device like FitByte if they get to customize its look to fit their style. When asked about future changes they would like to see in FitByte, most participants mentioned they would prefer if the device has thinner temples (or arms) and lighter weight. Regarding privacy concerns, participants mentioned that the placement of the on-board camera made wearing it in public less concerning, mainly because the lens is looking down to the side of the user's face and not to the front. Participants mentioned that people did not notice there is a camera unless the participant mentions it. One participant said "My wife asked me where is the camera looking at? After I showed her it was looking to the side of my face, she was fine with it". Another participant said, "If someone sits very close to my left side, I would mention that I'm wearing a camera, otherwise I see no need to bring it up". In addition, all participants expressed that they would prefer to have a way to manually turn the camera off in case they do not want to record clips during a specific activity. Also, two participants said they would prefer that the system would detect eating or drinking first before turning the camera on to ensure that it's only recording when they need it to.

DISCUSSION

FitByte was able to detect almost all eating events, irrespective of the amount of noise. The eyeglasses were able to recognize that the user was eating, on average, in 6.4 seconds. Thus, FitByte can enable fast notifications or interventions (*e.g.,* remind a person with diabetes to not eat a donut). Moreover, FitByte also accurately (96.3%) detected the duration of the eating episode and the number of intakes (93.8%). Using the performance of each individual sensor as a proxy for performance for the corresponding phenomenon (*e.g.,* proximity sensor for hand-to-mouth gesture), it is evident that combining multiple modalities outperforms individual ones (Figure 9 and 10). Besides, FitByte can capture visuals of the food in a privacy-preserving way. These visuals allow users to recall more important details about the event like food type, food amount, location, and the social context. We also showed that the captured footage was sufficient for crowd workers to identify food types in almost all cases despite a few challenges with lighting conditions.

Drinking Detection

Drinking can be a single sip, a series of short sips, or a long sustained series of gulps (e.g., chugging). FitByte can reliably detect the latter two, but detecting a single sip is hard as it is a very short event. FitByte fails to detect sporadic drinking while moving or talking, but it reliably detects repeated sips, as long as the sips are within 30 seconds. If the user drinks and moves or talks at the same time, the high-speed accelerometer gets inundated by noise (surface noise due to motion artifact or bone conduction due to speech) making it difficult to detect swallowing instances.

Ranging Sensor

For the realtime implementation, we introduce a set of modifications to the hardware to help improve battery life. This approach enabled the system to run for 16.5 hours on a single charge, which highlights the potential of using FitByte for everyday use. During an initial pilot, we found that the camera triggered with a rate of 20 times every hour. Upon investigation, we found that the proximity sensor (VCNL4040) was susceptible to ambient light changes, mainly when a user used their phone or computer. To address this issue, we added another moving-average filter (size=10 samples) to the proximity signal. The filter reduced the number of false triggers, but in the future, a time-of-flight ranging sensor will be better.

Privacy and Social Acceptability

Systems with a wearable camera usually raise privacy concerns for users and bystanders. Google Glass is a popular example of that. Although several precautions were taken in its design to ensure that the camera is not recording without a clear indicator to the user and bystanders, the ability to hack the device and record video and audio without consent has been a major concern for customers and governments [26, 31]. In our design, we tried to approach this challenge by eliminating some of the sources of concern. We removed the microphone from the camera module to ensure no audio is recorded and we pointed the camera downwards to only capture the user's mouth. We did a preliminary investigation of the perceived social acceptability and privacy implications of the device with participants. The outcome of this short investigation indicated that users and bystanders are generally tolerable to the on-board camera once they know it points at the wearer's mouth region and is not recording audio. In the future, we plan to more deeply investigate the privacy and social aspects of FitByte with a large and diverse group of users and bystanders.

FitByte Design

The design process involved building several iterations of the device and testing them with a diverse group of participants. One of the major trade-offs was in the placement of the 4 kHz accelerometer. Placing the sensor closer to the center of the throat provides the best swallowing signal, but having a sensor extend outside the glasses frame to the throat was socially unacceptable. Thus, we experimented with several locations around the ear and nose and found locations below the ear (B and C in Figure 2) to give a reasonable swallowing signal as seen in Figure 4.

Beyond Food Journaling

Overall, FitByte has usable performance for most aspects of diet monitoring. It helps the wearer in recalling what they consumed while maintaining a single commonplace form factor. Moving forward, we are working on deploying FitByte with different populations (*e.g.,* multiple sclerosis patients to detect depression and fatigue, obese teenagers, pregnant women). In conjunction with diet monitoring, we are also working on measuring physiology (such as sleep quality, blood pressure, instantaneous spikes in glucose levels) and better understand the relationship between diet and disease.

CONCLUSION

In this paper, we introduced FitByte — a one-size-fits-all pair of eyeglasses that tracks when a user eats or drinks, and automatically takes a video snippet of their food and aid identifying food type. We conducted two studies to evaluate FitByte's performance outside the lab. Fitbyte recognizes eating episodes with 94.1% accuracy and detects that the user is eating withing 7 seconds and estimates the duration of the eating episode with 96.3% accuracy. Using the footage captured by the on-board camera, volunteers were able to identify food type 37 of 40 eating and drinking episodes. When tested for multiple hours during the day, Fitbyte demonstrated similar performance. Participants who wore the device for extended periods during the day indicated that the eyeglasses form factor made the system more socially acceptable to wear in public. We hope FitByte becomes a catalyst in ensuring smart glasses of the future support automatic diet monitoring; similar to how first-generation wrist-worn computers supported step counting.

ACKNOWLEDGEMENTS

We are grateful to the Qualcomm Innovation Fellowship, Tata Consultancy Services, and Bosch for supporting this research. We also thank Nur Yildrim, Judy Kong, Jessica Wallace, Lynn Kirabo and Nelson Wong for their help in building the prototypes.

REFERENCES

[1] O. Amft, H. Junker, and G. Troster. 2005. Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*. 160–163. DOI:http://dx.doi.org/10.1109/ISWC.2005.17

[2] O. Amft*, M. Kusserow, and G. TrÖster. 2009. Bite Weight Prediction From Acoustic Recognition of Chewing. *IEEE Transactions on Biomedical Engineering* 56, 6 (June 2009), 1663–1672. DOI: http://dx.doi.org/10.1109/TBME.2009.2015873

[3] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of Chewing Sounds for Dietary Monitoring. In *Proceedings of the 7th International Conference on Ubiquitous Computing (UbiComp'05)*. Springer-Verlag, Berlin, Heidelberg, 56–72. DOI:http://dx.doi.org/10.1007/11551201_4

[4] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner,

and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article Article 37 (Sept. 2017), 20 pages. DOI:http://dx.doi.org/10.1145/3130902

[5] Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner. 2015a. Detecting Mastication: A Wearable Approach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 247–250. DOI:http://dx.doi.org/10.1145/2818346.2820767

[6] Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner. 2015b. A Wearable System for Detecting Eating Activities with Proximity Sensors in the Outer Ear. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. Association for Computing Machinery, New York, NY, USA, 91–92. DOI: http://dx.doi.org/10.1145/2802083.2808411

[7] Shengjie Bi, Tao Wang, Ellen Davenport, Ronald Peterson, Ryan Halter, Jacob Sorber, and David Kotz. 2017. Toward a Wearable Sensor for Eating Detection. In *Proceedings of the 2017 Workshop on Wearable Systems and Applications (WearSys '17)*. Association for Computing Machinery, New York, NY, USA, 17–22. DOI:http://dx.doi.org/10.1145/3089351.3089355

[8] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, and et al. 2018. Auracle: Detecting Eating Episodes with an Ear-Mounted Sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article Article 92 (Sept. 2018), 27 pages. DOI: http://dx.doi.org/10.1145/3264902

[9] Y. Bi, M. Lv, C. Song, W. Xu, N. Guan, and W. Yi. 2016. AutoDietary: A Wearable Acoustic Sensor System for Food Intake Recognition in Daily Life. *IEEE Sensors Journal* 16, 3 (Feb 2016), 806–816. DOI: http://dx.doi.org/10.1109/JSEN.2015.2469095

[10] Junghoon Chae, Insoo Woo, SungYe Kim, Ross Maciejewski, Fengqing Zhu, Edward J Delp, Carol J Boushey, and David S Ebert. 2011. Volume estimation using food specific shape templates in mobile image-based dietary assessment. In *Computational Imaging IX*, Vol. 7873. International Society for Optics and Photonics, 78730K. DOI: http://dx.doi.org/10.1117/12.876669

[11] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. 2018. Detecting Eating Episodes by Tracking Jawbone Movements with a Non-Contact Wearable Sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article Article 4 (March 2018), 21 pages. DOI:http://dx.doi.org/10.1145/3191736

[12] Jungman Chung, Jungmin Chung, Wonjun Oh, Yongkyu Yoo, Won Gu Lee, and Hyunwoo Bang. 2017. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Scientific reports* 7 (2017), 41690. DOI: http://dx.doi.org/https://doi.org/10.1038/srep41690

[13] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover. 2014. Detecting Periods of Eating During Free-Living by Tracking Wrist Motion. *IEEE Journal of Biomedical and Health Informatics* 18, 4 (July 2014), 1253–1260. DOI:http://dx.doi.org/10.1109/JBHI.2013.2282471

[14] M. Farooq and E. Sazonov. 2017. Segmentation and Characterization of Chewing Bouts by Monitoring Temporalis Muscle Using Smart Glasses With Piezoelectric Sensor. *IEEE Journal of Biomedical and Health Informatics* 21, 6 (Nov 2017), 1495–1503. DOI: http://dx.doi.org/10.1109/JBHI.2016.2640142

[15] J. M. Fontana, M. Farooq, and E. Sazonov. 2014. Automatic Ingestion Monitor: A Novel Wearable Device for Monitoring of Ingestive Behavior. *IEEE Transactions on Biomedical Engineering* 61, 6 (June 2014), 1772–1779. DOI: http://dx.doi.org/10.1109/TBME.2014.2306773

[16] David R. Jacobs. 2012. *Challenges in Research in Nutritional Epidemiology*. Humana Press, Totowa, NJ, 29–42. DOI: http://dx.doi.org/10.1007/978-1-61779-894-8_2

[17] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G. Yang. 2012. An Intelligent Food-Intake Monitoring System Using Wearable Sensors. In *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*. 154–160. DOI: http://dx.doi.org/10.1109/BSN.2012.11

[18] Christopher Merck, Christina Maher, Mark Mirtchouk, Min Zheng, Yuxiao Huang, and Samantha Kleinberg. 2016. Multimodality Sensing for Eating Recognition. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '16)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 130–137.

[19] Mark Mirtchouk, Drew Lustig, Alexandra Smith, Ivan Ching, Min Zheng, and Samantha Kleinberg. 2017. Recognizing Eating from Body-Worn Sensors: Combining Free-Living and Laboratory Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article Article 85 (Sept. 2017), 20 pages. DOI: http://dx.doi.org/10.1145/3131894

[20] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 2016. Automated Estimation of Food Type and Amount Consumed from Body-Worn Audio and Motion Sensors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 451–462. DOI:http://dx.doi.org/10.1145/2971648.2971677

[21] T. Olubanjo and M. Ghovanloo. 2014. Real-time swallowing detection based on tracheal acoustics. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 4384–4388. DOI:http://dx.doi.org/10.1109/ICASSP.2014.6854430

[22] Temiloluwa O Olubanjo. 2016. *Towards automatic food intake monitoring using wearable sensor-based systems*. Ph.D. Dissertation. Georgia Institute of Technology.

[23] T. Prioleau, E. Moore II, and M. Ghovanloo. 2017. Unobtrusive and Wearable Systems for Automatic Dietary Monitoring. *IEEE Transactions on Biomedical Engineering* 64, 9 (Sep. 2017), 2075–2089. DOI: http://dx.doi.org/10.1109/TBME.2016.2631246

[24] S. A. Rahman, C. Merck, Yuxiao Huang, and S. Kleinberg. 2015. Unintrusive eating recognition using Google Glass. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. 108–111. DOI:http://dx.doi.org/10.4108/icst.pervasivehealth.2015.259044

[25] Tauhidur Rahman, Alexander T. Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. 2014. BodyBeat: A Mobile System for Sensing Non-Speech Body Sounds. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '14)*. Association for Computing Machinery, New York, NY, USA, 2–13. DOI: http://dx.doi.org/10.1145/2594368.2594386

[26] Seyedmostafa Safavi and Zarina Shukur. 2014. Improving Google glass security and privacy by changing the physical and software structure. *Life Science Journal* 11, 5 (2014), 109–117.

[27] Giovanni Schiboni and Oliver Amft. 2018. *Automatic Dietary Monitoring Using Wearable Accessories*. Springer International Publishing, Cham, 369–412. DOI: http://dx.doi.org/10.1007/978-3-319-69362-0_13

[28] S. Sen, V. Subbaraju, A. Misra, R. Balan, and Y. Lee. 2018. Annapurna: Building a Real-World Smartwatch-Based Automated Food Journal. In *2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*. 1–6. DOI: http://dx.doi.org/10.1109/WoWMoM.2018.8449755

[29] Edison Thomaz, Irfan Essa, and Gregory D. Abowd. 2015. 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 (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1029–1040. DOI: http://dx.doi.org/10.1145/2750858.2807545

[30] Edison Thomaz, Aman Parnami, Irfan Essa, and Gregory D. Abowd. 2013. Feasibility of Identifying Eating Moments from First-Person Images Leveraging Human Computation. In *Proceedings of the 4th International SenseCam Pervasive Imaging Conference (SenseCam '13)*. Association for Computing Machinery, New York, NY, USA, 26–33. DOI: http://dx.doi.org/10.1145/2526667.2526672

[31] Michael S. Wagner. 2013. Google Glass: A Preemptive Look at Privacy Concerns Student Note. *Journal on Telecommunications High Technology Law* 11 (2013), 477. https://heinonline.org/HOL/P?h=hein.journals/jtelhtel11&i=505.

[32] Chang Xu, Ye He, Nitin Khannan, Albert Parra, Carol Boushey, and Edward Delp. 2013. Image-Based Food Volume Estimation. In *Proceedings of the 5th International Workshop on Multimedia for Cooking Eating Activities (CEA '13)*. Association for Computing Machinery, New York, NY, USA, 75–80. DOI: http://dx.doi.org/10.1145/2506023.2506037

[33] Koji Yatani and Khai N. Truong. 2012. BodyScope: A Wearable Acoustic Sensor for Activity Recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 341–350. DOI: http://dx.doi.org/10.1145/2370216.2370269

[34] R. Zhang and O. Amft. 2018a. Free-living eating event spotting using EMG-monitoring eyeglasses. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. 128–132. DOI: http://dx.doi.org/10.1109/BHI.2018.8333386

[35] R. Zhang and O. Amft. 2018b. Monitoring Chewing and Eating in Free-Living Using Smart Eyeglasses. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (Jan 2018), 23–32. DOI: http://dx.doi.org/10.1109/JBHI.2017.2698523

[36] R. Zhang, S. Bernhart, and O. Amft. 2016. Diet eyeglasses: Recognising food chewing using EMG and smart eyeglasses. In *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 7–12. DOI: http://dx.doi.org/10.1109/BSN.2016.7516224