Enabling Automatic Diet Monitoring Systems in Real-World Settings

Abdelkareem Bedri

February 2021

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Mayank Goel, ISR and HCII , CMU (Chair)
Jeffrey P. Bigham, LTI and HCII, CMU
Geoff Kaufman, HCII, CMU
Edison Thomaz, University of Texas at Austin

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Keywords: diet monitoring, eating detection, drinking detection, food journaling, wearable, earables, activity recognition, in-the-wild, free-living, automatic diet monitoring, smart glasses, just-in-time intervention, food type, and food amount

To the women in my life, Asma my dear wife and my mother Saadia.

Abstract

Chronic diseases such as diabetes, heart failure, and obesity are widespread globally. These diet-related diseases are mainly caused due to limited physical activity and poor eating patterns. Journaling and self-monitoring have been very effective tools in combating diet-related diseases as they help to discover undesired patterns at an early stage and motivate users to lead a healthy lifestyle.

Smartwatches are commonly used for fitness tracking. They can recognize different types of physical exercises and provide rudimentary measurements for health metrics such as heart rate variability, energy expenditure, and sleeping hours. While useful, these features do not provide users with a holistic view of how their daily activities influence their health and how their body reacts. For example, knowing how many calories we burn is insufficient unless we compare it to our calorie intake.

Current food journaling methods rely heavily on self-report, which suffers from self-bias, recall errors, and low adherence. In the last two decades, researchers have developed several automatic diet monitoring (ADM) systems to address the challenges of traditional journaling techniques. The focus of the diet monitoring research has been on detecting **when** people eat, and identifying **what** and **how much** they ate. Ecological validity has been a major issue in ADM research. While many ADM systems obtain high accuracy in lab settings their performance drops significantly when tested in the real world. The most cited reasons for this challenge are the difficulty to build generalizable models using data collected in the lab, the lack of reliable ground truth in free-living environments, privacy concerns, and the social acceptability of the device.

In my research, I tackle these challenges by developing and deploying a number of ADM systems (**EarBit** and **FitByte**). These trackers are hosted in commonplace form factors (i.e. headphones and eyeglasses) to ensure their social acceptability. I also worked on designing data collection techniques to build models that work reliably in the real world. The high performance obtained by these models has brought us closer to assessing the utility and usability of ADM systems in the field.

The final piece of my dissertation is a long-term field deployment for an ADM system (FitNibble) based on my previous ADM designs. In this study, I compared traditional self-report journaling and journaling with ADM. Through this evaluation, I assessed the factors influencing adherence to journaling like reducing missed events, social acceptability, usability, utility, and privacy concerns. Results have shown that *FitByte2.0* improved adherence by significantly reducing the number of missed events ($19.6\% improvement, p = .0132$). Results have shown that participants were highly dependent on the wearable in maintaining their journals. Participants also reported an increase in their awareness of their dietary patterns especially with snacking. All these results highlight the potential of ADM in improving the food journaling experience.

Acknowledgments

First, I would like to thank my advisor Mayank Goel for his guidance and support through these five years. I would also like to thank my Ph.D. committee for helping me navigate the last phase of my dissertation research. Thanks to all my collaborators Thad Starner, Gregory Abowd, Edison Thomaz, Temiloluwa Prioleau, Daina Li, Kunal Bhuwalka, Apoorva Verlekar, Valerie Avva, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Min Yan Beh, Sai Swaminathan, Scott Hudson, Sudershan Boovaraghavan, and Yuchen Liang.

My family and friends thank you for your endless support throughout this long journey. Finally, I would like to thank my research sponsors: Intel Labs, Google Research, Qualcomm, the CMU president, and TATA Consulting Services (TCS).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

What we eat has an immediate and long lasting affects on our health. Medical research literature has shown direct links between diet and chronic illnesses. On the flip side, a good healthy diet has always been helpful in fighting and preventing diseases [15]. A report published by the US Burden of Disease Collaborators showed that in 2016 dietary factors were associated with more than half a million deaths in the United States, putting it as a leading risk factor for mortality[47]. In the United States, almost half of the adult population suffer from at least one diet-related disease, such as heart failure, diabetes, high blood pressure, cancer, kidney failure, and obesity[23]. However, these diseases are no longer confined to resource-rich environments, diabetes and obesity are now considered global epidemics [33]. Diet-related diseases are mainly caused by limited physical activity and poor eating habits. To alleviate some of these issues, numerous research efforts have shown that journaling is one of the most effective ways to maintain a healthy eating regime [15]. Self-monitoring allows us to keep track of our physical and eating activities, understand how our bodies are reacting to these events, and open the door for us to reflect on our behavior and lead a healthy lifestyle. In the midst of our busy lifestyle, adherence to self-monitoring becomes very challenging. Many studies have shown low adherence rates to journaling due to the the tedious nature of the process[31]. As part of the quantified-self movement, academic and industrial communities have been actively trying to provide solutions to automate the journaling process, to raise adherence and improve usability. Currently, all smartwatches and smartphones come equipped with fitness tracker apps, while tracking physical activity is becoming increasingly popular, diet tracking is often overlooked. Diet monitoring is a hard problem. It involves recording when you eat, what you ate, and how much you're eating. This makes diet monitoring more challenging to automate than other activities. It is well known that diet has a higher influence on our well-being than physical exercise [73]. Therefore, there is a real need for solutions that can address the food journaling challenges.

1.2 Problem Statement

Food journaling requirements vary depending on what users and dietitians are interested to track. It involves tracking the eating event's time, location, food type, food amount, social context, mood, and calorie content. Current Food journaling solutions rely heavily on self-report either by using pen-and-paper forms like food frequency questionnaires and 24-hour recalls or through journaling applications. A recent study has shown that the majority (98%) of journalers use smartphone applications or web applications to log their dietary activities[20]. While the use of journaling applications has enhanced the user experience, it didn't address the major challenges with self-monitoring. Helander et al. analyzed logs from a food journaling mobile app and found only 3% of 190,000 downloads resulted in a person using the app for more than a week [31].

Cordeiro et.al.[21] Investigated the reasons behind the low adherence to food journaling. In their study, participants have cited the following reasons to explain why they stopped journaling before they reach their goals.

1. Requires too much effort.
2. Time-consuming.
3. Loss of motivation.

Generally, missing to log eating events results in incomplete journals making it difficult for users to understand their dietary patterns, and as a consequence, they lose their motivation and quit journaling. In the same evaluation, Cordeiro et.al found that participants missed logging because they simply forgot. Other reasons included lack of food nutrients information, stigma from journaling in front of others, or because they feel ashamed of the unhealthy meal choices they made. Lately, photo-based food journaling has become a popular journaling technique. It reduces the logging effort to just snapping a picture. The user can review the photos and add more information at their convenience. While effective studies showed that it doesn't significantly reduce the number of missed events because the majority of users still forget to log or/and reported experiencing stigma when they try to take food pictures in front of friends [20].

1.2.1 Automatic Diet Monitoring

In the last two decades, activity recognition researchers have worked on developing automatic dietary monitoring systems (ADM) to help mitigate some diet monitoring challenges. ADMs have the potential to reduce missed events by reminding users to log when they forget, and cutting off the manual journaling effort. Research in the field of automatic dietary monitoring (ADM) is focused on answering the following three questions: (1) **When** do you eat, (2) **What** type of food do you eat, and (3) **How much** of it did you consume.

Identifying food intake moments serves as an initial step towards identifying food type and amount. Therefore, detecting **When** eating and drinking events occur has been the focus of most ADM research [56, 63]. Most explored ADM systems are based on wearable form factors, and they use different sensing modalities to detect actions like chewing to identify food intake events. Many of these systems have obtained high accuracy when tested in controlled settings, but their performance significantly dropped when tested in the real-world. Building models in the lab to accurately capture how eating and drinking occur in-the-wild has become a major challenge.

(a) (b) (c)

Figure 1.1: This figure summarizes my thesis contributions starting with EarBit (a) an ADM system based on a sports headphone, which detects chewing. To detect drinking and identify food type and the amount, I designed FitByte (b), which is an ADM system based on an eyeglasses form factor. FitByte detects chewing, swallowing, hand-to-mouth gestures, and take images of the food. Finally, based on the lessons learned from EarBit and FitByte, I developed FitNibble (c) to evaluate ADM utility and usability in food journaling. This end-to-end ADM system provides just-in-time notifications to prompt users to do their logs.

Building these models with in-the-wild data is also difficult due to the lack of reliable ground truth for user activity in unconstrained environments.

Schiboni and Amft [63] discussed several challenges in building ADM systems, most prominent are the ability of these systems to **perform well in unconstrained environments** and to have a **socially acceptable form factor** for everyday use. Therefore, removing these technical and non-technical barriers has become a necessary step to evaluate the ADM impact on food journaling.

In my research, I address these two challenges by developing wearable ADMs that have commonplace form factors like headphones, wristbands, and eyeglasses. To improve in-the-wild accuracy I employed sensors that are less prone to environmental noise and I designed special data-collection techniques to effectively build models that work reliably in free-living environments. Addressing these challenges placed me in a good position to assess the utility and usability of ADM.

In my work I try to address the following research questions (RQs):

- **RQ1:** How can we develop **practical** ADM systems that **replicate** in-the-lab performance **outside** the lab?
- **RQ2:** Can ADM systems improve food journaling **compliance** and reduce **journaling difficulty**?

1.3 My Approach

In the effort to address **RQ1** and **RQ2** I started by developing EarBit (figure 1.1.a [11]), an ADM system based on a commonplace form factor, a sport's headphone. This ADM detects chewing with an IMU and use it as a proxy to detect eating. To build a naturalistic eating detection model we collected data in a home setting during a dinner party. When testing these models we found that they replicate their high lab performance outside the lab, but there were major challenges with this approach. First, Earbit fundamentally detects chewing, therefore we are missing on drinking events and eating semi-solids like ice cream and yogurt. Second, the way we collected the models' data was also difficult to replicate and participants found that using off-the-shelf wearable cameras is very restricting in free-living environments. Finally, EarBit takes on average 65 seconds from the beginning of an eating event to detect it. This delay means EarBit might not be suitable for just-in-time interventions.

In my next steps, I tried to address the challenges with EarBit by designing a new ADM system that can detect chewing, swallowing, hand-to-mouth gestures, and capture images of the food. I call this ADM system FitByte (figure 1.1.b [12]). The main challenge with this design was being able to capture all these actions from a single commonplace wearable. To achieve that I used a suite of sensing modalities including gyroscopes, proximity sensors, and a high-speed accelerometer. We used the onboard camera to also collect the ground truth of the user's activity in a free-living environment. We built our models using data collected while the user is completing a high-level task in free-living environments. These tasks require the user to engage in a variety of activities condensed in a short duration. With FitByte we were able to replicate high lab performance and accuracy outside the lab. We found that using the footage captured by the onboard camera can help us identify food types. FitByte was also able to recall 96% of the eating event duration, which helped us in estimating the food amount (Chapter 5). Finally, FitByte took only 7 seconds on average to detect the beginning of eating events, which makes it suitable for designing just-in-time interventions. This improvement in the average delay is attributed to the use of a proximity sensor to detect hand-to-mouth gestures

The findings of my research addresses **RQ1** and enables users to precisely track when eating moments occur. My approach also uncovers other aspects of food journaling like drinking events, food type, and food amount. All was done by collecting data in semi-controlled settings to build activity models, and by employing sensors commonly found in commercial wearables and smartphones. In the final part of my thesis work, I focused on addressing the second research question **RQ2** and make an initial assessment on the impact of ADM on the food journaling process.

In this investigation, I relied on Earbit and FitByte to inform the design of a new end-to-end ADM system (FitNibble) suitable for a field deployment (Figure 1.1.c). The setup used lightweight sensors attached to the user's glasses. The wearable has a BLE module that sends computed features to the user's phone which passes them to a server to generate eating predictions and pass it back to the phone. Using an iOS App FitNibble sends to the user just-in-time notifications to prompt the user. After deploying the system we found that FitNibble significantly improved compliance ($p = .013$) and reduced journaling difficulty ($p = .021$). These results directly address **RQ2** and demonstrate how ADM can improve the food journaling experience.

Through my research, I provided support for my thesis statement: ***Automatic diet monitoring enabled by wearable sensors can improve compliance to food journaling, by lowering the cognitive load required by users, and dropping the number of missed eating episodes.***

1.4 Document structure

In this document I use the terms *free-living environments*, *in-the-wild*,*in the real world*, and *unconstrained environments* interchangeably to describe uncontrolled, outside the lab studies. The next chapter provides some background on food journaling and reviews prior work in automatic diet monitoring. Chapter 3 discusses approaches I followed to address the first research question. Chapter 4 documents EarBit, my first attempt to address the free-living challenge. Chapter 5 describes my work on FitByte, an eyeglasses ADM capable of detecting eating and drinking events in-the-wild, captures images of food, and estimating food amounts. Chapter 5 discusses FitNibble, an ADM system based on FitByte design recommendations. In chapter 5, I also discuss how I used FitNibble to evaluate the utility and usability of ADM systems in food journaling. Finally, Chapter 6 is a general discussion about the findings of this research and possible future directions.

Chapter 2

Background

2.1 Food Journaling

Food journaling has been an effective method to combat diet-related diseases and help individuals lead a healthy lifestyle. Research has shown its positive effect in managing chronic illnesses such as diabetes [41, 42], kidney failure [17, 66], and obesity [32]. In addition, journaling plays an instrumental role in identifying and managing food allergies[30, 65]. In previous studies, researchers witnessed that journalers are more mindful of their diet and they are more encouraged to avoid unhealthy foods [34, 71].

2.1.1 Food Journaling Methods

Self-report is the most common diet monitoring method. Food frequency questionnaires and 24-hour recalls represent the most typical journalling methods employed by health experts. Self-report methods require users to keep track of many aspects related to their dietary activities such as when they eat, what they ate, the amount of food/drinks consumed, where did they eat, the social context, mood, and Calorie content [3, 25, 72]. In recent years, the smartphone has become a popular tool for food journaling; applications like MyFitnessPal and Weight Watchers have more than one million downloads. Most journaling apps help users populate their journal with the aid of a large food database that has calories and nutritional contents for meals.

2.1.2 Food Journaling Challenges

Despite the clear benefits of journaling it is not widely adopted. Food journaling requires a high level of engagement from the user to maintain their logs. The taxing nature of the journaling process causes fatigue and lead to reduced compliance [5, 22]. In [31] Helander *et al.*. found that of the 190,000 downloads of a food journaling app, only 3% used the app for more than a week. Cordeiro *et al.* [20, 21] have investigated the barriers and challenges for different food journaling methods and found that loss of motivation, time commitment, and the large effort required to maintain a journal are the most common reasons people cite to explain why they stopped journaling before they reach their goals. These reasons usually lead people to miss reporting eating events and overtime the journal loses its value because it is incomplete. When investigating further why journalers miss eating events, Cordeiro *et al.* found the following reasons to be the most common:

1. The most reported reason for missed events is forgetting to log.
2. The lack of dietary content information for a meal makes it difficult to report.
3. In some social contexts it's difficult to journal (e.g. in a party)and these events often get missed or partially logged.
4. People would intentionally miss logging some events because they feel ashamed or guilty of their unhealthy choices.
5. Stigma from journaling in front of friends is another reasoner why individuals deliberately miss to log.

2.1.3 Photo-Based Journaling

In [20] Cordeiro *et al.* propose the use of photo-based food journaling, which require the user to only collect images of the food they ate throughout the day and label these photos at a later time. Participants reported that photos were very helpful to recall food contents and the social context even if they just took a photo of their empty plates. This method reduced the number of missed events due to lack of food information by 38%. This is mainly because users were able to use the saved photos to lookup contents at their convenience. This method doesn't require users to count calories, which made participants less anxious and reduced missed events because of shame. While effective, photo-based journaling similarly suffers from some of the original challenges faced by traditional techniques like users forgetting to log and stigma.

Among all the reasons users reported for missed events forgetting to log is the most common. There is a clear need for methods that would help users recall to log eating events as soon as they happen. In the following sections, I introduce automatic diet monitoring as a potential solution to this problem and discuss its challenges.

2.2 Automatic Diet Monitoring

In the last two decades, researchers in the wearable community have developed many automatic diet monitoring systems (ADM) to help mitigate some of the challenges facing traditional journaling methods. Most of the ADM research has focused on detecting *when* eating events occur, while little have been done to automatically identify food contents [56, 63]. To identify eating moments, ADM systems use sensors to detect one or more actions that usually take place while eating like chewing, swallowing, and repetitive hand-to-mouth gestures. This section reviews ADM systems based on the actions they are designed to recognize.

2.2.1 Hand-to-mouth Gestures

Observing hand movements as a proxy to detect eating has been a well-studied approach. Amft *et al.* [1] instrumented two participants with four XSens-MT-9B motion sensors placed on the upper and lower arm of each hand, and asked them to perform several activities including eating in a controlled setting. Dong *et al.* [24] also instrumented participants with inertial sensors on the wrist for long periods (between 8.5 and 12 hours) to detect eating events. Sen *et al.* [64] used the accelerometer and gyroscope embedded in an off-the-shelf smartwatch to detect eating events and The smartwatch also has a camera that captures images whenever the classifier detects an eating event. All the images pass through a filtering pipeline to detect if the food was captured by the camera or not. Lin and Hoover [39] also used a smartwatch inertial sensor to monitor the number of bites taken during a meal. Thomaz *et al.* [67, 68] also had participants wear an inertial sensor on their wrist and asked them to engage in several eating and non-eating activities in a controller setting.

The ubiquity of wrist-worn motion sensor underscores the potential of using this method to detect eating. However, many food intake events may not require repetitive hand-to-mouth gestures like drinking from a bottle or eating a sandwich or a fruit while holding it close to the mouth. Besides the missed events, using hand-to-mouth motion can lead to many falsely detected events as it can easily confuse eating with other daily living activities like smoking and lifting weights. These reasons make diet monitoring more challenging with this approach, especially in free-living environments.

2.2.2 Chewing

Several sensing approaches have been employed to detect chewing. GlassSense [19] 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 [26] used a piezoelectric strain sensor placed on the temporalis muscle to detecting chewing bouts. Bedri *et al.* [9, 10] 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.* [18] 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 [58]. Bi *et al.* [14] put EMG gel electrodes and a contact microphone behind participants' ear. Zhang and Amft built custom 3D printed eyeglasses with EMG sensors [75, 76]. The EMG dry electrode is placed on the eyeglass's temples to capture the Temporalis muscle movement. The system achieved a 95% accuracy in for detecting eating episodes 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.

2.2.3 Swallowing

To detect liquids and solids, one of the most promising approaches is to listen to throat sounds using to detect swallowing. Rahman *et al.* [59] have used a piezoelectric microphone on the neck to detect sounds of drinking, eating, and other activities. In [51] Olubanjo and Ghovanloo have also used a throat microphone to detect swallowing and developed an algorithm to classify it from other tracheal events. Yatani and Truong have also used a similar approach to distinguish between a set of 12 activities including different ways of eating and drinking. Commercial products like breastfeeding monitors[1] have also used microphones to detect swallowing.

One of the primary challenges with swallowing sound detection is achieving usable performance in free-living environments as microphones (even surface-coupled) are extremely susceptible to environmental noise and motion artifacts.

2.3 ADM Performance in Free-Living Environments

In the previous section, I briefly discussed some of the challenges faced by different diet monitoring methods in free-living environments. In many studies, researchers cite reasons for why their ADM systems are underperforming in-the-wild. In [67] Thomaz *et al.* recruited 7 participants for a longitudinal unconstrained data collection using motion sensing on the wrist. Before this study, they collected data in-the-lab and used it to train the machine learning models, and then tested with the data collected in-the-wild study. Their results showed F1-scores between 71% and 76% in detecting eating episodes every hour. Thomaz *et al.* explain while it's easy to collect data in a controlled setting and annotate the data, it's very difficult to obtain reliable ground truth for in-the-wild activities. They also mention that activities collected in a controlled setting do not represent how these activities occur in-the-wild. In [10] Bedri *et al.* cited similar reasons for why their chewing-based ADM had low precision when tested in free-living environments. In [40, 51] researchers reported a significant drop in accuracy when testing their audio-based ADMs in-the-wild. This drop was explained by the low signal-to-noise ratio for chewing and swallowing signals as they got overwhelmed with environmental noise.

2.3.1 Challenges with field deployments

In automatic diet monitoring research, most efforts were exploratory. Researchers have investigated a wide range of sensing modalities and assessed how they can be used to detect eating and drinking events. When it comes to assessing the values these systems can provide to the end-user the research work is limited. The main reason behind that is the poor performance of ADM systems in free-living environments. The lack of systems that can reliably work in-the-wild has prevented the research community from exploring the utility and usability of ADM. In their review paper, Schiboni and Amft [63] discuss ADM challenges in free-living environments and attribute the lack of reliable performance to the difficulty of acquiring ground truth in unconstrained environments, lack of validation procedures, social acceptability of the device, and energy efficiency. In recent years, some researchers have worked on addressing these problems and provide solutions, which can mitigate these challenges. For example, Zhang and Amft built custom 3D printed eyeglasses with EMG sensors [75, 76] these glasses were specifically designed to fit the user to ensure the EMG dry electrode has good conductivity. The electrodes are placed on the user's temples to capture mastication from the Temporalis muscle movement. The system captures eating events based on the detected chewing rate. When tested, the system

[1] https://mymomsense.com

achieved 95% accuracy in detecting eating episodes in unconstrained environments. Another example is FitByte [12], on which we are basing our wearable setup. FitByte is an automatic diet monitoring system also based on an eyeglass form factor, but it doesn't require a custom fit frame for every user (one size fits all) and it utilizes a set of inertial sensors to detect chewing and swallowing and a proximity sensor to detect hand-to-mouth gestures. Detecting all these activities helped FitByte achieve 92.7% F1-score in detecting both eating and drinking episodes in free-living environments. We chose FitByte as a reference for our design because it detects all eating actions (chewing, swallowing, and hand-to-mouth) from a single device. It also doesn't require the glasses to be custom fitted for every user, which makes it easier to deploy. FitByte is also not hard to build because all the sensors it requires are commonly found in commercial wearables.

Up to my knowledge, only a few ADM systems were evaluated in long-term field deployments (a week or more). All these deployments have used an off-the-shelf smartwatch to detect eating from hand-to-mouth gestures. Thomaz *et.al.* [67] have conducted a field study with one participant for 31 days. The study was focused on evaluating the performance of the setup in free-living environments using an offline machine learning pipeline. The system achieved 71.3% F-score in detecting eating episodes. Turner-McGrievy *et.al.* [69] have deployed another watch-based ADM for 4 weeks with 12 participants. The goal of the study was to see the influence of ADM on users engaged in a weight loss program. The wearable tries to estimate the calorie count from the number of bytes detected. Participants had to remember to turn on the byte counting App every time they eat and see the estimated KCalorie count at the end of the meal. Participants lost 1.2 Kg on average after the study, but it wasn't clear how much influence the wearable had on the results.

Morshed *et.al.* [49] have also deployed a smartwatch based system with 28 college students for 3 weeks. This setup had a real-time recognition system that prompts participants every time it detects eating and asks them to answer a few questions about their meal. The evaluation was limited to assessing the system's accuracy in detecting main meals (not snacks), but the authors didn't thoroughly investigate the usability of the system nor the impact it had on the user experience with food journaling.

In this research, I present an in-depth analysis of the user experience with ADM and the impact it had on adherence to the food journaling process. My FitNibble system was capable of detecting meals as well as small snacking events in real-time, This feature had a great impact on the overall experience and significantly improved adherence to the food journaling process.

Chapter 3

My Approach

In general, understanding the value of automatic diet monitoring and its impact on users' health is difficult if the devices we use are inaccurate or socially unacceptable [63]. Faulty trackers either underreport events or annoy users with many false positives leading to a trust deficit and negatively impacts adoption rates. For these reasons, I shaped my thesis research around addressing the technical ADM challenges to be well situated to assess ADM's utility and usability. In this chapter I describe the general methodology I followed in addressing the first research question (**RQ1:** How can we develop **practical** ADM systems that **replicate** in-the-lab performance **outside** the lab?). I then show how answering this question situated me in a good position to address the second research question (**RQ2:** Can ADM systems improve food journaling **compliance** and reduce **journaling difficulty**)

Tackling **RQ1** requires understanding the nature of the roadblocks affecting ADM performance in free-living environments. The following sections will discuss these barriers and the approaches I followed to address each one of them.

3.1 ADM Challenges In-the-Wild

Through extensive evaluation of previous work for in-the-wild ADM's assessment, I identified the following reasons for poor free-living performance:

- **Social acceptability:** The form factor of wearable ADMs plays a big role in how people would embrace the technology. If the wearable is intrusive or its appearance doesn't adhere to the social norms, users will find it challenging to always wear it especially when they have company.
- **Environmental noise:** Many ADMs have relied on microphone data to detect chewing and swallowing sounds; these sensors were placed in-ear, on-throat, or in the environment. Researchers have reported a significant drop in accuracy when these devices are used in a noisy environment like restaurants. Further analysis showed that chewing and swallowing sounds are overwhelmed with other environmental sounds making the signal indistinguishable from noise [40, 51].
- **Easily confused actions:** All food intake monitoring techniques identify events by detecting one of the actions that occur during eating or drinking, such as chewing, swallowing,

and hand-to-mouth gestures. Depending on the employed sensing modality and its placement, these actions can be easily confused with other everyday actions. For example, when using an inertial sensor on the wrist to track hand-to-mouth gestures they can confuse eating events with smoking or lifting weights [50, 67]. A proximity sensor tracking jaw motion can confuse chewing with talking or walking instances [9, 10].

- **In-the-wild ground truth:** One of the main challenges to assess the performance of activity recognition systems in-the-wild is the difficulty to obtain reliable ground truth for the user activities. Many free-living ADM evaluations have relied on self-report [10, 29], which suffers from recall errors and user bias. Others have dedicated researchers to follow participants and note down their activities [18], and others have instrumented the user with bulky wearable cameras [67]. The latter two approaches introduce new variables that can affect the participants' behavior and the behavior of others around them, compromising the nature of the free-living environment.
- **Generalizable models:** Due to the challenges in obtaining reliable ground truth in-the-wild, most ADM models are built with annotated datasets recorded in controlled lab environments. Many studies have shown these models do perform poorly when tested in-the-wild [10, 27, 67] demonstrating that eating activities recorded in controlled settings don't represent the true nature of these activities when they occur in unconstrained environments. For example, when eating naturally the user tends to mix this activity with others like talking, using the phone, or watching TV. These natural interactions are difficult to simulate in controlled studies.

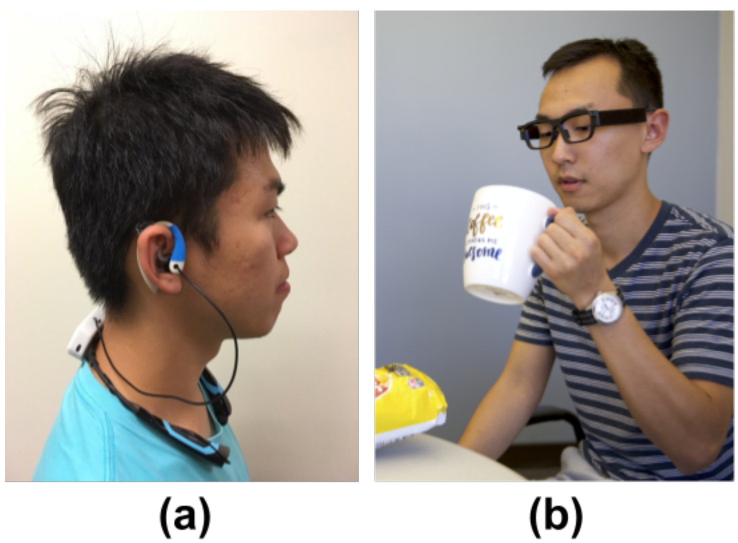

Figure 3.1: ADM systems I developed (a) EarBit: a diet monitoring system that detects chewing using an IMU behind the ear (b) FitByte: an eyeglasses-based ADM that detects chewing, swallowing, and hand-to-mouth gestures and has a camera on board that captures food images

3.2 My Approach to Improving ADM Performance

Understanding the factors that influence the ADM in-the-wild performance has informed the strategies I used to address each one of the above challenges.

- **Social acceptability:** My strategy to tackle social acceptability challenges is to build wearable ADM systems hosted in commonplace form factors like earpieces, headphones, wristbands, or eyeglasses. Given these form factors have well-established social acceptability, it makes it easier to adopt by users especially if they are used to wearing them regularly. While effective this approach comes with a set of design challenges, which include finding the appropriate sensing modalities that can be integrated into these form factors and ensuring the sensor(s) placement and fitting deliver repeatable clear food intake signals.
- **Environmental noise:** In my approach, I focused on employing sensors that are less prone to environmental noise and explored placements and fitting techniques to ensure clarity of the signal. The following chapters (4,5) describe in detail the design of three ADM systems and the sensing techniques they used to track eating and drinking instances.
- **Easily confused actions:** To reduce confusion with other activities, I've employed several techniques like introducing new sensing techniques to reliably detect eating actions, adding reference sensors to help reduce false positives by measuring the same signal from different locations, and utilized feature engineering techniques that specifically target confusing activities.
- **In-the-wild ground truth:** To obtain reliable ground truth I developed an eyeglasses ADM with an integrated miniaturized camera that points down towards the user's mouth. Due to the nature of the activities that we would like to track (i.e. eating and drinking) this placement allowed us to monitor the user behavior without invading their privacy or the privacy of who is around them. The microphone was removed from the camera module to prevent it from recording audio, and the user had the freedom to switch the camera off whenever they wanted. The camera's placement and size made it easy for users to forget its presence during the day and act normally. The footage recorded by the camera was also used for other purposes like identifying food types.
- **Generalizable models:** Finally to address the models' generalizability, I designed several methods to record and label activities in semi-controlled environments which proved to perform well in free-living environments. An example of this method is the recording of activities during a dinner party in a two-story house instrumented with cameras. In this scenario, the only instructions given for the users was not to leave the house during the party (75 minutes). Other than that the participants had no instruction on what to do and were acting normally (Figure 3.2.a). This method allowed us to obtain a dataset annotated with 1-second precision. Another method I developed to obtain annotated datasets was done with eyeglasses cameras. This method focused on recording short events where the user is engaged in eating, drinking, or other activities (e.g. grabbing a snack, watching TV, exercising, lunch meeting). Targeting these events allowed us to obtain short and easy to annotated datasets of condensed natural activities (Figure 3.2.b).

Figure 3.2: Semi-controlled approaches followed to build reliable in-the-wild food-intake models (a) Collect data in a house environment (b) Collect data for short events in-the-wild

3.3 My Approach to Evaluating ADM Utility and Usability

My work on improving ADM performance in real-world settings through Earbit and FitByte has paved the way for me to address **RQ2**. Informed by the design of my previous ADM platforms, I developed FitNibble an ADM that can be attached to the user's eyeglasses. The device is a part of an end-to-end system that includes an iOS app and a backend server. This system allows users to receive just-in-time notifications prompt them to do the logs as soon it detects they are eating (Figure 3.3). This feature was enabled by the proximity sensor that help detects hand-to-mouth gestures. The proximity sensor helped reduce the average detection delay to 7 seconds.

I deployed this setup in a long-term field study. The study required participants to try traditional self-report food journaling for a week. We then introduced them to FitNibble and asked them to use it for another week. Throughout the study, we tracked the user experience with daily surveys (experience sampling) and interviewed the participants at the end of each phase. The interviews focused on assessing the adherence to the food journaling process, the social acceptability and privacy concerns, and general feedback on the usability and utility of the two journaling methods and how they compare to each other.

Our data analysis showed that participants depended on the wearable to do their logs and journaling difficulty dropped significantly in the second phase. But The major outcome of this study was in *compliance*, as we saw a 19% ($p = 0.13$) increase in the number of days with no-missed events after using FitNibble. All these results highlight the potential of ADM in reducing the missed events and improving compliance to food journaling (**RQ2**).

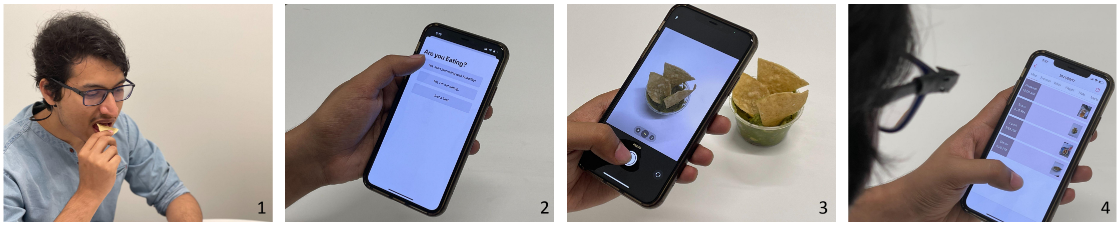

Figure 3.3: With FitNibble users can receive just-in-time notifications prompting them to log their meals and snack. This setup helped in addressing the food journaling challenges with compliance and missed events

The following chapters document the three ADM systems I developed: EarBit, FitByte and FitNibble. These chapters also discuss the different approaches I took to address ADM challenges in-the-wild and how they all formulated the answer to my research questions.

Chapter 4

EarBit

4.1 Introduction

EarBit is an experimental, head-mounted wearable system that monitors a user's eating activities while remaining resilient to the unpredictability noise in real-world settings (Figure 4.1). This system is based on a wearable developed to monitor jaw and tong motion for silent speech recognition [7, 8, 61]. The setup design is also driven from my previous work in diet monitoring using in-ear proximity sensors to detect chewing instances [9, 10].

EarBit uses chewing behavior as a proxy for eating, resulting in instrumentation of the head. As an experimental platform, EarBit's design allows for the collection of data from a number of sensing modalities (optical, inertial, and acoustic). We use these sensors to determine the combination of sensing modalities that is most effective for detecting the moment of eating. To reduce the gap between results from a controlled laboratory setting and the real world, the algorithms for these sensors (shown in Figure 4.1) were developed and evaluated in a semi-controlled home environment that acts as a living lab space. The results of this study indicated that an inertial sensor behind the ear (measuring jaw motion) in tandem with an inertial sensor behind the neck (monitoring body movement) produced good results in detecting eating activity, and was also the form factor considered most comfortable by the participants; particularly since the function of the inertial sensor behind the neck is used to detect activities like walking and could be replaced by a user's smartphone or wrist-mounted activity tracker.

Eating detection models trained on data from the semi-controlled study were then tested on a new dataset collected in a relatively relaxed "outside the lab" environment. We recruited a new set of 10 participants, and instead of asking them to come to our study location, we gave them the EarBit prototype and asked them to use in their own environments. We collected data for a total of 45 hours. EarBit's IMU is essentially a chewing sensor, and at a 1-second resolution, EarBit correctly recognized chewing activity with an accuracy of 93% and an F_1 score of 80.1%. When these Outside-the-Lab chewing inferences are aggregated into separate eating episodes, EarBit accurately recognized *all but one* recorded eating episodes (delay = 1 minute). These events ranged from 2 minutes snacks to 30 minutes meals.

The main contribution of this research is a demonstration of the experimental EarBit system that successfully recognizes eating episodes in a real world setting. This contribution comes in

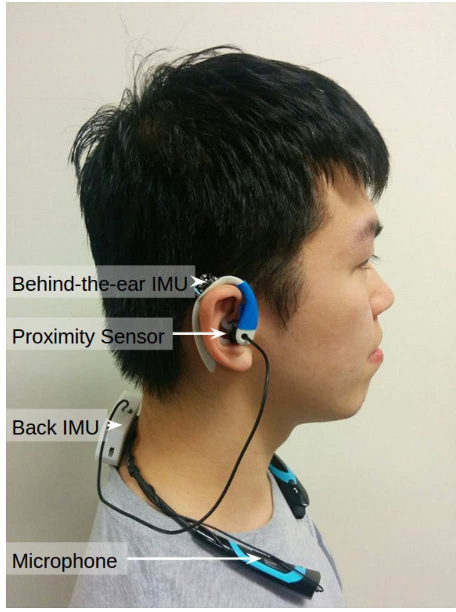

Figure 4.1: EarBit's data collection prototype with multiple sensors. Our semi-controlled and Outside-the-Lab evaluations show that the Behind-the-Ear IMU is enough to achieve usable performance. We envision such a sensor to be part of future eyeglasses or augmented reality head-mounted displays.

three parts:

1. An evaluation of a wearable setup for eating detection based on off-the-shelve form factors.
2. A novel, semi-controlled laboratory protocol used to judge the effectiveness of combinations of three sensing modalities for eating detection.
3. A machine learning model that uses inertial data collected in the semi-controlled environment to reliably recognize eating episodes in a real world setting.

4.2 Data Collection

McGrath identified three key factors when conducting a study: precision, generalizability, and realism [44]. However, it is difficult to collect data that has all three elements. At one extreme, laboratory experiments allow researchers to accurately measure behavior because the researcher can control when and where behaviors of interest occur [16], but this data often lacks realism. At the other extreme, *in situ* observations allow researchers to capture real life behavior. However, this data often lacks precision due to the lack of proper instrumentation or control, resulting in poor ground truth data. Consequently, the leap from a controlled study to the *in-situ* study often becomes intractable for machine learning models.

4.2.1 Semi-controlled Lab Study

In order to bridge the gap between controlled and real life studies, we collected our training data in a simulated natural environment. We observed participants interacting in a sensor-instrumented home (the Aware Home at the Georgia Institute of Technology) especially designed to support ubiquitous computing research [37]. This 3-storied building spans over 5,000 sq. ft., and is embedded with various sensors to support data collection.

Scenario

The participants were invited to the Aware Home for dinner. Once at the house, a researcher facilitated the group's activities over a 75-minute session. There were 3 to 4 different participants in each session. In an attempt to catalyze conversation, participants were chosen such that each participant was familiar with at least one other participant at the dinner. In total, sixteen participants (19-25 years, 9 female & 7 male) participated in a total of 5 sessions.

After completing a brief demographic survey, the participants were asked to wear the multi-sensor setup shown in Figure 4.1. Once the participants felt comfortable with the hardware, they either ate dinner, took a tour of the home, or engaged in free-flowing conversation while watching TV. Although the group had the freedom to chose the order of activities, all participants performed all activities in each session. Also, there was no restriction on the duration of each activity.

The tour required walking through the home, including walking up and down a flight of stairs. The group decided whether to eat their dinner either in the dining area or the living room. The participants chose their dinner entree from local restaurants with different cuisines. While watching TV, participants were also offered snacks, such as potato chips, chocolate candy, peanuts, apples, and bananas. Additionally, participants were provided bottled water and assorted sodas to drink. Since participants already knew each other, they were comfortable with spontaneous, free-flowing, natural conversations that rarely required any host facilitation. Additionally, familiarity allowed the participants to eat in a natural manner without being self-conscious about their manners. For example, participants often talked and ate simultaneously.

Of the sixteen participants involved in the study, only ten participants provided usable data. Four participants had to leave prematurely due to a personal emergency, and two participants had corrupted or missing sensor data. Nevertheless, our semi-controlled dataset had 12.5 hours of annotated data with almost 26% labeled as *chewing*.

Ground-truth

We used four video cameras to record participants' activity. Three stationary video cameras were located in the dining area and living room, and a handheld camera was handled by a researcher. This camera followed participants when they went to areas outside the range of the stationary cameras (e.g., the stairs, kitchen). In order to sync the devices' data with the cameras, each participant was asked to perform a gesture of tilting their heads from side to side. To sync the video cameras, we switched the house lights on and off three times at the beginning of each session.

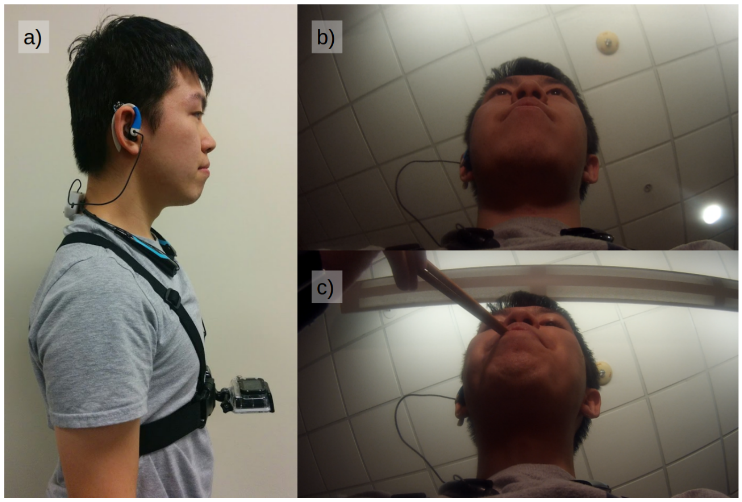

Figure 4.2: Outside-the-lab study configuration: a) A user wearing the EarBit system and GoPro camera. b) A picture from the GoPro camera of the user working at a desk. c) A picture from the GoPro camera of the user eating with a pair of chopsticks.

Over the course of the scenario, user behaviors included walking, standing, sitting, talking, eating, laughing, watching TV, etc. At the conclusion of the scenario, participants completed a post-study survey. The survey covered: (1) comfort ratings for different hardware components of the experimental device; (2) comfort ratings for different combinations of components; and (3) an open question about their experience with the experimental device. Following the post-study survey, we engaged the participants in an informal focus group and discussed usability, comfort, and practicality.

4.2.2 Outside-the-lab Study

The semi-controlled Aware Home study put the participants in a social group and aimed to collect the data in a realistic setting. While we largely succeeded in collecting realistic behavior, the participants were still aware of multiple cameras and the data recording focused on capturing eating events. For example, it would be uncommon for a user to spend 26% of their day eating. While a high percentage of eating episodes are an optimal approach to collect training data, it is not an ideal evaluation scenario. Hence, we decided to evaluate our algorithms in a slightly more relaxed and naturalistic environment. We outfitted 10 new participants (3 female and 7 male, aged 18 to 51) with EarBit and asked them to take it out of the lab and use in their natural environments. In this study, participants recorded data in diverse environments including houses, offices, cars, restaurants, prototyping workshops, streets and public transport. None of these participants were part of the previous study and the participants were advised to engage in at least one eating activity. We recorded up to two 3 hours sessions with each participant. The session length was limited by our groundtruth collection device: GoPro Hero 3.

Considering participants were going to use EarBit outside a controlled environment, groundtruth collection becomes hard. Traditionally, self-reporting any eating activity is a standard practice for determining ground truth for eating studies in unconstrained environments. However, a number of previous studies (e.g., [6]) and our own pilot study showed that self-reporting is not reliable. In an initial version of our study, several participants indicated that they forgot to write down eating times while they were eating. Instead, they wrote best guesses of time and duration. In other instances, participants did not remember to write down eating times until after the study was over. Hence, we revised the study to obtain ground truth via a chest-mounted GoPro Hero 3 camera. The camera faced upward towards the participant's face and continuously recorded activities around the participant's head (Figure 4.2a). Apart from asking participants to try and not occlude their mouth while eating (Figure 4.2c), there was no change to the instructions given to participants. They were told to conduct their normal, daily activities, and to self-report eating via manual logging. The GoPro sessions lasted for 3 hours, due to the battery constraints of the camera. In order to collect sufficient per-person data, participants were asked to complete two sessions. However, 5 of the participants were unavailable for a second session, and we had a total of 15 outside-the-lab sessions (3 hours each). 11% of the recorded data was identified as *chewing* and is representative of an average user's daily life [13]

4.2.3 Video Annotation

To acquire ground truth for each user's activities in both studies, we hand-annotated the video recordings from both studies. We used Chronoviz [28] to synchronize video and sensor data. Four coders annotated the data by manually inspecting the recorded audio and video. The annotations included six labels divided into two categories: body movement (*moving* or *stationary*) and jaw activity (*chewing*, *drinking*, *talking*, or *other*). Any labeling window can have one annotation from each of the two categories, but not two from the same. For example, a user could be walking (body movement) and eating (jaw activity), but cannot chew and talk (both jaw activities) simultaneously.

Moving included a wide variety of actions, like walking, body rocking, etc. *Stationary*, *chewing*, *talking*, and *drinking* are self explanatory actions. We used the *other* label for relatively infrequent but significant jaw actions such as laughing, coughing, and yawning. We did not annotate portions of the video when the participant could not be seen; though that was rare. We performed the annotation by considering non-overlapping 1 second window of video and labeling it as the activity that lasted the longest within the window. High granularity annotations allow us to learn from small, quick transitions. For example, Figure 4.3 shows a user having a meal over a 10 minute period. The user transitions through a number of activities while having his or her meal, and we are able to annotate small and sporadic periods of silence in addition to the main activity of chewing.

Additionally, a section of video can have more than one label, one from each of the two categories. For example, a person that is walking while eating simultaneously can have both of these labels for the same segment of data. A similar example is depicted at the end of the expanded subplot in Figure 4.3, in which the subject was labeled to be both *moving* (from the body movement category) and *talking* (from the jaw activity category). However, when selecting frames for training and testing we resolve the confusion in mixed activities labels by giving different priority level for each class. The *moving* label has the highest priority, followed by *other*, *drinking*, *chewing*, *talking*, and finally *stationary*, respectively. The activity with the highest priority becomes the dominant label for the frame. Mixed activities that had eating overlapped with labels

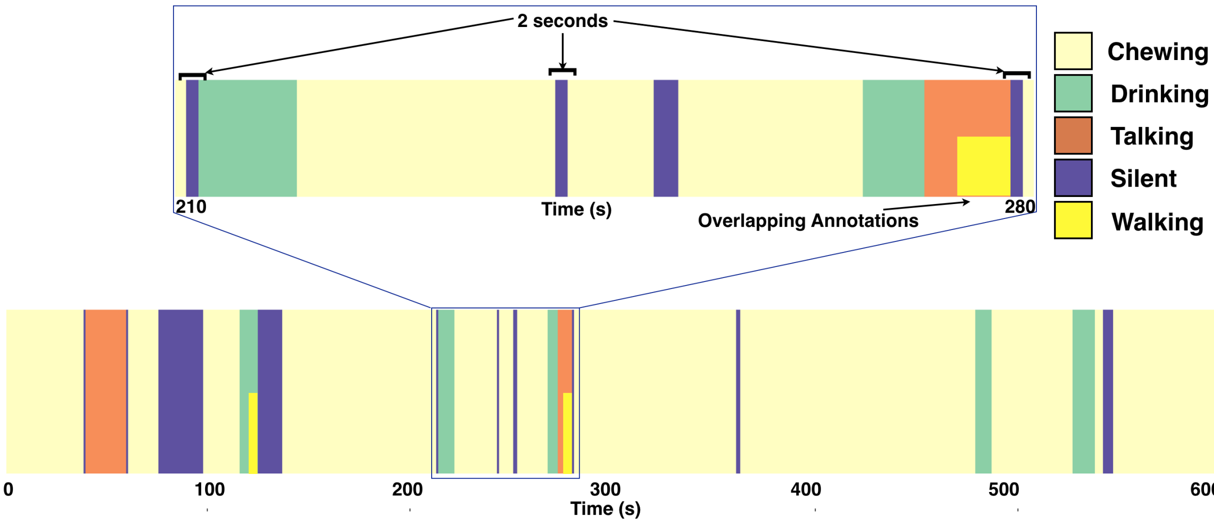

Figure 4.3: An example of annotations for eating activity. We annotated our video data at a 1-second resolution. In this 600-second example of a user having a meal, we capture all minute transitions and capture various 2-second intervals where the user stopped chewing. Mixed activities would have overlapping annotations as indicated in the example of walking and talking. For all the instances when the user is not moving a stationary label is also added.

with higher priority represented only 2.2% in the semi-controlled lab (Aware Home) data set.

To annotate the video for the outside-the-lab study, we chose to provide only two labels (*chewing* and *not-chewing*). Therefore, it is important that the multi-class machine learning models trained using the semi-controlled study be ultimately converted into binary classification models for the outside-the-lab study. We will discuss this in detail in Section 4.3.2.

For annotating the recorded videos, we employed 4 coders and then used Cohen's Kappa to compute inter-rater reliability [38]. Kappa (K) was computed using a 15-minute video sample from the Aware Home dataset. This video was chosen so as to encompass a wide range of activities. Because any subset of activities could take place simultaneously or individually, the annotations are not conditionally-independent. Hence, we computed the inter-rater reliability for each activity separately, where $0.60 < K \leq 0.80$ represents satisfactory agreement and $K > 0.80$ represents near-perfect agreement. Our worst inter-rater reliability was $K = 0.69$ (for *stationary*) and our best was $K = 0.99$ (for *other*); average agreement across all activity labels was 0.84.

4.3 System Description

In this section, we first describe the initial set of sensors identified to be suitable for detecting chewing/eating through instrumentation on/near the head. We then discuss the process of choosing an optimal subset of sensors leading to a revised design of EarBit and its machine learning algorithms.

4.3.1 Choosing the Right Sensor(s)

Our goal is to design a system that accurately detects the chewing activity as a proxy for food intake. We aim to achieve this using an optimal number of sensors, while considering the social acceptability and comfort of the form factor. To this end, we investigated a number of sensors and compared their performance and usability.

Sensor Selection

Previous research in food intake monitoring has focused on tracking various actions that occur during an eating activity, so-called proxies for eating. These include hand-to-mouth gestures, chewing, and swallowing. Although hand motion is involved in most eating activities, and has the advantage of leveraging common commercial sensing platforms that people already have (e.g., smartwatches or activity trackers), it has limitations (i.e., usually only one hand is instrumented) and we felt that detecting chewing and swallowing is more directly associated with eating and, therefore, sufficient to infer eating episodes. Fontana et al. support our claim, indicating that in a naturalistic environment, jaw motion can be more indicative of eating activities than hand gestures [27].

To detect chewing, we exploit two sensing modalities: optical and inertial. The optical sensor is the VCNL4020 fully-integrated **proximity sensor** with an infrared emitter. Bedri et al. have used this sensor to track jaw motion for detecting chewing in a controlled environment [9, 10]. The sensor is placed at the entrance of the ear canal and measures the degree of deformation at the canal caused by the movement of the mandibular bone. The sensor is fixed inside a Bose IE2 ear-bud, and it features a wing to tuck under the outer-ear flap. The system does not require any calibration for different users and we evaluated its adaptability to different users in the prototype testing phase.

Apart from the in-ear proximity sensor, we augmented the outer-ear flap of the ear-bud with a **9 Degree-of-Freedom IMU** (LSM9DS0). Rahman et al. used a similar sensor to detect eating events in a controlled setting [58]. The flap helps in coupling the IMU to the temporalis muscle. This is one of the four mastication muscles and links the lower jaw to the side of the skull covering a wide area around the ear. During chewing, the muscle continuously contracts and relaxes, and this movement can be picked up by the IMU. Figure 5.4 shows an example of sensor stream of the behind-the-ear IMU while the user was talking, eating, and then walking.

The system also includes a **microphone** around the neck; a HBS-760 Rymemo Bluetooth headset (Figure 4.1). A similar microphone-based approach has been used to detect swallowing [74]. These works recommended placing a microphone coupled to the throat with some level of acoustic shielding. With the aim to increase comfort, we modified the type and placement of the sensor to be slightly more socially-acceptable. It leads to slightly degraded signal-to-noise ratio, but we accept it as a reasonable compromise. In addition to these sensors, we also placed a 9-DOF IMU behind the user's neck (Back IMU in Figure 4.1). This IMU is used to measure large body motions, such as walking. In the future, such information could alternatively be extracted from a wrist-worn fitness device or a smartphone.

Data from the two IMUs and proximity sensor is sampled at 50 Hz using a Teensy 3.2 microcontroller, which stores the received data on an SD card. The microcontroller, back IMU, and battery are housed in a casing and attached to the back of the Bluetooth headset, as shown in Figure 4.1. Audio from the wireless Bluetooth microphone is recorded at 22.05 KHz and sent to an Android phone. We developed four copies of the prototype for instrumenting multiple users simultaneously for the semi-controlled lab study.

Sensors Comparison

Using only the dataset from the Aware Home semi-controlled study, we compared different sensing modalities on the basis of their recognition performance and usability.

The activity recognition processing pipeline was based on prior literature and compared the performance of different sensors and all combinations of sensors using leave-one-user-out (LOUO) user-independent testing. We used the approach suggested by Bedri et al. to develop the processing pipeline for the IMU and proximity sensor ([10], see Figure 4.5). Bedri et al. also recommended using Hidden Markov Models (HMMs) with 10 minute segments for the final classification. For the neck microphone, past work suggested using Mel-Frequency Cepstral Coefficients (MFCCs) to differentiate between speech and non-speech activities [43, 52]. Such a capability can be valuable to differentiate between talking and other activities. Therefore, we calculated 26 MFCCs from the microphone data (100 ms using 20-filter bank channels) before calculating further features from the audio.

Figure 4.6 shows a preliminary comparison across the sensing modalities. The IMU placed behind the neck (back-IMU) was used in all sensor conditions because it helped to filter out movement based on more gross body activities (e.g., walking). The behind-the-ear IMU (E) performs better than other combinations. The combination of behind-the-ear gyroscope and proximity sensor (E+P) has comparable results to E, but there are no clear benefits of using the additional sensor. Beyond this preliminary performance evaluation, we decided to focus our

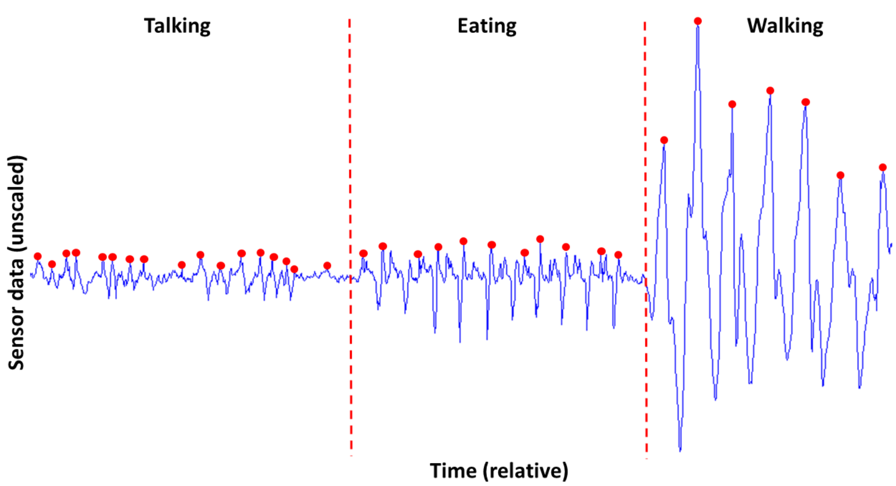

Figure 4.4: Example data from the y-axis of the behind-the-ear gyroscope. The dots indicate local maxima with high energy in the signal. As compared to talking, the peaks for eating are more periodic and ”spiky”.

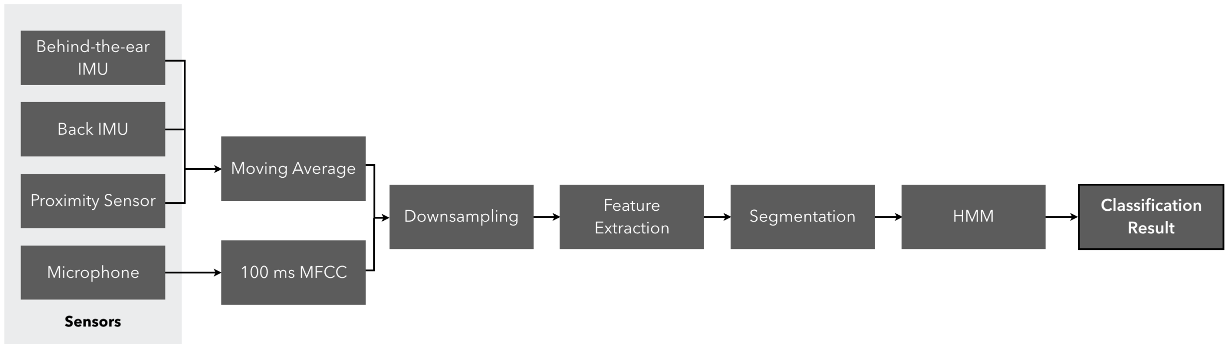

Figure 4.5: Flowchart for initial evaluation of the multi-sensor setup

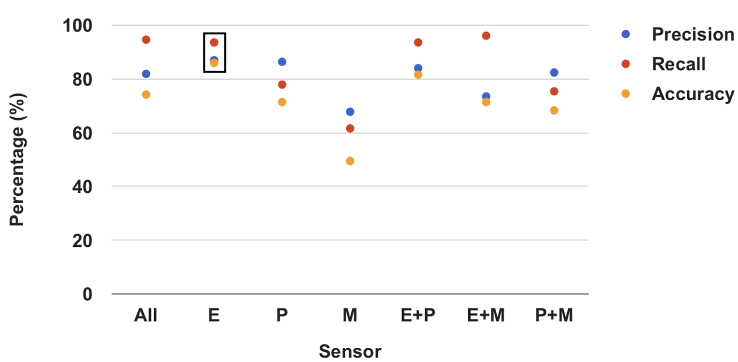

Figure 4.6: Comparison between sensing modalities. E = behind-the-ear IMU, P = outer-ear proximity sensor, M = neck microphone. The back IMU is used in all condition to detect if the user was walking. The performance of behind-the-ear IMU (E) was most consistent for all three metrics. It was also considered most comfortable to wear by the users.

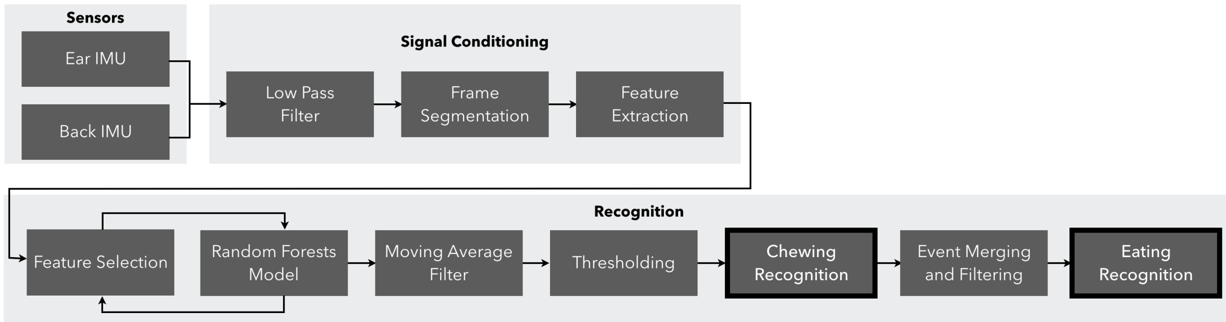

Figure 4.7: Flowchart for EarBit algorithm

attention only on the behind-the-ear gyroscope. While it had marginally better performance than other sensors, more importantly it was the most preferred sensor by the users.

Our post-session survey highlighted that the participants did not prefer using the in-ear proximity sensor. Respondents rated comfort and usability on a five point Likert scale. Wilcoxon Signed Rank Test showed that the users found back-of-the-ear IMU more comfortable than the in-ear proximity sensor (Md = 4 *vs.* Md = 3.5, p ¡ 0.05). In the informal focus group session as well, multiple users complained about the in-ear earbud.

”The [in] ear piece was uncomfortable. It felt piercing and itchy.”
”The Bose headphones felt uncomfortable after extended periods of use.”
”I’m not used to having something in my ear when I’m eating ”

Thus, we decided to limit the evaluation of the Outside-the-lab study to the behind-the-ear IMU and used the back-IMU to cancel large body motions.

4.3.2 Redesign of recognition pipeline

The processing pipeline described in Section 4.3.1 was based on prior literature and we used it to do a preliminary comparison of performance of various sensing modalities. Instead of opting to continue to optimize our Hidden Markov Models, we decided to switch to a different machine learning approach. In general, HMMs are more suited for discovering patterns and transitions in temporal data sequences. They are ideal when the model needs to develop an understanding of the *shape of the signal*. However, Figure 5.4 shows that the behind-the-ear IMU acts as a very direct sensor that captures the oscillation patterns of the temporalis muscle when a user is chewing. The behind-the-ear IMU simplifies the machine learning problem to primarily differentiate between magnitude and periodicity of motion from different activities. For this problem, we believe summary statistical features and an algorithm like Random Forests should suffice.

In the rest of this section, we provide full details of our machine learning pipeline, and provide explanations for various design decisions. Figure 4.7 shows the whole processing pipeline.

Signal Conditioning

The new pipeline starts with a **preprocessing** step to condition the raw signals. This step includes smoothing the 50 Hz gyroscope data using a Butterworth filter of order 5 (cut-off frequency = 20 Hz). Data is then segmented using 30 second windows sliding at 1 second.

Feature Extraction

Our feature set aims to encode the relevant information about the motion of the temporalis muscle when the user's jaw moves. For each 30 second window, we compute 78 features to characterize jaw movement while chewing. These features are essentially 13 features computed for each axes of the gyroscopes placed on the ear and back (i.e., 13 features $\times$ 3 axes $\times$ 2 sensors = **78 total features**).

When a user chews, the jaw moves, and the back-of-the-ear IMU picks up the motion. In an ideal case, energy or magnitude alone will be very high for such motions and low when the user is doing some other activity. However, a user performs many activities that can generate significant motion that gets recorded on the behind-the-ear gyroscope; walking and talking are common examples. Figure 5.4 shows example data from the *y*-axis of the gyroscope when the user was talking, then transitioned to eating, and then walking. One valuable insight captured by Figure 5.4 is that **chewing motion is more periodic than many other activities, such as talking.** On the other hand, walking and some other large motions (e.g., exercises) are also periodic. Though in some cases the overall magnitude of motion while walking is significantly larger, it won't always be true. For such cases, a separate IMU on the body (in our experiments behind-the-neck IMU, but in practice a wrist-worn or pocket-held device) can be used to detect these large motions, as shown in other research related to activity recognition [4, 35, 36]. Next, we list our 13 features that capture information about the magnitude and periodicity of motion for different axes and sensor locations. These features include time and frequency domain features that are commonly used in recognizing human activities from inertial data. Size of the FFT is same as the size of the feature calculation window (i.e, 30 seconds = 1500 samples). In [48]. Morris et. al. introduced a set of 5 features based on signal auto-correlation to reliably recognize repetitive strength-training exercises using inertial sensor. In general, the auto-correlation of any periodic signal with frequency f will produce another periodic signal with peaks at lag 1/f, while a signal that has no periodic component will produce no peaks when it's auto-correlated. Just like strength-training exercises, chewing produces repetitive motion that can be captured using same features. Hence, our features set also includes auto-correlation features, and were computed using the same methods as applied in [48].

1. **Magnitude of motion.**
   (a) **Root Mean Square** encodes the amount of energy in the signal.
   (b) **Variance** is square of RMS and encodes similar information. Having both RMS and variance can provide flexibility if there is non-linearity in some axes.
   (c) **Entropy** reflects the amount of information (or conversely noise) in the signal. Entropy tends to be a strong feature in detecting silent and noisy activities, such as silence and speech. The normal formula for Shannon's entropy was used to compute

the entropy feature, but the bins are predefined in increments of 10, ranging from -50 to 50. The outliers were assigned to a separate bin.

(d) **Peak Power** is the magnitude of the dominant frequency of the signal. If a signal is fairly repetitive (e.g., eating and walking in Figure 5.4), the magnitude of the main frequency can indicate the intensity of motion, and can help in differentiating between facial and whole-body motions.

(e) **Power Spectral Density** is magnitude of power spectrum in logarithmic scale.

2. **Periodicity of motion.**

(a) **Zero Crossing** captures the rough estimate of the frequency of the signal.

(b) **Variance of Zero Crossing**. Zero crossing is going to be high for any high-frequency data, and can be severely affected by noise. We calculate the variance in the times at which signal crosses zero, to record the periodicity of zero crossings.

(c) **Peak Frequency** is the dominant frequency of the signal, calculated through a frequency transformation.

(d) **Number of Auto-correlation Peaks**. Abnormally high or low number of peaks here indicate noisy signal.

(e) **Prominent Peaks** are the number of peaks that are larger than their neighboring peaks by a threshold (0.25). Higher number of prominent peaks suggest a repetitive signal.

(f) **Weak Peaks** are the number of peaks that are smaller than their neighboring peaks by the same threshold (0.25) as Prominent Peaks.

(g) **Maximum Auto-correlation Value** is the value of the highest auto-correlation peak. A higher value suggests very repetitive motion.

(h) **First Peak** is the height of the first auto-correlation peak after a zero crossing.

Feature Selection

Given the large number of computed features, we introduced a feature selection step in our pipeline. This step helps in avoiding the curse of dimensionality and enhances the generalizability of our eating detection models by reducing overfitting.

We implemented the feature selection process using the sequential forward floating selection algorithm (SFFS), which is proven to be very effective in searching for optimal feature set [57]. For feature evaluation, we used random forest classifiers to build models using out semi-controlled lab dataset. A leave-one-user-out cross validation was performed at each step, and the exclusion and inclusion criteria for features was based on the F1 score of chewing detection.

The SFFS algorithm selected 34 out of 78 features as most effective for eating detection. These 34 features came from all 13 feature types across different axes. The most common selected feature types are entropy, peak frequency, the number of auto-correlation peaks, and first peak after a zero crossing.

Recognition

We use Random Forests (implemented with the *Scikit-learn* toolkit in Python) and leave-one-user-out validation to avoid overfitting. Furthermore, we keep all Random Forest-specific parameters at their default values to avoid any manual overfitting. This is where Random Forests are especially useful because they do not need much manual tuning and the only major parameter is the size of the trees. However, with separate feature selection phase, we do not need to control the size of the trees as well in most cases. Therefore, we only optimize some of our windowing parameters and we will discuss those in detail later in this section.

Detecting Chewing

The labels in the Aware Home dataset included: chewing, walking, talking, stationary, drinking, and other. Due to the very low number of occurrences in the dataset, the latter two labels, which represented 5.3% and 1.2% of the dataset respectively, were removed from training and classification tasks. Completely removing these instances from the dataset would skew the timeline. Therefore, the algorithm simply skips these instances during training and classification tasks, but still uses the sensor information to calculate features for other instances (remember that the features are calculated over 30 second windows). In our dataset with 26% of data points labelled as chewing. This happens because our training data was collected in a social setting when the group of participants were socializing and a significant amount of time was spent eating. While this is not representative of an average day in a user's life, it provides us with some robust training data.

In contrast to the Aware Home dataset, the Outside-the-lab dataset only had two labels: *chewing* and *not-chewing*. However our machine learning models made a four-class classification: chewing, walking, talking, and stationary. Instead of changing the classifier's output classes to match the labels used in the Outside-the-lab dataset, we simply treat all non-chewing predictions as "not-chewing". Therefore, when we report results in Section 4.4, we convert our performance metrics to reflect the performance of a binary classifier. In the interest of uniformity, we do this conversion to binary classification for both the semi-controlled lab (Aware Home) data set and the Outside-the-lab datasets.

The machine learning model produces recognition results every 1 second (recall that we used 30 second windows sliding by 1 second). Since, there is seldom any need for 1 second resolution for chewing inference, we apply a moving average on the confidence value returned by the Random Forests. Consecutive values were averaged together to produce the new confidence value for each second. The moving average window was centered on the value to be predicted. The size of moving average window (optimal value = 35 samples) is tuned using the Aware Home dataset.

The output of the filter is converted into a binary decision by using a simple threshold of 0.5. An example of this post-processing is shown in Figure 4.8. The result of this tuning procedure will be discussion in Section 4.4.

Detecting Eating: Aggregating Chewing Inferences

Although EarBit acts as a chewing sensor, most users will be interested in identifying eating events. We aggregate individual chewing inferences into eating event inferences through a two-step process (shown as the last step in Figure 4.8): *merging of events* and *filtering short events*.

Merging of events helps in removing sporadic discontinuities in eating recognition. This is based on an assumption that a user won't have two meals within 10 minutes of each other.

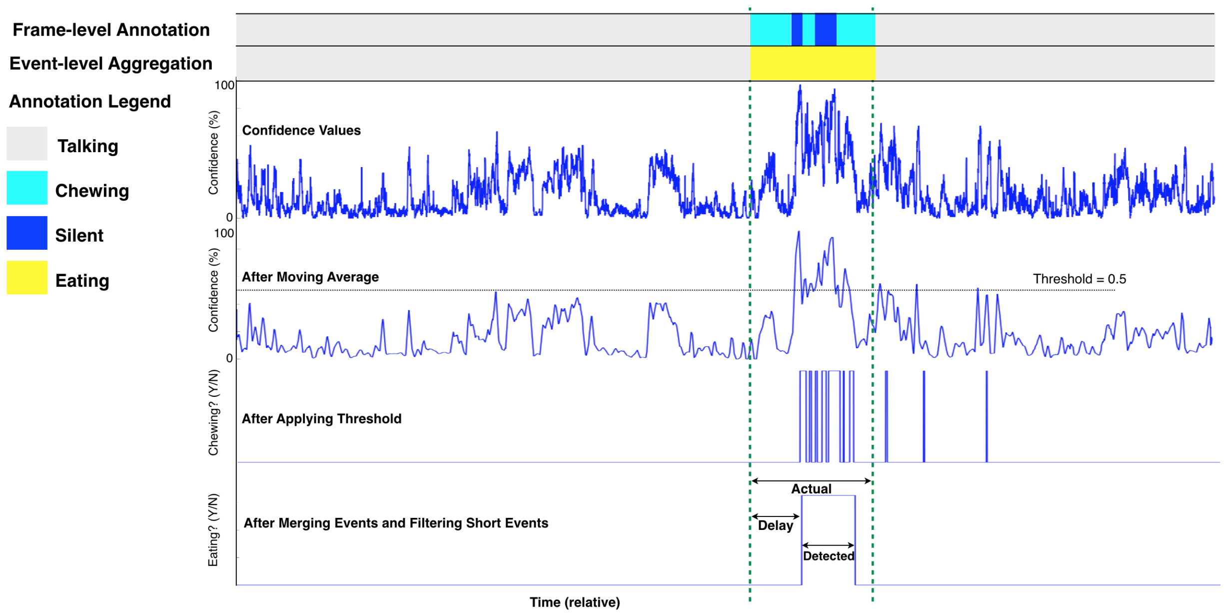

Figure 4.8: An example of conversion of confidence values from Random Forests to frame-level results (chewing) and then to event-level predictions (eating episodes).

Therefore, we merge all labeled and recognized eating events that occur within 10 minutes of each other. Here, we understand that time cannot be the only factor in segmenting meals. For example, a user might start eating an apple, leave for an urgent meeting, and then come back to continue eating the fruit. Perhaps a richer understanding of the user's activities and intent would be necessary, but that is not the focus of this evaluation.

In addition to the merging step, we added a second layer of filtering to remove small isolated events that are less than 2 minutes in duration. This filtering step comes at the cost of skipping very short snacks, which is a compromise we made to improve precision in detecting full meals and snacks that are longer than 2 minutes.

Overall, we minimize the number of tunable parameters in our approach; Random Forests also implicitly minimize the need of tuning parameters (as discussed earlier). Therefore, the only tuning parameter for EarBit is the size of the moving average filter. All other parameters were based on domain knowledge and assumptions about the user's behavior. For example, for merging events, we assume that a user won't have two separate meals within 10 minutes of each other. This assumption was also confirmed when we analyzed the video recordings. None of the tunable and human-set parameters were optimized using the outside-the-lab dataset. That dataset was collected to evaluate EarBit's performance and we made sure that none of EarBit's parameters were optimized on it.

4.4 Results

In this section we will discuss EarBit's performance in detecting eating in our two studies. We started by developing and validating our algorithm on the Semi-Controlled Lab dataset and then we used those models to evaluate performance of the Outside-the-lab dataset. We completely sequestered the data from the Outside-the-lab dataset and analyzed it only after the algorithm was ”frozen”, that is, after satisfactory validation on the Aware Home dataset. This was done to avoid any unintentional and manual overfitting on the test data.

For evaluation, we test the algorithm's performance on both frame-level (chewing detection) and event-level (eating episode detection). The main performance measures are F_1 score, precision, recall and accuracy. For the event level analysis, we also reported $delay$, which measures the time from the beginning of an eating event till EarBit starts recognizing it. Additionally, we also measure *coverage*, i.e., what percentage of actual event was recognized. For example, if a user spends 15 minutes having dinner, but EarBit predicts a 12 minute eating event, then Coverage is 80%. In cases where the predicted event starts before or ends after the actual event, $Coverage$ can give artificially high results. However, we did not have any case where the predicted event exceeded the time-bounds of the actual eating episode.

The main difference between *coverage* and *recall* as metrics in our evaluation is, recall is computed directly on prediction values produced by Random Forest. While coverage is computed after applying the filtering steps on the prediction results as shown in figure 4.8

4.4.1 Validation on Semi-controlled lab dataset

To validate the performance of EarBit's algorithm, we used leave-one-user-out cross validation. We used these validations to tune our only tunable parameter: size of the moving average window.

Figure 4.9 shows chewing recognition results for semi-controlled lab study as a function of the moving average window size. The results stabilize at 35 seconds mark. EarBit's cross-validation accuracy is 90.1%, F1 score is 90.9%, precision is 86.2%, and recall is 96.1%.

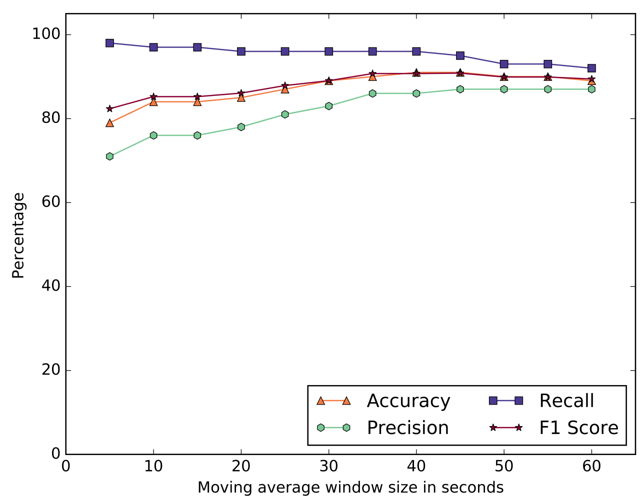

Figure 4.9: Chewing recognition results for semi-controlled lab

For the event-level performance, with a 35 seconds moving average window, EarBit captured all 15 eating events in the dataset, and falsely recognized one non-eating episode as eating. It achieved 89.6% coverage and the average delay in event recognition is 21.3 second. Once the moving average size and the machine learning models were final, we evaluated its performance on the Outside-the-lab dataset.

4.4.2 Outside-the-lab Study

For the Outside-the-lab data, with a 35 seconds moving average window, EarBit detects chewing with an accuracy of 93% (F1 score = 80.1%, Precision = 81.2%, Recall = 79%). When converted into eating episodes, EarBit successfully recognized 15 out of 16 eating episodes, and it only falsely recognized 2 additional eating episodes. The average delay is 65.4 seconds and the mean coverage is 72.2%. After reviewing the dataset we found that during the 2 falsely recognized events the participants were talking, and for the single miss-classified eating event the participant was eating a frozen yogurt. Since our models was trained on chewing instances, this explains why events that don't contain regular chewing such as eating ice cream or soup cannot be fully recognized.

As we discussed earlier, the filtering step was added to help reduce the number of false recognized eating events. To evaluate the effect of this filtering step, we also ran our analysis after excluding it from the pipeline. As expected, the number of false positives increased to 10 for the semi-controlled lab dataset and 20 for outside-the-lab dataset.

4.5 Discussion

The overall results from the semi-controlled lab study and outside-the-lab study show that EarBit was successful in detecting eating with high accuracy outside the lab. EarBit was able to recognize accurately almost all eating events in both environments we tested it on. The sole falsely recognized eating event was eating frozen yogurt, which doesn't contain the regular chewing activity that our model is trained on. The high event coverage values (89.6% in-the-lab and 72.2% outside-the-lab) indicate EarBit capability in automating the food journaling process with a precise logging of meals and snack duration's. EarBit also requires about a minute to recognize an eating episode. This low delay values allows EarBit to be used in applications that require just-in-time interventions.

4.5.1 What it means for the end user?

The outside-the-lab study has 45 hours of recorded data. In this duration, EarBit had only 2 falsely recognized eating events. If we assumed that a typical user sleeps for 8 hours a day, our dataset has approximately 3 days worth of daily activities. That means that EarBit generates 0.7 false positives per day. For a typical user who eats 3 to 6 means and snacks daily, the false positives do not pose a significant usability challenge. Although this extrapolation would not always be accurate, it provides a reasonable trend of the results.

By reviewing our outside-the-lab dataset, we found that the falsely recognized events are mostly due to talking activities. After visualizing the entire dataset, we found a total of 26 talking events. EarBit has only classified 7.6% of them as eating. We believe the features set we used helped in correctly recognizing most of these events as non-eating, but using EarBit with a modified user interface can improve its precision by incorporating more data from the user. For example, as soon as EarBit detects an eating event it can prompt the user with a question ”Are you eating?”, if the user’s response was positive the system carries on with the food journaling process, but if it was negative the system can ask the user for a label ”So what are you doing?” and then utilize this instance to generate a better user adaptive model.

4.5.2 Study design

Eating detection in most laboratory settings lacks ecological validity. At the same time it is often hard to collect accurate data in unconstrained environments. Our study design aimed to solve both problems. Researchers equipped the Aware Home for recording and monitoring various eating scenarios. At the same time, the nature of a house facilitates normal interactions and eating behaviors. Thus, the researcher is able to control the environment while the participant behaves in a more natural manner. However, it was obvious to the participants that they were video recorded and the researchers were present as well. These factors meant that the setting wasn’t entirely natural. Moreover, the proportion of eating events was higher than an average day in a user’s life. We addressed some of these issues in the outside-the-lab study. As the participants used the system in their own environments, the proportion of eating events was more natural in this study, but they had a chest-mounted camera for groundtruth. Hence, the data collection was not entirely naturalistic here as well. We believe our fine-grained labeling of activities, and the protocol of training and evaluating the model on data from significantly different settings produced repeatable and generalizable results. However, the quest for a true evaluation of eating activity in unconstrained environments remains unfinished.

We believe our study can serve as a good starting point for future studies on eating detection, and we hope other researchers use and improve our pipeline to detect activities - like eating - in unconstrained environments.

4.5.3 Self Reporting

Self-reported eating is the predominant method used to record eating in unconstrained environments. However, this method of reporting is known to be inaccurate. For example, in a laboratory setting [6] found that both people with and without eating disorders under-reported eating.

During our study, we found multiple issues with self reporting. When comparing ground truth between video footage and self report obtained from collecting data in unconstrained environments , we found that some participants forgot to report eating episodes, reported best guess eating times, and/or reported best guess eating duration. One participant reported the following:
”*1:00 A.M.: Snacking some during movie*
19:32 snacking some more
(There was probably more but I don’t remember how long it went)”
Another participant said, ”*I forgot I was wearing the device and got caught up in a conversation we were having over lunch, so I totally forgot to write down what time I started eating. I think I ate for about 30 minutes*”. Participants in the study were provided monetarily incentives to report

eating activity, yet on occasion they still forgot to report. From this discussion, it is probable that many studies involving self-reported eating suffer from inaccurate and incomplete data. Since our evaluation tests the system's performance on how accurately it recognizes chewing instances and eating events, we had to obtain more reliable ground truth. To overcome this issue, we decided to equip participants with a wearable camera to record their activities outside the lab. This condition imposed some limitation on the session duration due to the short battery life of the camera. The camera also can impose some restriction on the user behavior, but we believe this is a reasonable compromise for obtaining a reliable ground truth in unconstrained environments.

4.5.4 Form Factors

During our pilot study, we realized that in some cases the behind-the-ear IMU was not placed properly and was floating. Almost half of the earpiece was above the pinna, instead of being behind it. This issue meant that the sensor was not coupled to the temporalis muscle. We solved this issue by demonstrating the correct way to put the device to our participants and giving clear instructions to make sure that the sensor is placed properly. We largely succeeded in making sure there were no placement issues and a review of the video footage showed that there were no visible placement issues with the sensors. However, when a device like EarBit is used in the real world, it would be important for the system to be resilient and adaptive to placement issues. In our future prototypes, we are experimenting with embedding the sensor in eye-glasses and using firmer silicone mounts in case of earbuds.

Chapter 5

FitByte

5.1 Introduction

Based on the findings of the EarBit evaluation [11], we defined several challenges that are important to tackle. Addressing these challenges ensures better performance for ADM systems in real-world scenarios. First, it's difficult to collect naturalistic data without having access to a facility like the Aware Home. Second, EarBit doesn't detect drinking and eating events that don't require chewing, such as eating ice cream, yogurt, and soup. Last, EarBit doesn't provide users with additional information about food type or amount. All these uncovered challenges informed the design of my second ADM systems FitByte [12].

FitByte, is a pair of eyeglasses that tracks the wearer's food consumption using multi-modal sensing to capture all food consumption actions. FitByte (Figure 5.1) 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 5.3); 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

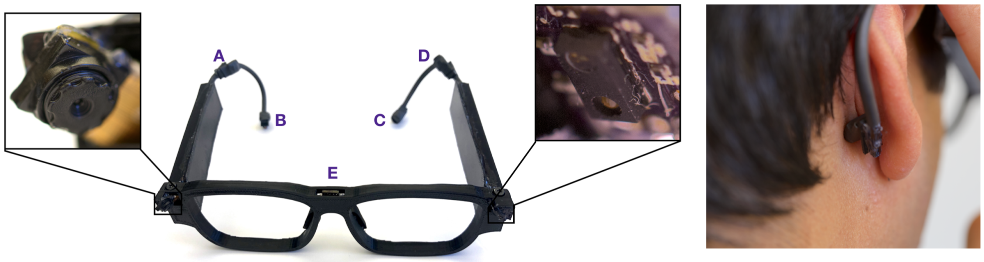

Figure 5.1: FitByte hardware. The device has one camera, one proximity sensor, and six IMUs. One IMU each at A, B, D, and E. At C, there are two IMUs: one gyroscope and one 4 kHz accelerometer to measure body vibrations behind the ear. The left temple houses the battery and the right one has the microcontroller. 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.

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.

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 [55, 63].

As Figure 5.2 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

	Jaw Motion	Hand Gesture	Swallow Sound	Chew Sound	Food Images
Amft *et al.* [3]				●	
Zhang *et al.* [34-36]	●				
Chung *et al.* [12]	●				
Farooq *et al.* [14]	●				
Bedri *et al.* [5,6]	●				
Thomaz *et al.* [29,30]		●			●
Olubanjo *et al.* [21]			●		
Rahman *et al.* [25]			●	●	
Yatani *et al.* [33]			●		
Sen *et al.* [28]		●			●
Liu *et al.* [17]				●	●
Bedri *et al.* [4]	●				
Bi *et al.* [8]	●				
Mirtchouk *et al.* [19,20]	●	●		●	
FitByte	●	●	●	●	●

Figure 5.2: 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.

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.* [45, 46] 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.

Figure 5.3: Samples from FitByte's on-board camera for food intake moments.

5.2 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 5.2). In this section, we describe the utility of different hardware components of FitByte.

5.2.1 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[1]. 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.

5.2.2 Sensors

Existing diet monitoring approaches have mostly focused on detecting one food intake action [54]. 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 [67]. 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 5.4). 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 [10, 11, 18]. 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 5.1: 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

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

a fifth gyroscope in the nose bridge (Figure 5.1: 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 5.1); 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 5.4). 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.

Camera

To help capture visuals of the consumed food, we use a miniaturized camera (Adafruit Mini Spy camera (480p video and 1280×720 photo)[2]. We placed the camera at the top-right corner of the frame to capture activities around the mouth region (Figure 5.3). 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.

5.2.3 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 5.1). 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.

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

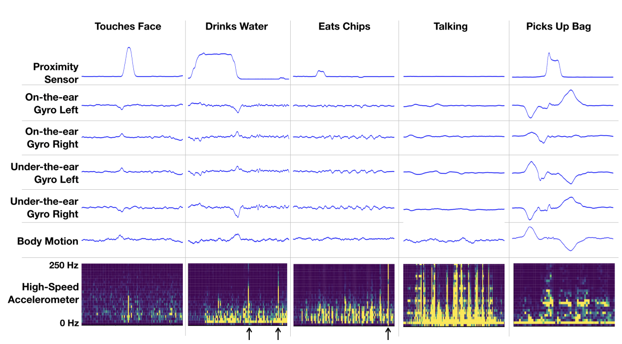

Figure 5.4: Signals from FitByte's sensors as the user performs different activities. The point in times marked by the vertical arrows at the bottom indicate swallows.

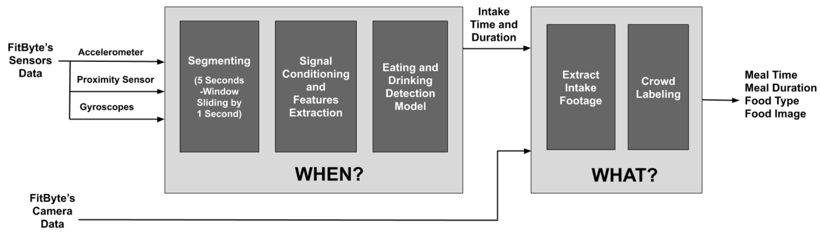

Figure 5.5: FitByte's machine learning pipeline

5.2.4 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 5.1 - *(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.

5.3 Algorithm

In this section, we explain our signal processing and machine learning approach to detect when food is consumed (Figure 5.5). 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.

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.

5.3.1 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 5.4 ”Eats Chips”). Bedri *et al.* [11] 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 5.4 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.

5.3.2 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, 45]

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.

5.3.3 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.

5.4 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, 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.

5.4.1 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 [11, 18]. 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 5.6). 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[3]. 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.

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

Figure 5.6: 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.

5.4.2 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.

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.

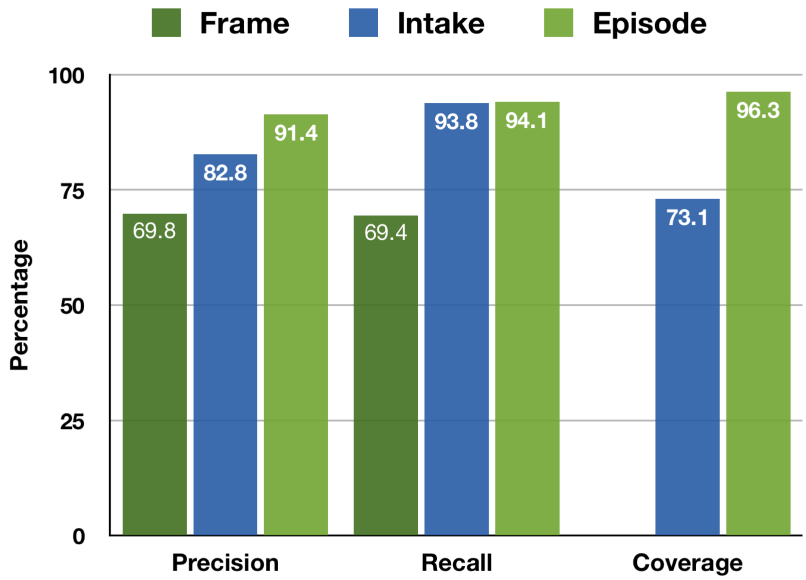

Figure 5.7: Eating Detection Results for the semi-constrained study. Frame-level results will not have coverage because those inferences are already made at a 1 second resolution.

5.5 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.

5.5.1 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 5.7 provides a summary of the results.

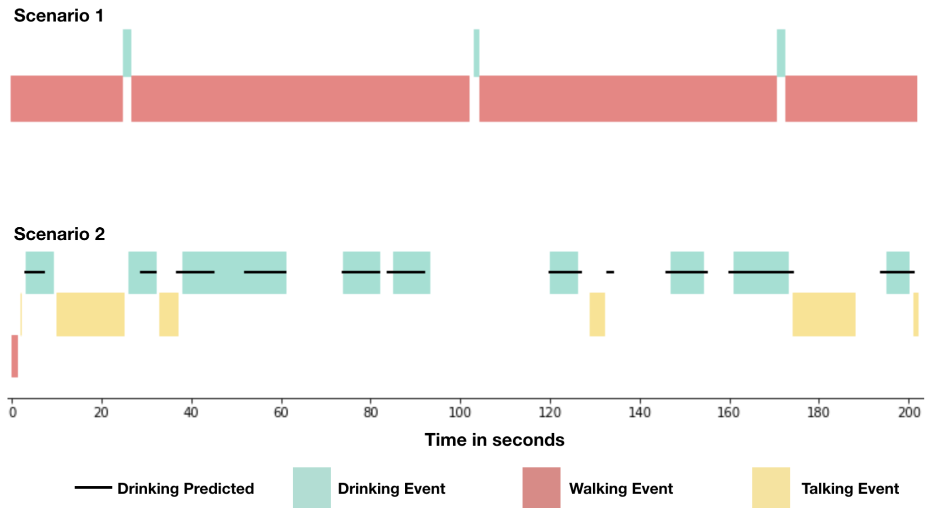

Figure 5.8: Shows timeline of two scenarios from the user study. *(Top)* FitByte fails to detect drinking activity when the user occasionally sips liquid while walking. *(Bottom)* However, FitByte succeeds in detecting drinking episodes when the user drinks for longer and drinking is not completely occluded by other activities.

5.5.2 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 5.8 shows an actual scenario from our data collection for the two cases. While FitByte fails at detecting situations like Scenario 1 in Figure 5.8, 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.

5.5.3 Identifying Food Type

We triggered the camera using FitByte's IMUs and proximity sensors to capture food visuals (Figure 5.3). 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[4].

5.5.4 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 5.9 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 5.10), 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 5.10). 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.

5.5.5 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.

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

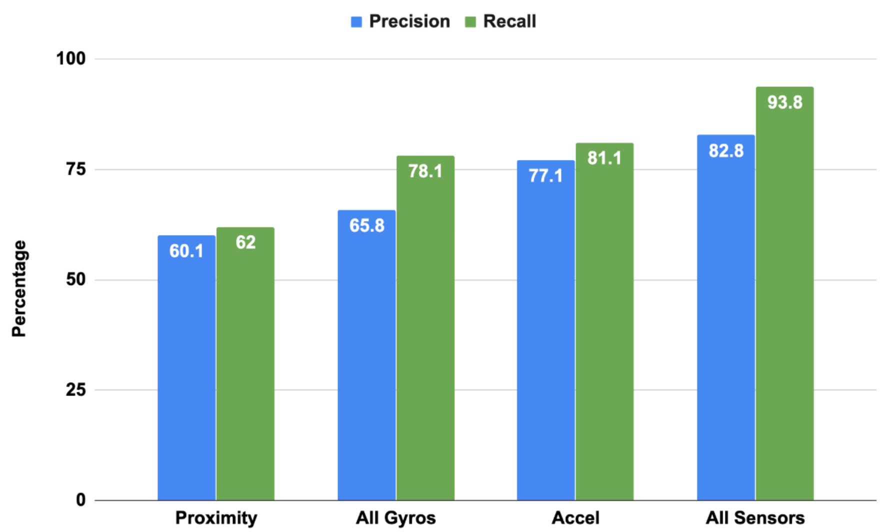

Figure 5.9: Performance of different sensing modalities compared to the performance of all sensors in the semi-constrained Study

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).

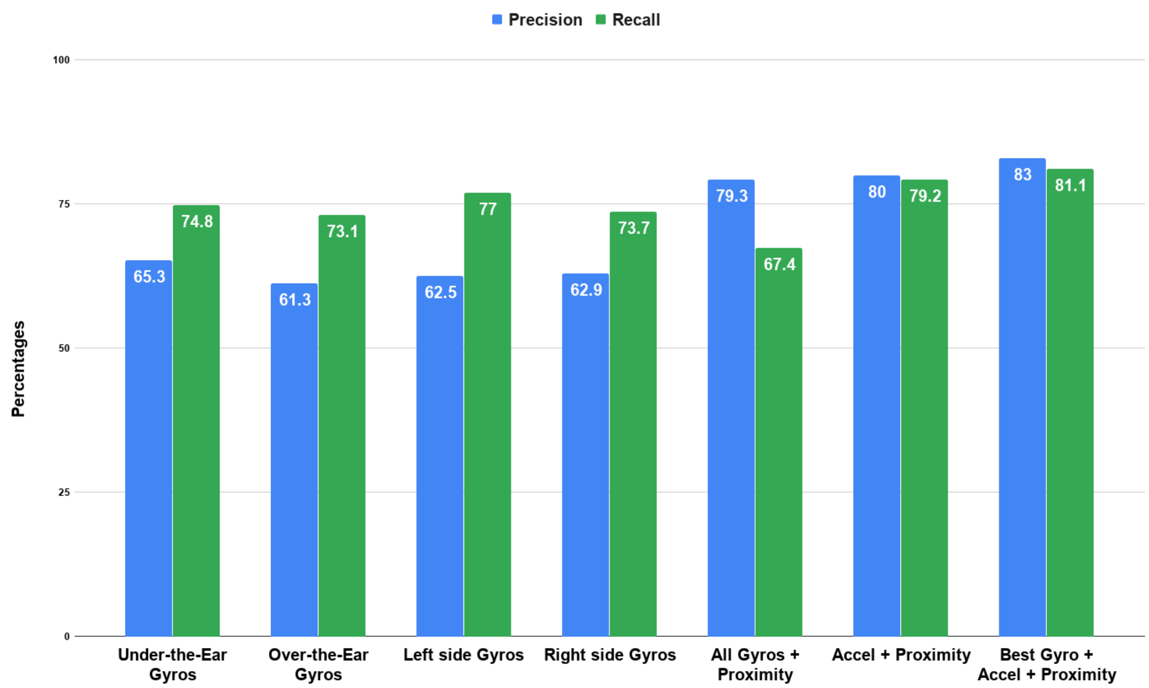

Figure 5.10: Performance of different subsets of sensors in the semi-constrained Study

5.5.6 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.

5.5.7 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.

5.6 Estimating the amount

To evaluate FitByte's ability to estimate the food amount we ran a separate study.

Food Amount Study Design

Given capturing ground truth information about food amount is hard, this study was significantly more controlled than the main study (described earlier). We limited the food items in this study to four types each has a different texture. We recruited 10 participants (7 males, 3 females) and provided them with 1 slice of cheese pizza (*soft*), half an apple (*crunchy*), one Activia yogurt (*saggy*) and one cup of water (*fluid*). The participants were free to eat any or all of the food items in any order they preferred. Most participants switched between food items in a random order. Other than positioning the food on the scales there were no other restrictions to the participant's behavior. In the study, participants talked to friends, worked on their laptop or used their phone. To obtain ground truth for the consumed food volume at every intake, we used two food scales with 0.01 gram sensitivity. One scale was dedicated to water cup, and the rest of the food items were on one plate placed on the second scale (Figure 5.11). A camera was used to record the food

Figure 5.11: The setup for estimating the food amount study

weight from the scale, and the onboard camera captured the participant's intake. Also, there were no restrictions on the order in which the participants should eat the food items. The study began by asking the participant to wear the FitByte system and ended once the participant finished their meal. The duration of the study ranged from 10 to 20 minutes.

5.6.1 Food Amount Annotations

Using the video footage recorded of the user and the scales all session were annotated using the Elan 5.2 software. Annotations were made for the duration of the intake instances defined by the moment the participant brings the food to their mouth and the moment they stop chewing (or swallowing in case of water). Each labeled intake were then annotated with the food type and the amount consumed. Other non eating activities were not annotated.

5.6.2 Food Amount Analysis

Estimating the food amount undergoes three steps

1. Identifying food consumption moments and food type using the pipeline described in Section 5.3.1–5.3.3.
2. Estimating the amount from the recognized intakes.

In the first step, we passed the data we collected in this study through the validated final model to detect eating and drinking instances at the intake level without any hyper-parameter tuning. The outcome of this phase provided us with the predicted intake instance and the coverage of each instance. Since the meal we prepared contained 4 different food items we had to identify the food type at every intake instance. For that we extract footage captured every time the proximity sensor detects a beginning of an intake gesture and had a research volunteer label the food item they saw in each footage.

In the second step we used our predicted intake duration to train regression models to estimate the amount of consumed food for each food type (*i.e.,* pizza, apple, yogurt and water).

The features we used to train the regression models are similar to the features we used in detecting food consumption moments but we added two new features: the duration of the intake and the estimated speed of chewing – computed by dividing the number of peaks per intake by the duration of the intake. We trained a Gradient Boosting Regressor with 100 estimators and the least squares loss function using the ground truth volume corresponding to every predicted intake. The models were validated using leave-one-user-out-cross-validation. We computed the average percentage absolute error at the intake level, the item level, and the meal level as shown in equation 5.1.

$$PAE_{\alpha} = \frac{1}{n}\sum_{i=1}^{n}\frac{abs(AW_{i,\alpha} - PW_{i,\alpha})}{AW_{i,\alpha}} \tag{5.1}$$

$Where$:

PAE: is the percentage absolute error.
AW: is the actual weight.
PW: is the predicted weight.
n: is number of participants.
α: can be intake, item or meal.

5.6.3 Food Amount Results

Using the trained and validated FitByte models for detecting food intake moments, we were able to detect 262 intakes from a total of 275 intakes in the dataset and had only 8 falsely recognized intakes (95.3% recall and 97.0% precision). The average coverage for every intake is 96.1%. A crowd worker annotated the food type at every intake using only a thumbnail image of the video recorded at the intake. The crowd worker had the option of visualizing a short video of the intake if not sure about the type but they did not use this feature and recognized the types from only one image captured at the beginning of the intake. The overall accuracy in identifying food type at every intake was 93.9%.

To estimate the amount (in terms of weight) of food consumed, Fitbyte had an average percentage absolute error (PAE) of 20.8% ($\pm 16.3\%$). The average weight of the meal in the study is 531.8 g. The average PAE for estimating the volume of each food item is shown in table 5.1 along with the average PAE for every intake.

Compared to other approaches for estimating the amount [2, 45] FitByte doesn't relay on manually identified chewing count or use the ground truth of intake duration because it infers this information from the data automatically.

5.7 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

Type	Item MAE %	Intake MAE %
Pizza	26.6	37.0
Apple	25.1	42.5
Yogurt	28.1	46.5
Water	24.1	39.8
Meal	20.8	41.1

Table 5.1: The average percentage absolute error at the item level and the intake level

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 5.9 and 5.10). 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.

5.7.1 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.

5.7.2 Estimating the Food Amount

Moreover, FitByte also accurately (96.3%) detected the duration of the eating episode and the number of intakes (93.8%). This result is very encouraging as it can be a stepping stone to detect how much a person ate. We believe that for each person, once we know the food type, the number of intakes and duration of their eating episode will have a strong correlation with quantity. Therefore, FitByte can enable very accurate estimates of quantity of food consumed. Using the same pipeline to detect food consumption moments and identifying the food type we ran a semi-controlled study to evaluate the ability of FitByte in estimating the amount of a predefined meal. With only 10 samples the Percentage Absolute Error for estimating the meal amount was 20.8% which indicates how the high accuracy of detecting food intake instance and estimating their duration can provide a reasonable estimate to the amount food consumed. This shows that by collecting a larger and richer dataset, FitByte has the potential of estimating the amount consumed for a wide range of food items with higher accuracy. By providing an easy way to identify food type and estimate meal amount, FitByte can enable users to estimate calorie intake using nutrition databases such as the USDA Nutrient Database.

5.7.3 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.

5.7.4 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 [60, 70]. 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.

5.7.5 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 5.1) to give a reasonable swallowing signal as seen in Figure 5.4. We chose to place the sensor at B to have the sensor closer to the processor and avoid inducing noise into the circuit by carrying high-frequency signals across the frame.

Fitting was another challenge. Making sure we always get a good signal from all sensors required several iterations on the design. For the IMU sensors, we tried adjustable 3D printed arms and adhesive silicone attachments, but they either broke or were not durable after long use. We finally chose the combination of the adjustable copper wire arms and the flexible 3D printed fulcrums to hold the sensors in place. In future designs, we plan to provide users with an app that allows them to visualize sensor data in real-time to verify proper sensor placement. We also plan on putting the proximity sensor and the camera on a simple gimbal that the user can adjust if needed.

Chapter 6

FitNibble

6.1 Introduction

In this chapter, I discuss my attempt to address **RQ2** and assess the utility and usability of ADM and its impact on food journaling compliance. In this effort, I developed an end-to-end system (FitNibble) that allows users to receive just-in-time notifications to prompt them with logging every time it detects they are eating. This system and its evaluation are based on the outcomes of the FitByte and EarBit studies [11, 12].

In academic research, several efforts have been made towards developing automatic diet monitoring systems (ADM), but these proposed solutions were harder to implement in a real-world product. Researchers have cited several technical and non-technical challenges, which represent clear barriers for this technology to reach the end-users [63]. The major technical challenge stems from the complex nature of the diet monitoring task, as fundamentally the user needs to keep track of *when* they eat, *what* they eat, and *how much* they eat. Such detailed tracking makes automating the food journaling process exponentially difficult when compared to step counting or sleep tracking. While detecting food type (*What?*) and amount (*How much?*) remain as open questions, many advancements have been made in detecting *When* people eat and for how long, using wearable ADM systems with different sensing approaches to track chewing, swallowing, and/or hand-to-mouth gestures using modalities like inertial sensors, microphones, EMG, and proximity sensors [56, 63]. These systems have been developed and tested in lab environments to validate their functionality, but ecological validity remains a great challenge for most of these ADM setups. Non-technical challenges have also manifested when evaluating these ADM systems in public, which include social acceptability of the form factor and privacy concerns. All these challenges have formed a barrier for researchers to evaluate the utility and usability of ADM systems because these metrics are difficult to assess without a reliable end-to-end system. To allow for just-in-time interventions the ADM system should also be accurate at detecting the onset of an eating event even if it's short. For instance, snacking is usually underreported with other journaling methods despite the high implications it may have on people's health. Finally, the ADM system should also have a practical and socially acceptable form factor, to help researchers capture the real impact ADM can have on the user experience.

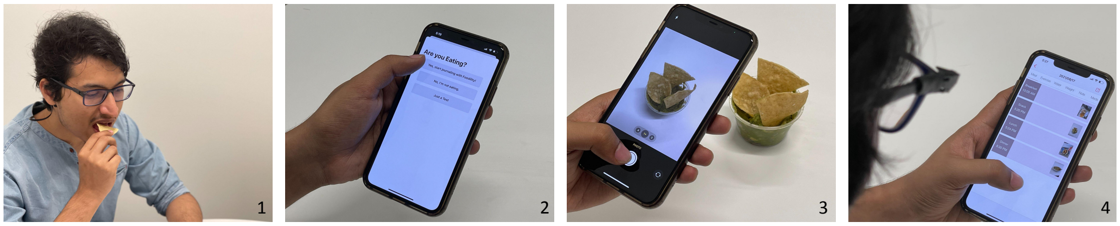

Figure 6.1: With FitNibble users can receive just-in-time notifications reminding them to log their meals and snack. This figure show the flow of the user experience with our system

6.1.1 Food Journaling Challenges

Research exploring the food journaling challenges found that the low adherence rates are caused by the perceived tediousness of the journaling task and loss of motivation due to the high rate of missed events. Missing to log several eating events results in an incomplete journal, which invalidates its use in understanding dietary patterns. When investigating why people miss logging events, the majority of participants said they simply forget, as remembering to log your meals requires constant attention to your daily activities, which to many individuals introduces an unbearable cognitive load [20, 21]. We believe that using a reliable ADM system to detect *when* people are eating, would allow for just-in-time intervention to prompt users to log. This feature should ease the food journaling process by offloading the user's attention from tracking their activity to other tasks. This system should have a positive impact on adherence allowing users to have more complete journals by reducing the number of missed events.

6.1.2 FitNibble Design

In this evaluation, we designed a field study to assess the utility and usability of *FitNibble*, an end-to-end wearable ADM system designed to recognize eating events and send just-in-time notifications to prompt users to record meals and snacks at the right time. *FitNibble* design is based on the FitByte [12] and EarBit platforms [11]

As explained in chapter 5, the FitByte platform has seven sensors to track chewing, swallowing, and hand-to-mouth gestures and use it as a proxy to detect eating and drinking events. The setup also has an onboard camera to capture images of the food.

To build a deployable setup, we made a few adjustments to the design and the system architecture. These decisions are based on the feedback received from FitByte and EarBit study participants. These design changes were required to ensure reliable performance in a free-living environment and to improve social acceptability.

First, we decided to use the user's glasses as the platform and attach to them our lightweight sensors. This change was recommended by the FitByte users, who preferred to wear their own glasses over adjusting to a new frame. We iterated on the design and ran pilots to ensures comfort and social acceptability and found the setup shown in figure 6.3 to be the most suitable.

Second, we decided to drop the high-speed accelerometer due to its high power requirements. This sensor played a major role in detecting drinking events. In this evaluation, we found it suitable to focus on eating events detection. After testing the setup in a short pilot, we also decided to drop the onboard camera. Participants found the setup bulky and heavy to carry, and some participants didn't want to have a camera on the setup for privacy reasons.

Finally, It was clear from our analysis that the proximity sensor played a major role in reducing the average prediction delay to 7 seconds. This feature is valuable to design just-in-time interventions therefore we decided to keep the proximity sensor. We also kept one of the 3D gyroscope sensors (bottom right) to detect chewing from jaw motion. For motion reference, we used another 3D gyroscope sensor attached to the back of the user. All these sensors are sampled at 10 Hz, which we found to be the best data rate to reduce power consumption and ensure reliable performance. For more details on the final design, please check section 6.3.2.

6.1.3 Approach

In this evaluation, we focus on the role ADM can play in mitigating some of the challenges of food journaling and in improving adherence to the process. Our evaluation was done on a long-term field study that has two phases. In the first phase, we introduce participants to traditional self-report food journaling methods using a custom iOS application. The application allows users to set reminders to log, take photos of the food, and review daily activities in a calendar format. We gave participants two days to get acquainted with the app and the journaling process. We then collected data for seven more days for the actual evaluation (9 days total). During the study, we asked participants to fill in a daily survey to reflect on their experience with the app(i.e. Experience sampling). After finishing this phase we interviewed participants to assess their overall experience. In the second phase, we introduce participants to the ADM wearable setup, which can be linked to the journaling app, and send notifications to the user to prompt them to log in whenever it detects they are eating. In a similar fashion to the first phase, participants have also used the new setup for 9 days and were asked to fill in a daily survey to assess their experience with this new method. At the end of the study, we conducted a second round of interviews with the participants to assess their overall experience in phase 2 (with FitByte) and how it compared to phase one (without FitByte).

For analysis, we followed a mixed-methods approach on the qualitative and quantitative data we collected. We evaluated the user experience with different measures including utility, usability, effect on adherence to food journaling, social acceptability, and privacy concerns. Our analysis has shown that in this short period *FitNibble* has improved adherence by significantly reducing the number of missed events (19.6% improvement, $p = .013$), and participants have exhibited clear dependency on the wearable as soon as they started using it. Journaling difficulty have also dropped significantly after using FitNibble ($p = .005$). The device also helped the majority of our participants discover new dietary patterns especially with the amount of snacking, and we started to see signs of behavioral change due to increased awareness of eating habits. All these outcomes underscore the importance of ADM in improving the food journaling experience.

The main research contributions of this effort can be summarized in the following points:

1. Improved machine learning models for real-time eating detection using the publicly available FitByte dataset.
2. An end-to-end open-sourced system (including wearable schematics and firmware, smartphone app, and backend).

3. A field study to assess the utility and usability of a diet monitoring wearable, and provide a list of recommendations for future ADM system's design.

6.2 Method

As many technical ADM challenges have been addressed in previous work [11, 12, 75, 76], a long-term field study can help ADM researchers explore the nontechnical challenges that can influence the adoption of this technology.

The main goal of this evaluation is to assess the value ADM can provide to the food journaling experience, and understand the influence of the system performance on usability and utility.

In this evaluation, we aim to use an adapted version of FitByte to send just-in-time notifications to prompt the users to do the logs when it detects they are eating. We hypothesize that by using this approach we can reduce the cognitive load required by self-report journaling methods, and reduce the number of logging errors especially when it comes to missed events. Having this tool should significantly improve the user experience and help them adhere to the food journaling process.

In our analysis, we used several metrics to assess the experience with food journaling using ADM. These metrics include utility, usability, social acceptability, and any privacy concerns raised from using this system. We will use a combination of quantitative and qualitative methods to assess these metrics and use the results to inform new design recommendations for ADM.

6.2.1 Study description

To evaluate the utility and usability of our ADM setup we designed a field study that allows participants to experience food journaling with and without ADM. In this study we targeted individuals who are interested in understanding their dietary behavior in general and not focused on specific goals like weight loss. The second criteria we had was to only recruit individuals who wear eyeglasses regularly because our ADM setup is based on that form factor and we didn't want the experience of the participants to be influenced by their unfamiliarity with wearing eyeglasses all day.

The study has two phases, each phase should last for 9 days (18 days total). In this study, participants will be introduced to a traditional photo-based food journaling method. This method requires the user to just take photos of their meals and snacks throughout the day and review them before they go to sleep. This approach is found to deliver sufficient information to help them understand their dietary patterns without focusing on minute details like calorie count. We chose this approach because it requires minimum effort and for users who are not focused on specific goals like weight loss, asking them to keep track of many details can be overwhelming and in some cases anxiety producing [20].

Phase1: Photo-based journaling without ADM

In the first phase, participants will be asked to use the FitByte app (check description in section 6.3.3) to log their meals and snacks. In this phase the app won't be linked to the wearable, which will require the user to remember by themselves to log every time they eat. In the app there is a feature that would allow users to set reminders at specific times. We added this feature to help users to remember to log if they know at what times they are most likely to eat. To help participants distinguish between snacks and meals we defined snacks as any eating event outside the main meals breakfast, lunch, and dinner. Participants can also log drinking events but it is not required.

We require participants to use this journaling method for 9 days. In this period, they will have two days to get familiar with the journaling method and 7 days for the actual data collection. At the end of each day participants are asked to fill a short daily survey for experience sampling. The survey is designed to encourage participants to review and reflect on their logs for the day. For the survey questions, check appendix A & B. At the end of this phase we conducted a semi-structured interview with each participant to understand their experience with this journaling method with a specific focus on utility, usability, social acceptability, and impact on adherence. For the interview questions, check Appendix D.

Phase2: Photo-based journaling with ADM

In the second phase, participants are introduced to our wearable ADM setup. At the beginning of this phase we install the hardware setup on the user's glasses as described in section 6.3.2. We then explain to the participant how to put on the wearable setup and explain to them how to connect the wearable to the phone app via Bluetooth. This new setup should allow the app to send reminder notifications to the user every time it thinks they are eating and ask them to log.

Before we start the study we ensure the system is working properly by performing a functionality test. In the test, we ask the user to simulate an eating event by chewing and performing several hand-to-mouth gestures at the same time. We run this test multiple times to ensure that the system is detecting eating events reliably.

Similar to the first phase, participants will use this journaling method for 9 days (2 days to get familiar with the method and 7 days for actual data collection). Participants are also asked to perform the same tasks of phase by logging their eating events and filling in the daily survey. In this phase, we still asked participants to remember to log by themselves and not rely on the wearable notifications. Participants were still able to set reminders on the app at a specific time.

At the end of this phase participants are invited again for a semi structured interview to reflect on their experience with the new journaling method and how it compares to the first one. For the interview questions, check Appendix D. To analyze the interview data we used an emergent thematic coding method.

6.2.2 Measures and study facts

In addition to the data we collected from the interviews and the daily surveys, we also collected app usage data in both phases. This helped us understand how frequently participants were using the app and how often they used specific features, such as the time reminder. We also recorded their responses to the wearable notifications to help us track the number of true positives and false positives per day. In order to preserve participants' privacy, none of the logs data and photos were recorded .

In addition to the app usage data, we asked participants to fill in a standard self efficacy scale survey at the beginning and end of the study [62]. We collected this data to gauge the influence of self efficacy on the user experience and see if their self efficacy rating changed by the end of the study.

We recruited 13 participants (5 female and 8 male) with an average age of 34 years, ranging between 21 years and 54 years. The majority of the participants (6) were university students but the remaining participants (7) were from diverse backgrounds including school teachers, artists, and stay at home parents. Since we conducted this study during the COVID 19 pandemic and as lock down restrictions were lifted in the city, we asked participants to report their expected level of activity for the 18 days of the study including if they will work from home or office, how many times they will leave their homes and how many times they will eat outside (Appendix C). In our recruitment we didn't require participants to have a certain level of activity.

6.3 System Description

6.3.1 System architecture

This section discusses the overall system architecture of the *FitNibble* Deployment. The system has three fundamental components: the wearable, journaling App and the backend server (Figure 6.2). The wearable handels sensor data, preprocesses them, computes the model features, and sends it to the smartphone via Bluetooth. On the phone the custom iOS App we developed handels the features and sends it to a server which will run predictions on it and send the results back to the smartphone App. The following subsections will explain each component and its functionality in detail.

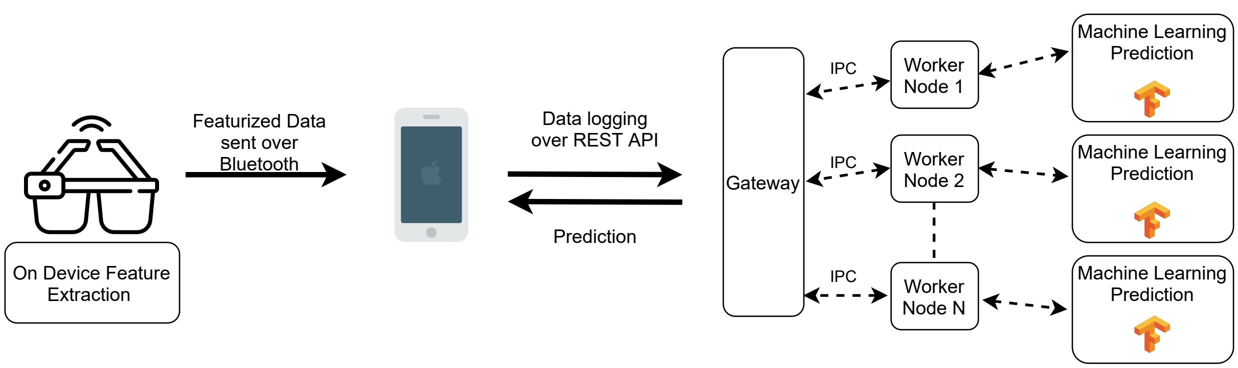

Figure 6.2: System architecture: FitByte2.0 wearable sends the extracted features over Bluetooth to our IOS mobile App. Our FitByte backend obtains this data over REST API through an iOS app and then forwards it for data logging and model prediction of eating.

6.3.2 Wearable

The FitNibble wearable we developed is informed by the original FitByte design. We also adopted the design recommendations published in the FitByte paper [12], which are based on the user feedback they received after using their wearable in a short deployment. In this feedback participants recommended that the wearable should be lightweight and less bulky similar to the regular eyeglasses they wear.

The original FitByte wearable served as a research platform to explore the use of several sensing modalities to detect eating and drinking. The FitByte platform has 5 gyroscopes to help detect chewing from different locations around the ear, a high speed accelerometer sampled at $4kHz$ to detect swallowing, a proximity sensor pointing at the mouth region to detect hand-to-mouth gestures, and an on-board camera to capture food images This platform allowed us to understand how different sensor combinations can influence the overall accuracy and at what expenses.

For the FitNibble design we were considering a light weight design that delivers high accuracy in detecting eating without consuming a lot of energy. These guidelines were set to ensure we have reliable hardware that can be deployed for several days without close supervision. For these reasons we decided to drop the high speed accelerometer and camera because of their high power consumption. The high speed accelerometer was originally added to improve drinking detection, but in this study, we are only focused on detecting eating events which can be done reliably with other sensors in the platform (check FitByte sensors combination results). Since the 5 gyroscopes were used to explore the best placement for the chewing sensor, we only kept two sensors, one at the bottom right side to detect chewing from the lower jaw bone and we used the second gyroscope as a reference sensor to filter out body motion. We also kept the proximity sensor due to the value it provides in detecting the eating onset, which will help in introducing just-in-time interventions.

We experimented with different placements for the gyroscope sensor including on-the-temple or above-the-ear, but we found they cause headaches after wearing them for a few hours, so we decided to keep the chewing sensor at bottom of the ear and the reference sensor was placed with the microcontroller board and battery in small pack attached to the back of the users shirt.

In the first design iterations we considered adding the camera to the FitNibble platform and dedicate a separate battery pack for it, but we ran a pilot with this version, and participants found the device to be heavy and bulky. Also a few participants requested the camera to be removed at the beginning of the study for privacy reasons. For all these reasons we decided not to include the camera on the final deployed version.

The final FitNibble setup has a proximity sensor (VCNL 4040) hosted in a small 3d printed holder. The holder is attached to the right hinge of the user's glasses as shown in figure 6.3.left. On the same side a 3-dimensional gyroscope (MPU9250) is attached to a flexible adjustable arm linked to the glass's temple tip (Figure 6.3.right). Similar to the original FitByte setup we used a 12 gauge solid copper cable, which provided a wide range of fitting possibilities. These two sensors are connected to the same I2C cable, which extend to a cloth pocket attached to the back of the user collar (Figure 6.3.right). The pocket hosts the Bluetooth Low Energy module board (Rigado BMD 350, nRF52832,Arm Cortex-M4), the reference IMU (MPU9250), a 2000 mAh battery, and a battery charging board. We chose this configuration to ensure that the eyeglasses weight is as light as possible, and place all the heavy components on the back which was inspired by the Earbit design [11]. The FitNibble setup is designed to be attached to any pair of eyeglasses.

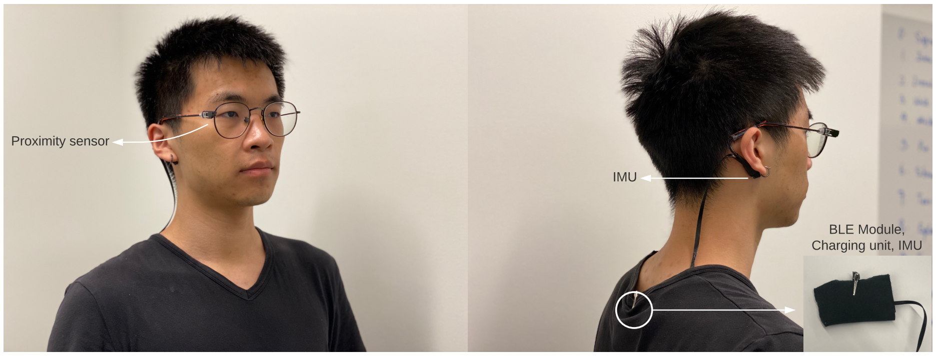

Figure 6.3: FitNibble wearable setup. The system has a proximity sensor to detect hand-to-mouth gestuers (left), an IMU in contact with the lower jaw bone to detect chewing (right), and small cloth pocket clipped to the back of the user's shirt containing the BLE module, battery, and the reference IMU (bottom right)

In the study we chose to use the participants' personal glasses to avoid any discomfort they may have from wearing a different frame.

The BLE module firmware collects data from all three sensors at 10 Hz, preprocesses it, computes the features, and sends it to the phone via bluetooth. Similar to the original FitByte pipeline, the features are computed from a 5 second window sliding by 1 second. The BLE module will send the feature vector to the phone every second. We chose to implement a feature extraction step in the module to reduce the data sending rate and conserve power. Similar to the original FitByte pipeline we preprocessed the data and computed the following features: Entropy,variance, absolute median, zero crossing count, zero crossing variance, and the RMS of each channel (7 channels x 6 features = 42 feature).

The only major change we made to the pipeline was to reduce the sampling rate for these sensors from 50Hz to 10 Hz to extend the battery life. We compared the accuracy for the two sampling rates using the original FitByte dataset and found that it doesn't significantly reduce the frame level accuracy (80% at 50 Hz 77% at 10Hz). Figure 6.4 illustrates all the steps done in each component of the system. The overall power consumption of the system is 25 mAh.

6.3.3 iOS App

On the mobile side, we developed the *FitByteApp* . This iOS app communicates between the *FitNibble* wearable, the backend server, and the user. The App allows users to set time reminders for different meals and snacks. To preserve participant privacy we linked the *FitByteApp* to a secure off-the-shelf journaling App, *Foodility* . The App allows users to do their logs and save the information away from the FitByteApp so the research team doesn't get access to participants' private data. *Foodility* is a simple food journaling app on the App Store that allows users to securely track their food consumption. With *Foodility* , users can select meal types, take short

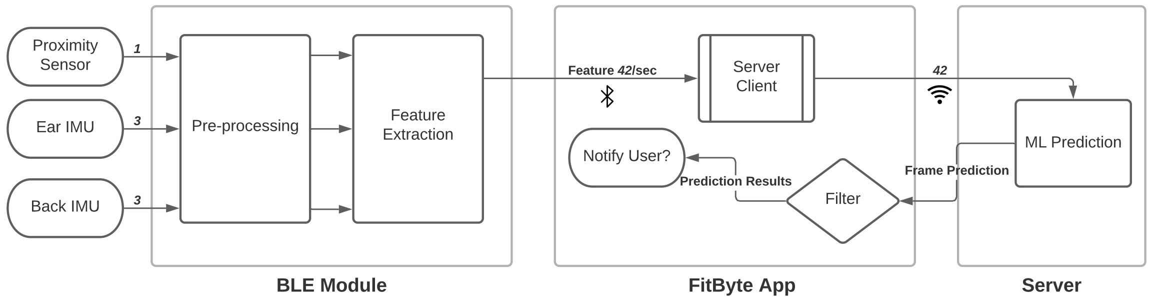

Figure 6.4: Pipeline of FitNibble

notes, and manually log their estimated calorie intake. Moreover, Foodility possesses the feature of taking a photo of the food, which is not offered by most other journaling apps. In this way, the participants can reflect upon their diet at the end of the day by looking at the photos they have taken during that day in the daily view, where all the pictures of their meals and snacks are in one place. On the *FitByteApp* participants can directly launch the *Foodility* app to do the journaling by clicking a button. The app also directly links participants to the required daily survey, and they can also set a reminder that would prompt them to do the survey at a specific time of the day (usually in the evening). The app also possesses the feature of setting daily journaling reminders at specific times, although the participants are not required to use them. After participants have the wearable installed on their glasses in the second phase, the app also handles Bluetooth connections to the wearable. Participants can find a list of Bluetooth devices that fit the characteristics of the wearable after turning on bluetooth pairing in the app, and they can connect to the wearable by tapping their device in the list. If the wearable gets disconnected at any point, the app will try to reconnect with the wearable once it rediscovers the wearable. Behind the scenes, the app receives preprocessed features from the wearable, and sends an API request to the server to get prediction results of whether a participant is eating. If there are 5 consecutive responses that predict the participant is eating, the app would send an instant journaling notification to prompt the user to do the journaling. A device detection notification would stop detecting eating status for 5 minutes, and would resume detection if the user did not interact with the notification within this timeframe. Clicking on the journaling notification (either manually set or through detection) will bring the users to a confirmation page, which has three buttons, “starting journaling”, “no, I’m not eating”, and “just a test.” The users would click “start journaling” if they are eating, and they will be redirected to the *Foodility app* to do the journaling. The user would click “no” if they are not eating when they receive the notification. In the scenario of device detection notifications, clicking no means that the user is receiving a false positive. The “just a test” button is there for the user to do the device checking test every time they put their glasses on. During the second phase, when participants receive a device detection notification, if they click “start journaling” or “no, I’m not eating,” the device would stop detection for 30 minutes so that they would not be bothered with more notifications during the rest of their meals (or snacks). For the purpose of the study, *FitByteApp* tracks users’ high level activities, such as launching the food journaling app, launching the daily survey, bluetooth connectivity, and users' selection on the confirmation page after getting a notification. These activity logs are then sent to the server.

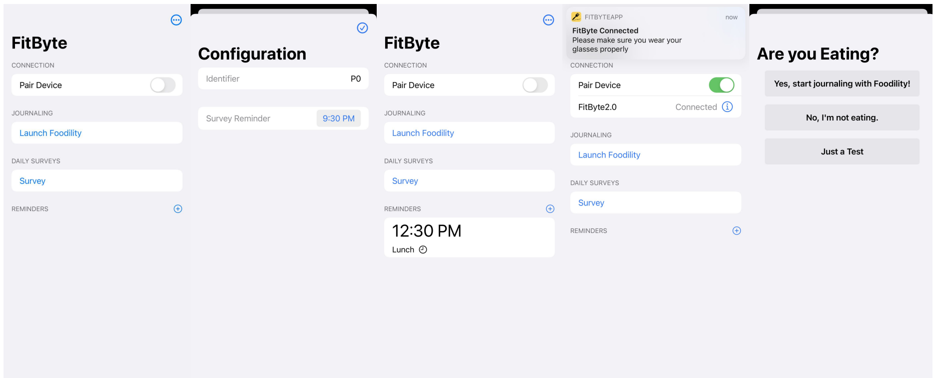

Figure 6.5: FitByteApp features and functionality. This figure shows the main page of the App and explains the main functionalities available on it, including setting reminders and receiving notifications when eating is detected

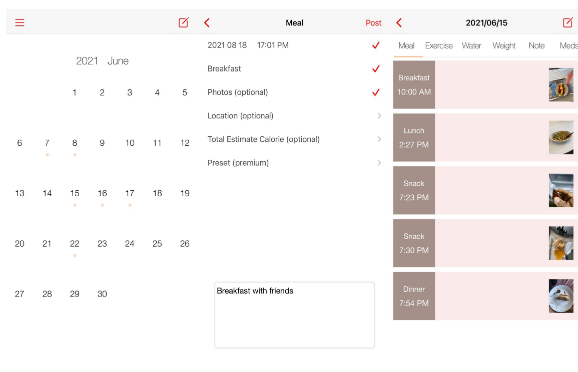

Figure 6.6: Foodility App layout

6.3.4 FitNibble Backend Server

Our backend server is based on a Python-based Flask framework which was custom built to have several functionalities store the device logs from the FitNibble App, obtain the prediction requests from the wearable device, and predict based on the requests. Although some cloud services provide several Rest API ecosystems ([**?**]), they come with several limitations which prevented us from using such platforms. First and foremost, the highest allowed frequency for sending requests is limited within a minute, and the maximum allowed requests (10000) are restricted for a month for the free tier. In addition, such services do not provide the functionality to serve machine learning model prediction requests at scale. This is suboptimal given our high-frequency ML prediction request and the number of participants that we would like to have in our study. Second, our IRB has policies against sending potentially privacy-sensitive data to a server that is not controlled in-house. In fact, such existing frameworks are ambiguous on whether data is visible to third parties. Further, such a system does not encrypt stored data and requires to be connected to one or more of their services to function correctly. Finally, posting data at the scale that we anticipated would have made using such services unnecessarily expensive. Given these requirements, we built our own custom Python-based Rest API server with enhancements optimized for scale and accepting model serving API requests at higher frequencies as shown in (Figure 6.2)

Rest APIs for Data Logging and Eating Detection

We implemented server REST interfaces to our FitNibble backend that were focused on logging the user interactions with the FitByte App. This logging information includes interaction events the participants made when they clicked the daily survey, set up the daily journaling reminders for notifications, and launched the Fidelity App to log their meal events. Each interaction event is stored with participant ID, device ID, the type of interaction, and the timestamp when the participants interacted with the App. This logged information is then used to correlate their interactions with the app as mentioned in 6.3.3.

In addition to these data logs, in Phase 2 of the study, the processed features received from the wearable to the App are then sent as a Rest API request to the server to get prediction results of whether a participant is eating. These requests are sent to the machine learning model serving to process the prediction requests and send the prediction back to the App, which is then shown as a notification.

Scalable data processing and Load Balancing

The rate of prediction request received on the FitNibble Backend from each wearable device through the App is at 1Hz. Such a high frequency of requests needs to have data processing at scale so that the server is able to handle multiple devices simultaneously in the study. To process such requests at low latency, we added a separate load balancer on our server (Figure 6.2) so that it would process all the incoming POST requests from the wearable devices. We investigated several load balancing mechanisms and settled on using the Gunicorn module, as it provides several key features. We configured our server to spin up multiple identical Flask-Python processes when it first starts up, each running on its processor for parallelism. All the API request events are load-balanced across these worker processes, thus allowing many requests at low latency. In addition, to ensure the stability and reliability of our FitNibble server for the duration of the study, we implemented mechanisms to restart any processes that may have failed or aborted. Further, we add vital statistics and logs related to load on the server and resources used that helped us gauge and improve system scalability.

Machine Learning

This machine learning model is trained on the dataset released from the earlier work [12]. The dataset consists of several activities such as eating, drinking, walking, talking, and silence. The previous work [12] achieved a Frame-level Recognition of 69.8% where we detect whether the user is consuming food at a 1 second resolution. This model was trained using a Random Forest classifier with an Scikit-learn implementation (default parameters, 100 trees). However, this model was trained with default parameters and limited configuration. To ensure that there is higher accuracy for eating detection in real world settings we worked on exploring the best model and parameters.

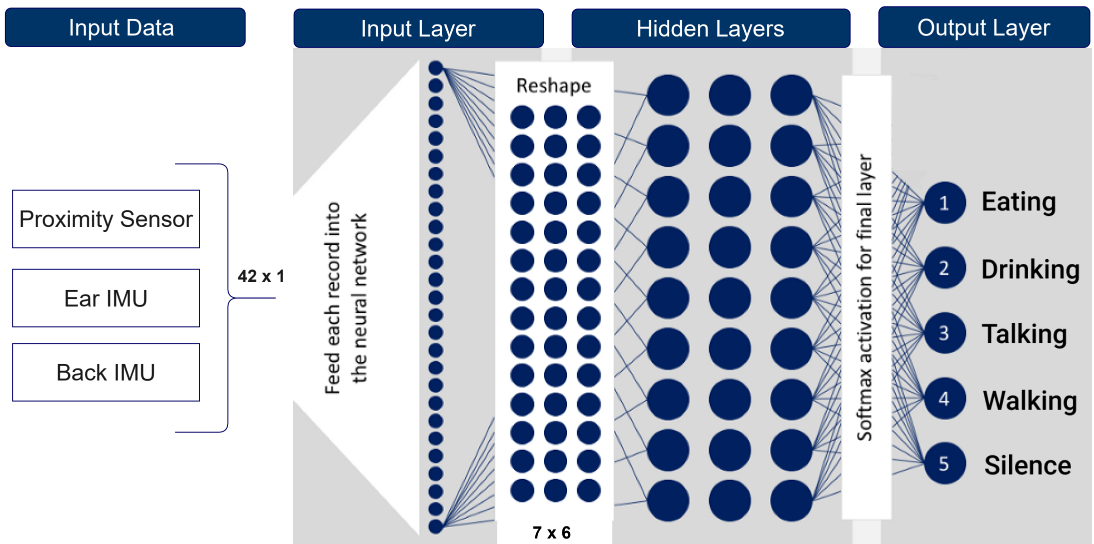

Figure 6.7: DNN Model architecture

In addition to accuracy, we also explored other ML frameworks such that we can run the predictions directly on the wearable or on the phone. We trained several classical models such as SVM, GBT and other models to recognize eating and drinking events and distinguishes them from other everyday activities such as movement, talking, and no-activity. Based on this evaluation, we found that the accuracy of the optimized random forest achieved and accuracy of 72% while the our customized DNN model we achieved an accuracy of 80% and hence, we decide to go with the DNN model. Our custom built DNN model as shown Figure 6.7. The input layer is a vector with 42 elements which is a flattened representation of the six dimensions of data. Inorder to feed the data into our neural network we shape it in such a way that each person has multiple two dimensional records which holds the data for each of the sensors from the FitByte wearable device. Each record is also associated with one label and this is fed to the neural network during the training process.

6.4 Results

In this section, we will present our findings from the data we collected, which includes the daily surveys, interviews, and App usage data. As we gave participants two days to get acquainted with the journaling method, in our analysis we only included data from the last 7 days of each phase. We also present results from 12 participants as one participant ($P2$) dropped out after a few days because they didn't feel the wearable prototype was comfortable.

We categorized our results under three main topics: adherence; utility and usability; and social acceptability and privacy concerns. In each topic, we present a summary of our findings from the qualitative and quantitative data analysis.

6.4.1 Adherence

The main challenge with food journaling is the low adherence rate. ADM systems are designed to improve the adherence rate by easing the journaling effort and reducing recall errors. Up to our knowledge, there are no published evaluations for the impact of ADM on journaling adherence. In this section, we present our results on how our ADM system FitNibble impacted the overall adherence by comparing results from phase 1 (without FitByte) and phase 2 (with FitByte).

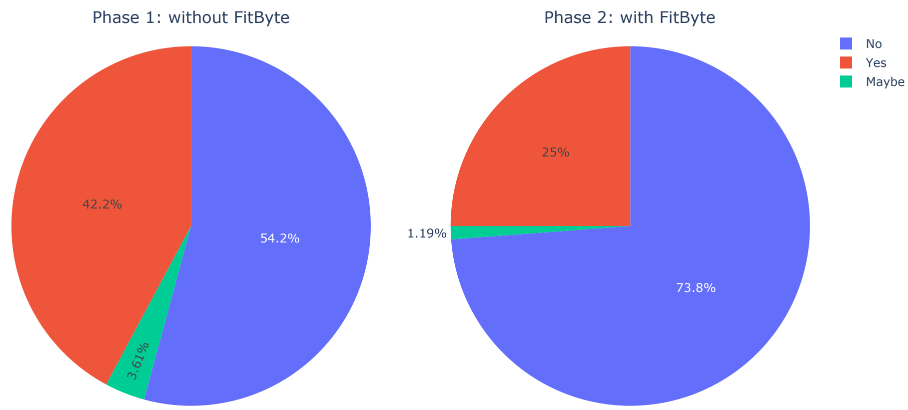

Figure 6.8: Percentage of days with missed logs before and after participants starts wearing FitNibble

Figure 6.8 presents the percentages of days with missed logs as reported by participants in their daily survey. As shown in the pie charts, participants in Phase 2 were less likely to miss logs: the percentage of *No* responses (i.e., No missed events today) increased by 19.6%. This is a clear sign of improved adherence while using FitNibble. To statically assess this improvement we ran a *chi-square* test of independence between the two phases and it showed that the improvement in adherence was significant X^2 (2, N =163)=6.1478 $p = 0.013158$.

To understand the reason behind this improved level of adherence, we asked participants in the exit interview why they started to report more *Nos* in the second phase. One assumption that can be made is this improvement was due to the familiarity with the journaling method, as participants have been doing it for more than a week, but the response we got from the interviews was completely different. Most participants attributed this improvement to the wearable. $P10$'s response is representative in this regard: "*I believe it was because the wearable was sending me notifications every time I eat.*". Another metric we used to assess FitNibble contribution to the improved adherence was the number of times the participants used the wearable notifications or the time reminders to do the log $vs.$ directly opening the App. Figure 6.9 illustrates that in the second phase participants relied more on the wearable notification and time reminders than in the first phase (48% increase). When investigating the contribution of wearable notification and time reminders to the total number we found that participants didn't use the reminder feature that much as explained in the Utility and Usability section. These results support the assumption that the wearable notifications played a big role in explaining the improved adherence, as participants were using them nearly fifty percent of the time when they logged eating events.

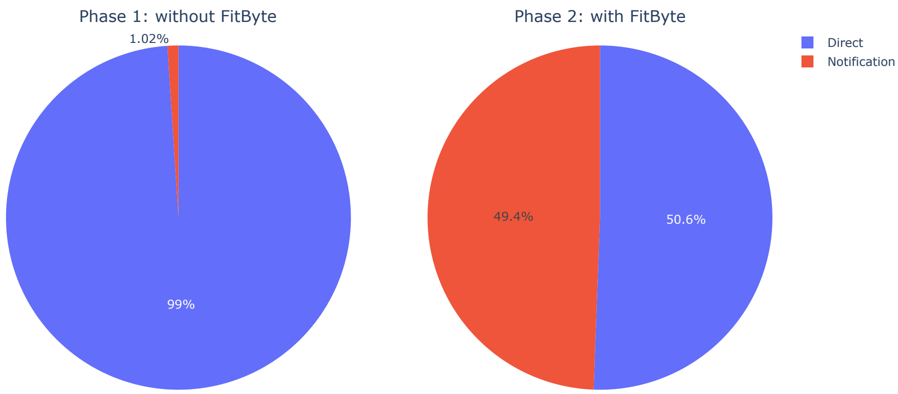

Figure 6.9: Logging methods used before and after participants starts wearing FitNibble. Direct, indicate self initiated logs, and Notification, refer to logs initiated after receiving a notification from the FitByte App, including time reminders and wearable notifications.

After analyzing the interviews' data we found the following emerging themes related to adherence.

Forgetting to Log

Among all the reasons participants reported in phase 1 for missing to log *Forgetting to log* was the most cited. This coincides with the findings of Cordeiro *et al.* [21] in manual food journaling studies. $P8$ mentioned this quote "*Most of the time I forget, and there's some time. I'm just too busy. I don't have time to record snacks because I just grab a banana and go out.*", and $P9$ said "*Ah, I think it's like unconsciousness. I just forgot that I need to journal while I am eating*". Participants also cited other reasons like being *busy or distracted*. For example, $P8$ also said "*Most of the time I forget, and there's some time. I'm just too busy. I don't have time to record snacks because I just grab a banana and go out.*". Others have cited *changes in routine* to be the reason for them missing. "I visited my girl in college for a few days, so my routine has changed and I missed a couple of events because of that " ($P1$). In phase 2, participants' adherence improved and the reasons they cited for missing events were different. One common reason was the *The wearable didn't notify me* For example, P1 mentioned "*"Uh, I missed. I didn't do it at lunch and I didn't do it because for some reason it did not detect."*" and $P13$ "*like one day because I was wearing my contact lens and then I just have back to back meetings so I didn't have the chance to put my glasses on till like late afternoon and I missed to log my lunch*"

Missing snacks

Another common theme from phase 1 interviews is that most participants realized that they are missing to journal small meals such as snacks more than main meals. $P3$ mentioned "*when I have a little snack that's like really easy for me to miss 'cause I won't be thinking about it.*" and $P12$ said "*Yeah, I think I've missed pretty much all the snacks.*". In phase 2 we saw the complete opposite, Participants started to be aware of their snacking habits. For example $P4$ mentioned in one of the daily surveys "*Today the device recognized that I was eating a few almonds. This was a snack that I didn't plan or realize that I was eating, it was somewhat automatic behavior after visiting the kitchen. I wouldn't log that normally, but it was nice that it could catch it.*" and also in their daily survey $P11$ mentioned the following memorable experiences "*The device reminded me to log both snacks when I didn't even think about it*", and "*I think for snack I am relying on the glass now.*".

Looking at the daily survey data, we notice an increase in the number of reported snacks in phase 2, but the number of reported meals was almost the same between the two phases (6.1). To evaluate the difference between journaling with and without FitNibble wearable, we ran a repeated measure ANOVA test for both meals and snacks for both phases and the differences were not significant ($F_meal(2, 83) = 0.39, p_meal = 0.844, F_snacks(2, 83) = 0.39, p_snacks = 0.99$).

Depending on the wearable's notifications

In the second phase, the participants depended on FitNibble in the journaling process. Keep in mind, we explicitly instructed participants not to rely on the device and to continue to log events when they remember, but most participants soon after they started using the wearable didn't follow our instructions. When asked how much they depended on the wearable in this phase $P13$ answered "*I think like 90% of the time. I do the logs after it notifies me*", and $P8$ said "Yeah, in general, I feel like it frees me from keeping paying attention to whether I'm eating or not. For the journal, I can just rely on it. I fully rely on this device. So if the notification is not on time, I just miss it." These quotes support the hypothesis that the improved adherence in the second phase was mainly due to the wearable and the notifications it sends and not any other factors.

Finally, we referred to the self efficacy score of participants, which we evaluated at the beginning and the end of the study. We noticed a slight increase on the average scores (beginning: 3.7, end: 3.85), but we didn't find any correlation between these scores and the particpants level of adherence.

6.4.2 Utility and Usability

In this section, we discuss the usability of all the features introduced in the FitByte App/Foodility App and the FitNibble wearable. We will also discuss the perceived utility of the setup and highlight some of the emerging themes from the data.

Time reminders has low value

One of the features we introduced in the FitByte App is allowing users to set logging reminders at a specific time if they know they are most likely to eat at a certain time. This feature was available in both phases of the study, but the general feedback we received was that participants didn't find it useful most of the time. This is evident in the daily survey responses when asked participants to state how often they used the reminder feature that day. From the data we found that this feature utility rating is trending low in both phases 6.1. We noticed that more participants found the feature to be useful in phase 2, but when investigating further the interview data showed that the participants were confusing the use of the wearable notifications and the time set reminders. We ran a repeated measures ANOVA to evaluate the difference between the two phases and the results were significant($F(2, 83) = 8.324, p = 0.005$).

When asked about the reasons behind the low utility of the feature, Some participants in the first phase mentioned that they would like to set the reminder once and have it repeat every day. We implemented this change to the feature but the utility didn't improve by much. Participants explained that usually, they don't have a fixed schedule, which makes it difficult to plan when they will have a meal or a snack. For example, $P13$ mentioned ""*It might be useful, but not to people who are students because students they have different schedules every day, so we don't have a fixed time for eating.*". $P7$ said "*I didn't use the reminders because my meal time is not fixed*".

Positive experience with the wearable

When evaluating the user experience with the FitNibble wearable, most participants said it improved their experience and attributed that to the smooth experience the wearable provides to do the logs. $P1$ said "*Much better than the first. I like that I didn't have to remember to log. The device prompted me with the notification and then it automatically opens that app so that I could easily log, that was very, very special*", and $P11$ said "*I do journal more now. it definitely reminds me most times so I don't miss*". We can also realize this from the daily surveys that the journaling difficulty has dropped in the second phase 6.1. Figure 6.10 shows the trends between the two phases. We ran a repeated measures ANOVA to evaluate the difference between the two phases and the results showed that the change in using remainder functionality was significant ($F(2, 83) = 5.524, p = 0.021$).

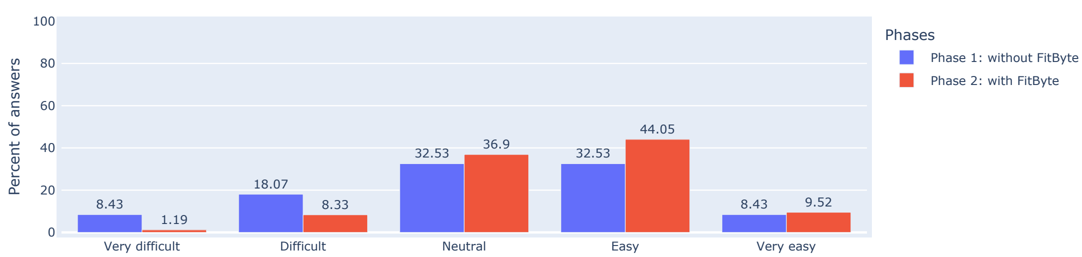

Figure 6.10: Journaling difficulty before and after participants start wearing FitNibble

In the second phase, we asked participants to report in the daily survey how often the wearable notifications helped them today. The average rating for this feature was 3.3$\pm$ 0.8 (above midpoint)6.1.

All these results point to the positive experience with the wearable, but a few of the participants didn't view the experience as positive as others. We noticed that the common attribute for participants in that group is they are very punctual at journaling even before using the wearable. For these 4 participants, the wearable wasn't helping them because it sends notifications after the user starts eating, and they are used to doing the log before they eat, so the wearable notifications bother them because it comes after they have already made the logs. $P4$ said "*I wouldn't like to keep using it, because I do all the work and it's not giving me back too much, because I have to remember to log before the meal, and I'm good at it.*", and $P6$ said "*I don't think it ever reminded me in a way that I would have forgotten. It was mostly just background noise.*" . This feedback highlights that ADM journaling provides value to users who are forgetful and less punctual, but for users who don't have these issues, journaling with ADM negatively affects their experience.

Variable perception of accuracy

When it came to how participants perceived the accuracy of the wearable notifications there was a split between good and bad responses. To understand the reasons behind this split we looked at the app usage data as we keep track of how many times participants responded to notification with *Yes* (true positives) and how many times they responded with *No* (False positives). After analyzing the data we found that half of the participants were receiving a few false positives per day ($Avr = 2.5 \pm 2$), and the remaining half of the participants were receiving many false positives per day ($Avr = 15 \pm 5$), Figure 6.11 shows an example of data logs from each group. Our understanding this is probably caused by variation in sensor noise across devices. We built 11 FitNibble wearable prototypes for this study. By visualizing the sensor streams for some of these prototypes we found significant variation in the noise level, especially in the MPU9250 gyroscope data.

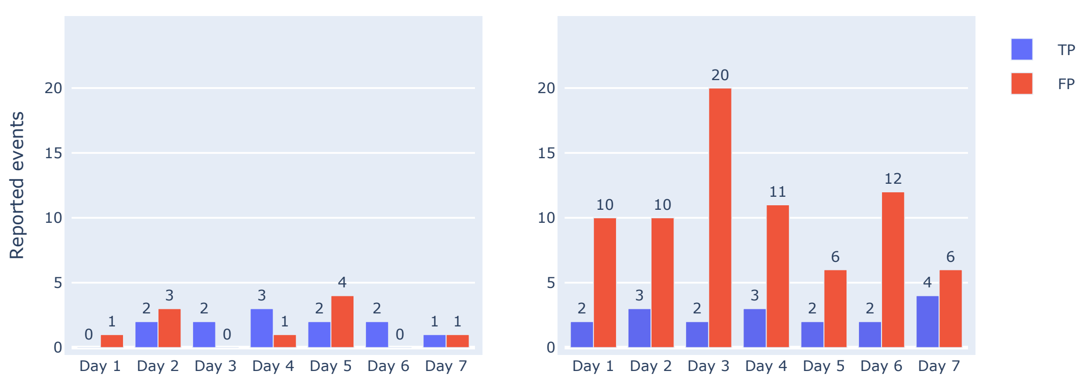

Figure 6.11: Number of true positives (TP) and false positives (FP) received across 7 days. Figure on the left is from a participant with high accuracy whereas the figure on the right is from a participant with low accuracy.

Despite the variation in accuracy across devices 61.9% of the daily response found the accuracy to be average (34.5%) or above average (27.4%) , while 38% of the responses found it to be below average. Figure 6.12 shows the distribution of the daily perceived accuracy for all

Metric	Phase 1	Phase 2
Number of reported meals	$Avr = 2.2 \pm 1.01\, meals/day$	$Avr = 2.2 \pm 0.74\, meals/day$
Number of reported snacks	$Avr = 0.54 \pm 0.83\, snacks/day$	$Avr = 0.77 \pm 0.92\, snacks/day$
Journaling difficulty	$Avr = 3.1 \pm 0.9\, points$	$Avr = 3.5 \pm 0.7\, points$
Time reminder's utility	$Avr = 1.5 \pm 0.5\, points$	$Avr = 1.8 \pm 0.3\, points$
Wearable notification utility	—	$Avr = 3.3 \pm 0.8\, points$
Perceived accuracy rating	—	$Avr = 2.8 \pm 0.6\, points$

Table 6.1: Summary of the daily survey results (all rating questions has a 1-to-5 likert scale)

participants. These results indicate that most participants found the wearable to be reasonably accurate despite the high false positive rate.

Since it's difficult to monitor false negatives in a long-term field study we relied on participants' reports in the daily survey and the interviews. Most participants didn't report false negatives and they think the system recall was high. For the few reported false negative incidents, participants indicated that there was a connection or a fitting problem most of the time. $P13$ said "*one thing I notice for the false negatives it's probably just because I'm not wearing my glasses properly.*"

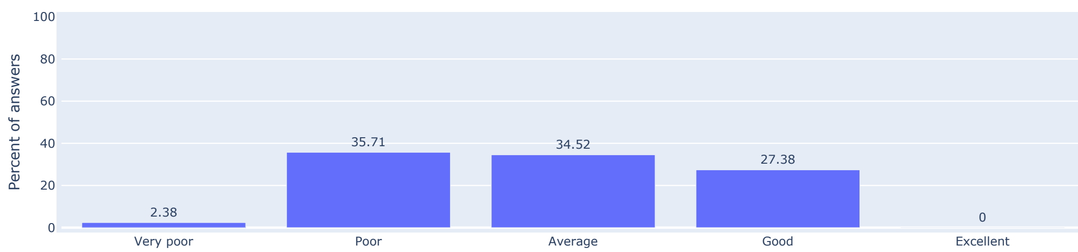

Figure 6.12: Percentage of user perceived accuracy of FitNibble

Increased awareness of dietary patterns

One major utility theme in the study was increased awareness. Participants have reported in both phases of the study that they are becoming more aware of their diet after they started journaling. For example, $P3$ said "*I am more mindful and aware of what I was eating, and I guess a little bit more so with this (the wearable) because I wouldn't really think about snacks until this thing would notify me.*" and $P5$ said "*I'm trying out a variety of food that I wouldn't really think about earlier. I think that's also attributed to the food journaling activity and to the device.*". Many participants also indicated that they learned something new about their diet. For example, $P12$ said "*I realized that I eat irregularly at night which will range from 5:00 PM to 9:00 PM*", and $P3$ said "*I definitely don't eat as much as I should. Uhm, at least during this point in the summer*". We also saw some trends of behavioral change due to increased awareness. $P5$ said " *It makes me more conscious about what I eat throughout the day, and when I want to snack, I think that I'll have to keep track of it in my food journal, and then I see myself not really following my diet so I don't snack.*"".

Desires for a finished product

During the exit interview, we asked participants "*What would it take for you to use this device in your daily life?*" The aim behind this question was to understand what barriers can prevent ADM to be widely adopted. Most participants said they would use FitNibble if the prototype was improved to a finished product quality. Some suggested changes to the form factor to improve comfort. For example, $P1$ said "*If this device didn't have this cable coming out of it. If I could just slip it into my glasses, or there's just a little clip here, so it's easy if I wanted to take it off.*", and $P3$ said " *it's not waterproof. So I remember that it was raining one day and I came to campus. So I had to wait for the resin to stop so I could go home*". Participants also suggested changes to the "Foodility" journaling App, such as reducing the number of clicks required to do the logs, and adding a weekly view to help users capture trends that happen across multiple days. One other desired feature was the integration with other health-related journals like fitness tracking, glucose level, calorie counting, and mood logs.

6.4.3 Social Acceptability and Privacy Concerns

In this topic, we mainly relied on the interview and the daily survey data. The main emerging themes from the data are summarized below.

Social acceptance

The major theme under this topic was the wide acceptance of the FitByte app and wearable. Most participants didn't perceive any discomfort in the social setting as people around them were either indifferent about the setup or they thought it was cool. $P1$ mentioned " *Really cool, yeah. They all think it's really cool!*", and $P9$ said " *Yes, I did it in front of my friends and they feel normal about it. I just told them I was in a research study*". Not all participants had the same experience. $P11$ mentioned that doing the food journaling added some social pressure on his girlfriend because she wasn't paying attention to her diet " *When I did this food journaling with my girlfriend, I became more careful not to create any social pressure*". $P6$ found it awkward to pull their phone every time they eat in a social setting " *It was just a little bit too weird to have someone says "hey you want some fries", and for me to say "OK but I'm gonna take a photo of them first"*"

In the second phase, most participants found the current FitByte design makes it invisible to others. For example, $P7$ said " *Most of the time I think people don't even notice*" and $P5$ said " *No one really looks at you and hiding the wire makes it even less noticeable. If it is more noticeable or larger, it will make a total change*".

Social collaboration

This is one of the interesting themes we found in the data. Some participants mentioned that in the first phase they depended on their friends and family to prompt them to log meals. $P3$ and $P1$ talked about that "*I told some of my friends about it, and some of them actually reminded me a few times.*", "*I would say this maybe twice a day to whomever I was with. " I have to remember to do this, I have to remember"*". After using the wearable in the second phase there was no mention of social collaboration. $P3$ mentioned that he started to rely more on the wearable "Usually, my friends prompt me because when I'm with them it's when I'm mostly distracted, but in this phase they didn't, because this thing tells me to do it anyways".

Have full control over the wearable camera

As we mentioned in section 6.3.2 we were considering including an on-board camera in the FitNibble design to capture images of the food when eating is detected. In the exit interview, we talked to participants about this feature and asked if they would consider using it or not. The majority of participants found the feature to be useful as it minimizes the effort to do the log. One good example is $P7$ who refused to wear the device because he thought it had an on-board camera and demanded it to be removed from the wearable. After using FitNibble in phase 2 his opinion was changed "*It would be good to have if it can be used in a more controlled way, like only when I'm eating*", "*I see why it would be useful. I don't want to take my phone from my pocket when my hands have food on them*". Most participants agreed to use the camera if they have full control over this feature. For example, $P13$ said "*If the camera is there we need to be very cautious about it. I would prefer not to use it unless I know where it saves the data and when it's on. I don't want it to accidentally trigger in the bathroom*".

6.5 Discussion

In this evaluation, we tried to assess the value an ADM system can provide to the user. While ADM research aims to automate different parts of the food journaling task like detecting eating events, identifying the food type, and estimating the amount; in this evaluation, we focus on detecting *"When"* a user starts to eat to allow for just-in-time interventions. In our use case, we targeted sending notifications to the user as soon as they started eating to prompt them to do their logs. The goal was to help reduce the number of missed events due to recall errors, which has a great impact on adherence to food journaling[21]. To tackle this challenge, we developed an end-to-end system, including a custom wearable (Fitbyte 2.0), an interface App, and a server backend. We then conducted a field deployment to compare traditional self-report journaling and journaling with ADM. The results of our evaluation highlighted the potential of ADM in increasing adherence to food journaling and improving the user experience with the process.

6.5.1 Main outcomes

Our field deployment allowed us to assess the influence of ADM on food journaling adherence. Our results showed around a 20% drop in days with missed events, and we saw a significant improvement in adherence with $p = .013158$. The main reason participants cited for this improved adherence was the reduced cognitive load on the user after using the wearable ADM. For example, $P5$ mentioned "*Compared to the first phase, I don't have to think about it*". After using Fitbyte 2.0 participants became more aware of their dietary behavior especially when it came to snacking as most of them missed small eating events.

While the rate for false positives was high for some participants and low for others, the overall perception of accuracy was leaning towards the positive side (Figure 6.12). Most participants said they would use Fitbyte 2.0 if it was redesigned to have a more compact form factor and finished product features (e.g. waterproof, no cables, and lightweight). This feedback indicates that there are no major barriers to adopting wearable ADM products. Many participants felt that doing food journaling in public is socially acceptable and even some participants relied on their friends and family to prompt them to log. This is a sign that there is less stigma associated with photo-based journaling. Our participants believe this can be attributed to the wide adoption of this type of journaling in social media platforms like Instagram and Snapchat.

Finally, in our evaluation, we investigated the use of a wearable camera to help users take photos of their food. While all participants felt this feature would raise many privacy concerns, after using Fitbyte 2.0, most of them saw the value of using it to reduce the journaling effort, $P5$ said "*You still have to take the picture yourself with your phone, it's not really cutting that part cuz it's not taking a picture for me*", but the participants demanded full control over its activation. So when the device detects they are eating it should ask for permission to turn on the camera and take the photo.

6.5.2 Design Recommendations

In this section, we discuss the lessons we learned from this study and how it can inform the design of the next generation of wearable ADM systems.

Targeted users

In our evaluation, we found a clear difference in responses between participants who are punctual with journaling and those who are not. Participants who regularly missed to log events have benefited the most from the wearable ADM system. On the other hand, participants who didn't suffer from this issue found low value in using it. This group was also more sensitive to false positives and found it to be annoying. Therefore, we recommend designers keep these differences in mind when assessing the utility and usability of their ADM system, and understand the value it will deliver to the targeted end-user.

Acceptable range of error

One other question we tried to answer in this evaluation was *"What is the acceptable number of false positives per day?"*. The feedback we received from participants was very similar despite the discrepancy in the false positives rates they received. Most participants recommend a maximum of 5 false positives per day, and they believe if this number gets close to 10, they would be annoyed. One other thing ADM designers should keep in mind is to clearly explain to the user what counts as a false positive and what doesn't. For instance, in our study, a few participants indicated that they received false positives when they were chewing gum, biting their nails, or drinking. In our model design, we considered all these actions to count as eating events. These participants were not fully aware of that, which in turn influenced their experience with the wearable.

Improving interaction

Through the interviews, we conducted we received many recommendations on how we can improve the interaction with the wearable. For example, some suggested that we use voice recognition to communicate between the App and the user. In this scenario, the user can respond to notifications with voice commands like *"Yes, I'm eating"* or *"No I'm not"* also they can do their logs by recording a short audio message that can be converted to text on the journal. Another proposed feature was *on wearable notifications*, which means the user prefers to get the notifications on a wearable and not on the phone. Participants mentioned when they eat at home, they don't usually have their phone with them so they would miss the notifications. One participant liked that Fitbyte sent notifications on her watch, and another suggested receiving the notifications on the glasses. Several participants also recommended we review the 30-minute snooze notifications rule. One participant mentioned that sometimes she would get a notification 5 minutes before she starts eating and would mark it as a false positive, but when she actually eats she doesn't get a notification because the device was soonzed. Worth noting that this type of false positive was useful to the participant, because it reminded her about journaling a few minutes before she ate, so she still remembered to log. Another participant mentioned that she has long eating events that can extend for more than an hour, and found the repeated notifications every 30 minutes to be annoying. One way to solve this issue is to give users the choice on how long they would prefer to snooze the notifications.

Finally, the experience our participants had with time set reminders showed low value for this feature. Many participants had flexible schedules and setting a recurring reminder didn't help them most of the time. This finding highlights the value of an ADM system such as Fitbyte 2.0, in improving the user experience with food journaling.

Chapter 7

General Discussion

This chapter summarizes the findings of my research and provides a design guideline for future wearable ADM research. I finally conclude my Ph.D thesis with a list of future directions that I and other researchers in the field can pursue.

7.1 Research Findings

In this dissertation, I worked on address two research questions:

- **RQ1:** How can we develop **practical** ADM systems that **replicate** in-the-lab performance **outside** the lab?
- **RQ2:** Can ADM systems improve food journaling **compliance** and reduce **journaling difficulty**?

My work on Earbit and FitByte was directed to address **RQ1**. The solutions I proposed required fundamental changes to the wearable design and the data collection protocol. In the ADM systems I developed, I focused on selecting sensors less prone to environmental noise and fitting problems. These sensors were hosted in commonplace form factors to make them easier to adopt and use in social settings. To address challenges with obtaining ground truth in unconstrained environments, I proposed asking participants to perform a high-level task (e.g. attend dinner party) in semi-controlled environments. In my research, I ran studies in simulated home environments and collected short-term sessions in free-living environments. Both methods provided rich naturalistic datasets annotated with high precession. All these factors collaborated in improving performance in free-living environments. Both setups achieved ¿90% accuracy in detecting eating episodes inside and outside the lab.

Building on these results I developed a new ADM system (FitNibble) and used it in a field deployment to address the second research question **RQ2**. The goal of the study was to evaluate the utility and usability of the ADM system. I considered evaluating factors such as adherence to food journaling, social acceptability, and privacy concerns. Our findings showed that FitNibble has helped participants adhere to food journaling protocol, it also significantly dropped the number of missed events and reduced journaling difficulty. All these findings support my thesis statement that:

Automatic diet monitoring enabled by wearable sensors can improve compliance to food journaling, by lowering the cognitive load required by users, and dropping the number of missed eating episodes.

7.2 Wearable ADM Design Guideline

In this section, I provide a guideline to inform the design of future ADM systems. This guideline is based on my research findings and the latest ADM literature.

7.2.1 Test your sensor in realistic noisy conditions

Many research efforts have evaluated their sensing modalities in controlled lab environments. This type of evaluation doesn't allow researchers to understand the nature of the signals in real-world settings. The signal-to-noise ratio varies between environments, so to pick the right set of sensors you need to anticipate the potential noise sources and test your setup against them. This test can be done in simulated or real-world scenarios. For example, researchers exploring the use of microphones in detecting chewing and swallowing sounds should test their setups in noisy environments like restaurants, train stations, or a football stadium. In [53] and [40], researchers found that environmental noise has a great negative impact on prediction accuracy. Audio pollution can also be simulated in controlled environments by playing sounds of noise captured in different real-world environments.

7.2.2 Pay attention to sensor placement and fitting

When designing a wearable setup, one of the main tasks that have a great influence on performance is selecting the best wearable form factor for the actions your system is trying to capture (i.e. chewing, swallowing, and/or hand-to-mouth gestures). The selected form factor becomes a platform to test different sensor placements to produce the best performance. Finding the best sensor placement requires a deep understanding of the human anatomy and the physical phenomena the ADM is trying to sense.

Fitting has also a great influence on ADM performance. Having a loose fit system can prevent the sensor from capturing body signals and make it more prone to motion artifacts and environmental noise [7]. If the selected form factor doesn't provide the appropriate sensor placement or fitting, then modification has to be made to the design or a new form factor should be selected.

7.2.3 Try capturing multiple eating actions to improve performance

Eating and drinking activities usually involve multiple actions including biting, chewing, swallowing, and hand-to-mouth gestures. Capturing only one of these actions and using it for ADM doesn't provide good coverage for different food intake activities. For example, if your ADM only detects chewing then your system will miss out on detecting drinking events and eating activities that may not involve chewing like eating yogurt, ice cream, and soup. Detecting multiple eating actions allows your ADM to cover a wider range of food intake events and improves its precision (low false positives) and recall (low false negatives). Capturing multiple actions also improves the system onset detection and provides good coverage for the eating episode duration. Capturing multiple actions from a single platform is a hard design problem. Researchers have explored using multiple wearable devices to capture all eating actions but the proposed setup was impractical, socially unacceptable, and difficult to maintain [45, 46]. The FitByte platform represents a good example of an ADM that captures multiple eating actions from a single platform, the wearable also allows users to take photos of their food using an onboard camera. FitByte's eyeglasses form factor enabled the system to tap into different regions of the user's head to capture a variety of eating actions without compromising on practicality and social acceptability [12].

7.2.4 Select a practical and socially acceptable form factor

In this dissertation, I've thoroughly discussed the impact of social acceptability on compliance to ADM. Researchers should aim to select commonplace form factors, which users wear throughout the day to ensure better coverage for the user activities. Form factors can be socially acceptable but not practical. For example, if your setup requires users to wear a backpack, users might find it socially acceptable but impractical to wear all day. Platforms like eyeglasses, wrist bands, belts, and shirts can be good candidates for practical and socially acceptable form factors

7.2.5 Collect data in semi-controlled environment

Building machine learning models that work reliably in real-world settings has been a long-standing challenge for ADM. As has been demonstrated in EarBit [11] and FitByte [12] we found one of the most effective methods to building realistic eating detection models comes from loosening the restrictions on the participant's behavior during data collection. The main objective is to give the user a high-level task (e.g. attend a dinner party, go buy yourself a snack and eat it), and give them the freedom to do it as they please. Dietary activities normally get mixed with other activities like talking and walking, to ensure the machine learning models don't get fed data with wrong labels a granular annotation scheme is required. This level of granularity usually requires having video footage as ground truth for participants' activities.

Requiring video recordings of the session introduces another challenge to the data collection process as participants may behave unnaturally when they know they are being observed. Another challenge is where to place cameras to ensure getting a good view of the user's activities without invading the privacy of others around him. EarBit and FitByte provide good examples for semi-controlled data collection protocols. They both focus on obtaining realistic data with minimum restriction on the participants' behavior or location. To record video footage of the activities I proposed instrumenting the environment with cameras or instrumenting the user with minimally invasive wearable cameras that can be integrated with the ADM platform.

7.2.6 Give users full control over privacy-invasive modalities

In some ADM systems, sensitive data is collected especially with the use of cameras and microphones. Incorporating these modalities in your setup raises privacy concerns to the user and people around them. You can try to limit your sensor use to specific occasions detected by other low-power sensors, but users might not want to receive false positives at sensitive moments (e.g. in the bathroom). The usability evaluation of FitNibble showed that users are willing to use sensors that capture sensitive information in public if it delivers a great value to their user experience and if they have full control over it. One way to implement this protocol is to always ask for permission from the user before turning on the camera or microphone, and ensure they run for a very short duration (a few seconds) to prevent it from invading the privacy of people around them. You can also reduce the risk by using a non-front-facing camera, like in FitByte, or sample the audio information with low frequency and/or compute features on the device to avoid concerns about speech and image reconstruction.

7.3 Future directions

In this dissertation, I've aimed to push the state of the art in the field of automatic diet monitoring. While my work provided an example for how we can overcome challenges with ecological validity, and how we can assess the utility and usability of wearable ADM systems, many ADM challenges remain standing. In this section, I discuss these challenges and propose future research directions.

7.3.1 Data annotation

Similar to any activity recognition system, ADM systems need rich and accurately annotated datasets. This is an essential resource that is difficult to acquire. Collecting data in naturalistic settings requires granular annotations as people tend to mix different activities in an unpredicted way. Annotating data with that level of precision represents the foundation for building machine learning models and evaluating their performance.

In general, this annotation task is very arduous. Researchers are required to spend hours and hours going through the data, checking ground truth, and labeling. While this is a challenge for all machine learning systems, the data collected for ADM is usually complex and sensitive information making it difficult to annotate by crowd workers.

One solution I propose is that ADM researchers can use a well-vetted ADM system in their data collection for ground truth. The ADM literature has good examples of systems that provide eating predictions with reasonably high accuracy. These predictions can be used as a first pass on annotations, which reduces the researchers' role to reviewing the predictor labels, rather than annotating the data from scratch. This method can significantly reduce the time and effort required to perform the labeling task.

7.3.2 Detecting food type and amount

Many food journalers use self-report Apps to help them track how many calories they are gaining or burning, and assess what impact that has on their weight and overall health. An ideal ADM system should help users answer these questions with minimal effort. While the ADM community is making progress in tracking eating moments, detecting food type and the amount is still an open challenge. In my research, I've tried to provide users with enough information to identify and track information about the food they consumed. This was mainly in the form of providing the users with footage of their meals and snacks. This can be an effective method to estimating calorie count given the limited technology we have in hand.

Advances in the computer vision field have the potential to provide a better solution to this problem, but vision-based systems have their own set of challenges when it comes to estimating food type and mount. For example, vision systems can not distinguish between foods that have the same appearance but different nutritional continents like Greek yogurt and fat-free yogurt. Vision systems might also have a difficult time identifying uncommon cuisines especially if they didn't get trained on them. One way to address these challenges is to supplement vision systems with modalities that are sensitive to material contents like spectroscopy systems or surface acoustic waves (SAW) sensors.

7.3.3 Exploring ADM utility

Finally, I would like to encourage researchers in the ADM field to run more long-term field studies and assess the utility and usability of different systems on different populations. For example, targeting senior citizens with ADM can provide their caregivers a better understanding of their dietary behavior and how it influences their health without constantly monitoring them. Another example can be to target athletes and help them measure the influence of dietary activities on their performance. Generally, ADM can also provide better diet tracking tools for medical research which still base their findings on self-reported data.

Bibliography

[1] O. Amft, H. Junker, and G. Troster. Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 160–163, Oct 2005. doi: 10.1109/ISWC.2005.17. 2.2.1

[2] O. Amft*, M. Kusserow, and G. TrÖster. Bite weight prediction from acoustic recognition of chewing. *IEEE Transactions on Biomedical Engineering*, 56(6):1663–1672, June 2009. ISSN 1558-2531. doi: 10.1109/TBME.2009.2015873. 5.3.2, 5.6.3

[3] Edward Archer, Gregory A Hand, and Steven N Blair. Validity of us nutritional surveillance: National health and nutrition examination survey caloric energy intake data, 1971–2010. *PloS one*, 8(10):e76632, 2013. 2.1.1

[4] Ling Bao and Stephen S Intille. Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer, 2004. 4.3.2

[5] Elizabeth Barrett-Connor. Nutrition epidemiology: how do we know what they ate? *The American journal of clinical nutrition*, 54(1):182S–187S, 1991. 2.1.2

[6] Lindsay T Bartholome, Roseann E Peterson, Susan K Raatz, and Nancy C Raymond. A comparison of the accuracy of self-reported intake with measured intake of a laboratory overeating episode in overweight and obese women with and without binge eating disorder. *European journal of nutrition*, 52(1):193–202, 2013. 4.2.2, 4.5.3

[7] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. Stick it in your ear: Building an in-ear jaw movement sensor. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 1333–1338, 2015. 4.1, 7.2.2

[8] Abdelkareem Bedri, Himanshu Sahni, Pavleen Thukral, Thad Starner, David Byrd, Peter Presti, Gabriel Reyes, Maysam Ghovanloo, and Zehua Guo. Toward silent-speech control of consumer wearables. *Computer*, 48(10):54–62, 2015. 4.1

[9] Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner. 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, page 91–92, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335782. doi: 10.1145/2802083.2808411. URL `https://doi.org/10.1145/2802083.2808411`. 2.2.2, 3.1, 4.1, 4.3.1

[10] Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner.

Detecting mastication: A wearable approach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, page 247–250, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450339124. doi: 10.1145/2818346.2820767. URL `https://doi.org/10.1145/2818346.2820767`. 2.2.2, 2.3, 3.1, 4.1, 4.3.1, 4.3.1, 5.2.2

[11] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. Earbit: Using wearable sensors to detect eating episodes in unconstrained environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), September 2017. doi: 10.1145/3130902. URL `https://doi.org/10.1145/3130902`. 1.3, 5.1, 5.2.2, 5.3.1, 5.4.1, 6.1, 6.1.2, 6.2, 6.3.2, 7.2.5

[12] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. Fitbyte: Automatic diet monitoring in unconstrained situations using multimodal sensing on eyeglasses. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020. 1.3, 2.3.1, 5.1, 6.1, 6.1.2, 6.2, 6.3.2, 6.3.4, 7.2.3, 7.2.5

[13] Marianne Bertrand and Diane Whitmore Schanzenbach. Time use and food consumption. *The American Economic Review*, 99(2):170–176, 2009. 4.2.2

[14] Shengjie Bi, Tao Wang, Ellen Davenport, Ronald Peterson, Ryan Halter, Jacob Sorber, and David Kotz. Toward a wearable sensor for eating detection. In *Proceedings of the 2017 Workshop on Wearable Systems and Applications*, WearSys '17, page 17–22, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349598. doi: 10.1145/3089351.3089355. URL `https://doi.org/10.1145/3089351.3089355`. 2.2.2

[15] Lora E Burke, Jing Wang, and Mary Ann Sevick. Self-monitoring in weight loss: a systematic review of the literature. *Journal of the American Dietetic Association*, 111(1):92–102, 2011. 1.1

[16] Sheelagh Carpendale. Evaluating information visualizations. In *Information Visualization*, pages 19–45. Springer, 2008. 4.2

[17] Jei-Fuu Chen, Ling-Ling Wu, and Seng-Cho Chou. Effects of journaling dietary intake app on the health outcomes of chronic kidney disease stage 3b-5. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 3379–3388. IEEE, 2016. 2.1

[18] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. Detecting eating episodes by tracking jawbone movements with a non-contact wearable sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1), March 2018. doi: 10.1145/3191736. URL `https://doi.org/10.1145/3191736`. 2.2.2, 3.1, 5.2.2, 5.4.1

[19] Jungman Chung, Jungmin Chung, Wonjun Oh, Yongkyu Yoo, Won Gu Lee, and Hyunwoo Bang. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. *Scientific reports*, 7:41690, 2017. doi: https://doi.org/10.1038/srep41690. 2.2.2

[20] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages

3207–3216, 2015. 1.2, 1.2, 2.1.2, 2.1.3, 6.1.1, 6.2.1

[21] Felicia Cordeiro, Daniel A Epstein, Edison Thomaz, Elizabeth Bales, Arvind K Jagannathan, Gregory D Abowd, and James Fogarty. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1159–1162, 2015. 1.2, 2.1.2, 6.1.1, 6.4.1, 6.5

[22] Mary Ruth Craig, Alan R Kristal, Carrie L Cheney, and Ann L Shattuck. The prevalence and impact of 'atypical'days in 4-day food records. *Journal of the American Dietetic Association*, 100(4):421–427, 2000. 2.1.2

[23] Karen B DeSalvo. Public health 3.0: applying the 2015–2020 dietary guidelines for americans. *Public health reports*, 131(4):518–521, 2016. 1.1

[24] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover. Detecting periods of eating during free-living by tracking wrist motion. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1253–1260, July 2014. ISSN 2168-2208. doi: 10.1109/JBHI.2013.2282471. 2.2.1

[25] Zoe Falomir, María Arregui, Francisco Madueño, Dolores Corella, and Oscar Coltell. Automation of food questionnaires in medical studies: a state-of-the-art review and future prospects. *Computers in biology and medicine*, 42(10):964–974, 2012. 2.1.1

[26] M. Farooq and E. Sazonov. 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):1495–1503, Nov 2017. ISSN 2168-2208. doi: 10.1109/JBHI.2016.2640142. 2.2.2

[27] J. M. Fontana, M. Farooq, and E. Sazonov. Automatic ingestion monitor: A novel wearable device for monitoring of ingestive behavior. *IEEE Transactions on Biomedical Engineering*, 61(6):1772–1779, June 2014. ISSN 0018-9294. doi: 10.1109/TBME.2014.2306773. 3.1, 4.3.1

[28] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D. Hollan. Chronoviz: A system for supporting navigation of time-coded data. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 299–304, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0268-5. doi: 10.1145/1979742.1979706. URL `http://doi.acm.org/10.1145/1979742.1979706`. 4.2.3

[29] Yang Gao, Ning Zhang, Honghao Wang, Xiang Ding, Xu Ye, Guanling Chen, and Yu Cao. ihear food: eating detection using commodity bluetooth headsets. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 163–172. IEEE, 2016. 3.1

[30] William D Heizer, Susannah Southern, and Susan McGovern. The role of diet in symptoms of irritable bowel syndrome in adults: a narrative review. *Journal of the American Dietetic Association*, 109(7):1204–1214, 2009. 2.1

[31] Elina Helander, Kirsikka Kaipainen, Ilkka Korhonen, and Brian Wansink. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. *Journal of medical Internet research*, 16(4):e109, 2014. 1.1, 1.2, 2.1.2

[32] Jack F Hollis, Christina M Gullion, Victor J Stevens, Phillip J Brantley, Lawrence J Appel, Jamy D Ard, Catherine M Champagne, Arlene Dalcin, Thomas P Erlinger, Kristine Funk, et al. Weight loss during the intensive intervention phase of the weight-loss maintenance trial. *American journal of preventive medicine*, 35(2):118–126, 2008. 2.1

[33] Parvez Hossain, Bisher Kawar, and Meguid El Nahas. Obesity and diabetes in the developing world—a growing challenge. *New England journal of medicine*, 356(3):213–215, 2007. 1.1

[34] Frederick H Kanfer. Self-monitoring: Methodological limitations and clinical applications. 1970. 2.1

[35] Nicky Kern, Bernt Schiele, and Albrecht Schmidt. Multi-sensor activity context detection for wearable computing. In *European Symposium on Ambient Intelligence*, pages 220–232. Springer, 2003. 4.3.2

[36] A. M. Khan, Y. K. Lee, S. Y. Lee, and T. S. Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1166–1172, Sept 2010. ISSN 1089-7771. doi: 10.1109/TITB.2010.2051955. 4.3.2

[37] Julie A. Kientz, Shwetak N. Patel, Brian Jones, Ed Price, Elizabeth D. Mynatt, and Gregory D. Abowd. The georgia tech aware home. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3675–3680, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-012-8. doi: 10.1145/1358628.1358911. URL `http://doi.acm.org/10.1145/1358628.1358911`. 4.2.1

[38] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons, 2010. 4.2.3

[39] B. Lin and A. Hoover. A comparison of finger and wrist motion tracking to detect bites during food consumption. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–4, 2019. doi: 10.1109/BSN.2019.8771066. 2.2.1

[40] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G. Yang. An intelligent food-intake monitoring system using wearable sensors. In *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks*, pages 154–160, May 2012. doi: 10.1109/BSN.2012.11. 2.3, 3.1, 7.2.1

[41] Lena Mamykina, Elizabeth D Mynatt, and David R Kaufman. Investigating health management practices of individuals with diabetes. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 927–936, 2006. 2.1

[42] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. Mahi: investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 477–486, 2008. 2.1

[43] S. Matos, S. S. Birring, I. D. Pavord, and H. Evans. Detection of cough signals in continuous audio recordings using hidden markov models. *IEEE Transactions on Biomedical*

Engineering, 53(6):1078–1083, June 2006. ISSN 0018-9294. doi: 10.1109/TBME.2006.873548. 4.3.1

[44] E Mcgrath. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction: Toward the Year 2000 (2nd ed*. Citeseer, 1995. 4.2

[45] Mark Mirtchouk, Christopher Merck, and Samantha Kleinberg. 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, page 451–462, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450344616. doi: 10.1145/2971648.2971677. URL `https://doi.org/10.1145/2971648.2971677`. 5.1, 5.3.2, 5.6.3, 7.2.3

[46] Mark Mirtchouk, Drew Lustig, Alexandra Smith, Ivan Ching, Min Zheng, and Samantha Kleinberg. Recognizing eating from body-worn sensors: Combining free-living and laboratory data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), September 2017. doi: 10.1145/3131894. URL `https://doi.org/10.1145/3131894`. 5.1, 7.2.3

[47] Ali H Mokdad, Katherine Ballestros, Michelle Echko, Scott Glenn, Helen E Olsen, Erin Mullany, Alex Lee, Abdur Rahman Khan, Alireza Ahmadi, Alize J Ferrari, et al. The state of us health, 1990-2016: burden of diseases, injuries, and risk factors among us states. *Jama*, 319(14):1444–1472, 2018. 1.1

[48] Dan Morris, T Scott Saponas, Andrew Guillory, and Ilya Kelner. Recofit: using a wearable sensor to find, recognize, and count repetitive exercises. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 3225–3234. ACM, 2014. 4.3.2

[49] Mehrab Bin Morshed, Samruddhi Shreeram Kulkarni, Richard Li, Koustuv Saha, Leah Galante Roper, Lama Nachman, Hong Lu, Lucia Mirabella, Sanjeev Srivastava, Munmun De Choudhury, et al. A real-time eating detection system for capturing eating moments and triggering ecological momentary assessments to obtain further context: System development and validation study. *JMIR mHealth and uHealth*, 8(12):e20625, 2020. 2.3.1

[50] Le T Nguyen, Ming Zeng, Patrick Tague, and Joy Zhang. I did not smoke 100 cigarettes today! avoiding false positives in real-world activity recognition. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1053–1063, 2015. 3.1

[51] T. Olubanjo and M. Ghovanloo. Real-time swallowing detection based on tracheal acoustics. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4384–4388, May 2014. doi: 10.1109/ICASSP.2014.6854430. 2.2.3, 2.3, 3.1

[52] T. Olubanjo and M. Ghovanloo. Tracheal activity recognition based on acoustic signals. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1436–1439, Aug 2014. doi: 10.1109/EMBC.2014.6943870. 4.3.1

[53] Temiloluwa Olubanjo and Maysam Ghovanloo. Real-time swallowing detection based on

tracheal acoustics. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4384–4388. IEEE, 2014. 7.2.1

[54] Temiloluwa O Olubanjo. *Towards automatic food intake monitoring using wearable sensor-based systems*. PhD thesis, Georgia Institute of Technology, 2016. 5.2.2

[55] T. Prioleau, E. Moore II, and M. Ghovanloo. Unobtrusive and wearable systems for automatic dietary monitoring. *IEEE Transactions on Biomedical Engineering*, 64(9):2075–2089, Sep. 2017. ISSN 0018-9294. doi: 10.1109/TBME.2016.2631246. 5.1

[56] Temiloluwa Prioleau, Elliot Moore II, and Maysam Ghovanloo. Unobtrusive and wearable systems for automatic dietary monitoring. *IEEE Transactions on Biomedical Engineering*, 64(9):2075–2089, 2017. 1.2.1, 2.2, 6.1

[57] Pavel Pudil, Jana Novovičová, and Josef Kittler. Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125, 1994. 4.3.2

[58] S. A. Rahman, C. Merck, Yuxiao Huang, and S. Kleinberg. Unintrusive eating recognition using google glass. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 108–111, May 2015. doi: 10.4108/icst.pervasivehealth.2015.259044. 2.2.2, 4.3.1

[59] Tauhidur Rahman, Alexander T. Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury. 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, page 2–13, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327930. doi: 10.1145/2594368.2594386. URL `https://doi.org/10.1145/2594368.2594386`. 2.2.3

[60] Seyedmostafa Safavi and Zarina Shukur. Improving google glass security and privacy by changing the physical and software structure. *Life Science Journal*, 11(5):109–117, 2014. 5.7.4

[61] Himanshu Sahni, Abdelkareem Bedri, Gabriel Reyes, Pavleen Thukral, Zehua Guo, Thad Starner, and Maysam Ghovanloo. The tongue and ear interface: a wearable system for silent speech recognition. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, pages 47–54, 2014. 4.1

[62] Charles A Scherbaum, Yochi Cohen-Charash, and Michael J Kern. Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and psychological measurement*, 66(6):1047–1063, 2006. 6.2.2

[63] Giovanni Schiboni and Oliver Amft. *Automatic Dietary Monitoring Using Wearable Accessories*, pages 369–412. Springer International Publishing, Cham, 2018. ISBN 978-3-319-69362-0. doi: 10.1007/978-3-319-69362-0_13. URL `https://doi.org/10.1007/978-3-319-69362-0_13`. 1.2.1, 2.2, 2.3.1, 3, 5.1, 6.1

[64] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Balan, and Youngki Lee. 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)*, pages 1–6. IEEE, 2018. 2.2.1

[65] Scott H Sicherer and Hugh A Sampson. Food allergy. *Journal of allergy and clinical immunology*, 125(2):S116–S125, 2010. 2.1

[66] Katie A Siek, Kay H Connelly, Yvonne Rogers, Paul Rohwer, Desiree Lambert, and Janet L Welch. When do we eat? an evaluation of food items input into an electronic food monitoring application. In *2006 Pervasive Health Conference and Workshops*, pages 1–10. IEEE, 2006. 2.1

[67] Edison Thomaz, Irfan Essa, and Gregory 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*, pages 1029–1040. ACM, 2015. 2.2.1, 2.3, 2.3.1, 3.1, 5.2.2

[68] Edison Thomaz, Abdelkareem Bedri, Temiloluwa Prioleau, Irfan Essa, and Gregory D. Abowd. Exploring symmetric and asymmetric bimanual eating detection with inertial sensors on the wrist. In *Proceedings of the 1st Workshop on Digital Biomarkers*, DigitalBiomarkers '17, page 21–26, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349635. doi: 10.1145/3089341.3089345. URL `https://doi.org/10.1145/3089341.3089345`. 2.2.1

[69] Gabrielle M Turner-McGrievy, Alycia Boutté, Anthony Crimarco, Sara Wilcox, Brent E Hutto, Adam Hoover, and Eric R Muth. Byte by bite: Use of a mobile bite counter and weekly behavioral challenges to promote weight loss. *Smart Health*, 3:20–26, 2017. 2.3.1

[70] Michael S. Wagner. Google glass: A preemptive look at privacy concerns student note. *Journal on Telecommunications High Technology Law*, 11:477, 2013. URL `https://heinonline.org/HOL/P?h=hein.journals/jtelhtel11&i=505`. 5.7.4

[71] Mary H Wilde and Suzanne Garvin. A concept analysis of self-monitoring. *Journal of Advanced Nursing*, 57(3):339–350, 2007. 2.1

[72] Walter Willett. *Nutritional epidemiology*. Oxford university press, 2012. 2.1.1

[73] Michal Yackobovitch-Gavan, Nessia Nagelberg, Moshe Phillip, Liat Ashkenazi-Hoffnung, Eli Hershkovitz, and Shlomit Shalitin. The influence of diet and/or exercise and parental compliance on health-related quality of life in obese children. *Nutrition Research*, 29(6): 397–404, 2009. 1.1

[74] Koji Yatani and Khai N Truong. Bodyscope: a wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 341–350. ACM, 2012. 4.3.1

[75] R. Zhang and O. Amft. Monitoring chewing and eating in free-living using smart eyeglasses. *IEEE Journal of Biomedical and Health Informatics*, 22(1):23–32, Jan 2018. ISSN 2168-2194. doi: 10.1109/JBHI.2017.2698523. 2.2.2, 2.3.1, 6.2

[76] R. Zhang and O. Amft. Free-living eating event spotting using emg-monitoring eyeglasses. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 128–132, March 2018. doi: 10.1109/BHI.2018.8333386. 2.2.2, 2.3.1, 6.2

Appendix A

Daily Survey Phase 1

Please review your journaling app and tell us about it. Please provide a numerical answer if the question ask you to do so.

* Required

1. Email *

2. How many meals have you logged today? *

3. How many snacks have you logged today? *

4. Did you miss logging any events today? *

   Mark only one oval.

   - Yes
   - No
   - Maybe

5. If you think you've missed an event, what was the reason?

6. Did the reminders you received on the App help you with logging events today? *

Mark only one oval per row.

	Never	Rarely	Sometimes	Very Often	All the time
Answer	⬭	⬭	⬭	⬭	⬭

7. How difficult was it to maintain your food journal today? *

Mark only one oval per row.

	Very difficult	Difficult	Neutral	Easy	Very Easy
Answer	⬭	⬭	⬭	⬭	⬭

8. Can you tell us one thing you noticed about your dietary habits today? *

9. Can you share with us one memorable experience you had with food journaling today? *

Appendix B

Daily Survey Phase 2

Please review your journaling app and tell us about it. Please provide a numerical answer if the question ask you to do so.

* Required

1. Email *

____

2. Please charge your device and confirm by checking the box below. *

Check all that apply.

☐ I confirm

3. How many meals have you logged today? *

____

4. How many snacks have you logged today? *

____

5. Did you miss logging any events today? *

Mark only one oval.

◯ Yes

◯ No

◯ Maybe

6. If you think you've missed an event, what was the reason?

____

7. Did the reminders you set on the App help you with logging events today? *

Mark only one oval per row.

	Never	Rarely	Sometimes	Very Often	All the time
Answer	◯	◯	◯	◯	◯

8. Did the notification you get from the wearable device help you with logging events today? *

Mark only one oval per row.

	Never	Rarely	Sometimes	Very Often	All the time
Answer	◯	◯	◯	◯	◯

9. How do you rate the accuracy of the wearable in detecting your eating events today? *

Mark only one oval per row.

	Very Poor	Poor	Average	Good	Excellent
Answer	◯	◯	◯	◯	◯

10. How difficult was it to maintain your food journal today? *

Mark only one oval per row.

	Very difficult	Difficult	Neutral	Easy	Very Easy
Answer	◯	◯	◯	◯	◯

11. Can you tell us one thing you noticed about your dietary habits today? *

12. Can you share with us one memorable experience you had with food journaling today? *

Appendix C

Pre-Study Questionnaire

We will start with questions about goals and level of activities.

* Required

1. Email *

2. Do you agree with the following statement: I'm interested in monitoring my diet. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

3. Do you have previous experience with food journaling? *

Mark only one oval.

Yes

No

4. If you answered Yes to last question, please tell us about it.

5. What is your experience with self-tracking apps (fitness tracking, sleep tracking, glucose monitoring, etc.)? *

____

6. What is your experience with wearables (smart watches, smart glasses, bluetooth headphones, etc.)? *

____

7. What is your goal(s) from monitoring your diet? *

____

8. For the next two weeks, will you be mainly working from home? *

Mark only one oval.

- ◯ Yes
- ◯ No

9. For the next two weeks, how often do you expect to leave the house? *

Mark only one oval.

- ◯ Every day
- ◯ A few days a week
- ◯ Once a week
- ◯ Never

10. For the next two weeks, how often do you expect to eat outside? *

Mark only one oval.

- Every day
- A few days a week
- Once a week
- Never

Pre-Study Questionnaire (General Self-efficacy)

11. I will be able to achieve most of the goals that I set for myself. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

12. When facing difficult tasks, I am certain that I will accomplish them. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

13. In general, I think that I can obtain outcomes that are important to me. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

14. I believe I can succeed at most any endeavor to which I set my mind. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

15. I will be able to successfully overcome many challenges. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

16. I am confident that I can perform effectively on many different tasks. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

17. Compared to other people, I can do most tasks very well. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

18. Even when things are tough, I can perform quite well. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer					

Pre-Study Questionnaire (Diet-monitoring Self-efficacy)

19. I will be able to achieve most of the diet-monitoring goals that I set for myself. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer	⬭	⬭	⬭	⬭	⬭

20. In general, I think that I can obtain dietary outcomes that are important to me. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer	⬭	⬭	⬭	⬭	⬭

21. I believe I can succeed at monitoring my diet. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer	⬭	⬭	⬭	⬭	⬭

22. I am confident that I can effectively monitor my diet. *

Mark only one oval per row.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Answer	⬭	⬭	⬭	⬭	⬭

Appendix D

Interview Questions - Phase 1

Utility and adherence:

- How do you rate your overall experience with this journaling method?
- Can you mention any benefits of this method?
    - Can you mention any disadvantages of this method?
    - Have you used the reminder feature and do find it helpful?
    - how often you are getting FPs and what do you think the reason was? (fitting, user adaptive, high level of activity ...)
- How often did you miss logging eating events with this method?
    - Why did you miss these events?
    - Do you remember logging after you had the meal?
- Did you achieve your food journaling goal(s), and can you tell us more about that? (check pre-study survey)
- What things did you learn about your diet and daily activities during the study?
- Have you used the app/device for something other than food journaling?

Usability and Features

- What are features you like the most about the app and why?
- What are features that you didn't like and why?
- What features do you suggest we add to make your experience better?

Social acceptability and Privacy:

- Describe an experience journaling in a social setting.
    - What was the context?
    - Who was present?
    - How did you feel?
- How do you feel about having a camera on the device to capture images for you? Triggers automatically? manually?

Closing

- If applicable, ask them to explain or elaborate on some of their daily survey responses.
- Anything else you would like to share with us?

Interview Questions - Phase 2

Utility and adherence:

- How do you rate your overall experience with this journaling method?
- Can you mention any benefits of this method?
    - Can you mention any disadvantages of this method?
    - Have you used the reminder feature and do find it helpful?
    - how often you are getting FPs and what do you think the reason was? (fitting, user adaptive, high level of activity ...)
- How often did you miss logging eating events with this method?
    - Why did you miss these events?
    - Do you remember logging after you had the meal?
    - Compared to manual logging, did you notice any change in the rate of missed events?
    - How accurate do you think Fitbyte was in detecting eating?
    - How comfortable was FitByte to wear?
    - Generally, how would you compare the manual logging to using Fitbyte?
    - How much did you depend on the FitByte device notifications to do your logs?
- Did you achieve your food journaling goal(s), and can you tell us more about that? (check pre-study survey)
- What things did you learn about your diet and daily activities during the study?
- Have you used the app/device for something other than food journaling?

Usability and Features

- What are features you like the most about the app and why?
- What are features that you didn't like and why?
- What features do you suggest we add to make your experience better?

Social acceptability and Privacy:

- Describe an experience journaling in a social setting.

- What was the context?
    - Who was present?
    - How did you feel?
- How do you feel about having a camera on the device to capture images for you? Triggers automatically? manually?
- Did the device raise any privacy concerns to you or others?
- What would it take for you to use this device?

Closing

- If applicable, ask them to explain or elaborate on some of their daily survey responses.
- Why do you think your adherence was better in the second phase? (If applicable)
- Anything else you would like to share with us?