FitNibble: A Field Study to Evaluate the Utility and Usability of Automatic Diet Monitoring in Food Journaling Using an Eyeglasses-based Wearable

Abdelkareem Bedri
Carnegie Mellon University
Pittsburgh, USA
bedri@cmu.edu

Yuchen Liang
Carnegie Mellon University
Pittsburgh, USA

Sudershan Boovaraghavan
Carnegie Mellon University
Pittsburgh, USA

Geoff Kaufman
Carnegie Mellon University
Pittsburgh, USA

Mayank Goel
Carnegie Mellon University
Pittsburgh, USA

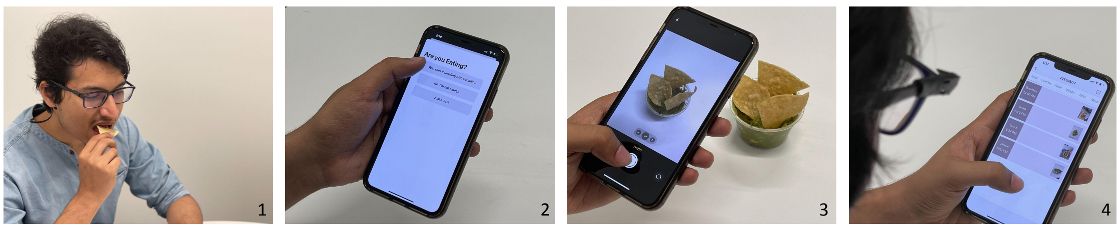

Figure 1: FitNibble continuously tracks a user's diet and provides just-in-time notifications reminding users to log their meals and snack. This figure shows the flow of the user experience. (1) As soon as the user starts eating, (2) the system sends a just-in-time notification, (3) prompting the user to fill the food journal and, (4) review their dietary activity for the day.

ABSTRACT

The ultimate goal of automatic diet monitoring systems (ADM) is to make food journaling as easy as counting steps with a smartwatch. To achieve this goal, it is essential to understand the utility and usability of ADM systems in real-world settings. However, this has been challenging since many ADM systems perform poorly outside the research labs. Therefore, one of the main focuses of ADM research has been on improving ecological validity. This paper presents an evaluation of ADM's utility and usability using an end-to-end system, *FitNibble*. FitNibble is robust to many challenges that real-world settings pose and provides just-in-time notifications to remind users to journal as soon as they start eating. We conducted a long-term field study to compare traditional self-report journaling and journaling with ADM in this evaluation. We recruited 13 participants from various backgrounds and asked them to try each journaling method for nine days. Our results showed that *FitNibble* improved adherence by significantly reducing the number of missed events (19.6% improvement, $p = .0132$). Results have shown that participants were highly dependent on *FitNibble* in maintaining their journals. Participants also reported increased awareness of their dietary patterns, especially with snacking. All these results highlight the potential of ADM in improving the food journaling experience.

IUI '22, March 22–25, 2022, Helsinki, Finland

ACM ISBN 978-1-4503-9144-3/22/03.
https://doi.org/10.1145/3490099.3511154

KEYWORDS

diet monitoring, eating detection, food journaling, wearable, utility, usability, and compliance

ACM Reference Format:

Abdelkareem Bedri, Yuchen Liang, Sudershan Boovaraghavan, Geoff Kaufman, and Mayank Goel. 2022. FitNibble: A Field Study to Evaluate the Utility and Usability of Automatic Diet Monitoring in Food Journaling Using an Eyeglasses-based Wearable. In *27th International Conference on Intelligent User Interfaces (IUI '22), March 22–25, 2022, Helsinki, Finland.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3490099.3511154

1 INTRODUCTION

What we eat has an immediate and long-term impact on our health. The medical research literature has shown direct links between diet and chronic illnesses. On the flip side, a healthy diet has always been helpful in fighting and preventing diseases [8]. As the psychological response to diet varies between individuals, nutritionists and physicians have always encouraged patients to keep track of what they eat and how their body reacts to it to understand better the impact it has on their health and well-being.

Despite all these benefits, self-monitoring of diet is rare, mainly because tracking daily activities at high granularity is an arduous and mundane task [17]. Fitness and sleep tracking have become increasingly popular and widely adopted in the last decade. Unlike sleep and fitness, diet monitoring products have mainly relied on self-report, and automation has proven hard. There are many research prototypes, but most either do not perform well in a real-world setting or are, currently, impractical to productize. Researchers have cited several technical and non-technical challenges, which represent clear barriers for this technology to reach the end-users [26]. 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 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, this was mainly attributed to wearable ADM systems which employee a wide range of sensing modalities to track chewing, swallowing, and hand-to-mouth gestures and use it as a proxy to detect eating and drinking events. Most of these systems have been developed and tested in lab environments to validate their functionality, but ecological validity remains a great challenge for most ADM setups. Non-technical challenges had also manifested when these ADM systems were evaluated in public, mainly concerning the social acceptability of the form factor and privacy concerns. [22, 26].

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 accurately detect 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 on people's health. Finally, the ADM system should also have a practical and socially acceptable form factor to help it make a real impact.

While there is still ongoing research on addressing the *when*, *what*, and *how-much* questions, this paper asks the question what values the current state of the art ADM may deliver to the end-user. In this work, we assess ADM's role in mitigating the challenges of typical food journaling techniques.

In this paper, we designed a field study to assess the utility and usability of *FitNibble*, a wearable ADM setup designed to recognize eating events and send just-in-time notifications to remind users to record meals and snacks at the right time. *FitNibble* design is based on recently published research on improving ADM performance in unconstrained environments. [4, 5, 33]. We mainly based our design on FitByte, an ADM platform based on eyeglasses form factor [4]. For this evaluation, we built an end-to-end ADM system using a modified version of FitByte and used an undated version of its eating detection ML models for real-time classification.

The field study discussed in this paper presents a preliminary investigation of the utility and usability of ADM systems through a small-scale deployment intended to demonstrate how augmenting typical food journaling procedures with our system can reduce journaling difficulty and improve compliance to the food journaling process.

From this small-scale deployment, our analysis shows that in one week *FitNibble* has improved compliance significantly by reducing the number of missed events (19.6% improvement, $p = .013$), and participants have exhibited explicit dependency on the wearable as soon as they started using it. Journaling difficulty has also dropped significantly after using FitNibble ($p = .005$). The device also helped most of our participants discover new dietary patterns, especially with the amount of snacking. Moreover, 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 paper can be summarized in the following points:

(1) An end-to-end open-sourced system includes wearable hardware schematics and firmware, smartphone app, ML model, and back-end server code. (link).
(2) A field study to assess the utility and usability of a diet monitoring wearable and provide a list of recommendations for the design of future ADM systems.

2 RELATED WORKS

This section will discuss the basic concepts of food journaling and discuss the common approaches used to monitor diet. We then review the previous work on automatic diet monitoring and discuss examples from the literature. We conclude by reviewing previous ADM field deployments and discussing the need for field studies that focus on the utility and usability of ADM.

2.1 Food journaling

Food journaling has been an effective method to combat diet-related diseases and help individuals lead a healthy lifestyle. In previous studies, researchers witnessed that journalers are more mindful of their diet and they are more encouraged to avoid unhealthy foods [18, 31].

Self-report is the most common diet monitoring method. Food frequency questionnaires and 24-hour recalls represent health experts' most typical journaling methods. 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 [2, 15, 32]. 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.

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 leads to reduced compliance [3, 13]. In [17] 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.* [11, 12] have investigated the barriers and challenges for different food journaling methods and found that loss of motivation, time commitment, and the considerable effort required to maintain a journal are the most common reasons people cite to explain why they stopped journaling before they reach their goals. When investigating further why journalers miss eating events, Cordeiro *et al.* found that the most common reasons are: forgetting to log, lack of nutrients information, the shame of unhealthy choices, stigma from journaling in front of others, and social context where it is challenging to use phones (e.g., meetings, or class)

In [11] Cordeiro *et al.* propose the use of photo-based food journaling, which requires 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 pictures 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 could use the saved photos to look up content conveniently. This method doesn't require users to count calories, making participants less anxious and reducing 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 to help users recall log eating events as soon as they happen. This paper introduces a wearable automatic diet monitoring system, FitNibble, which automatically reminds users to log their food by detecting intake episodes. Our field deployment shows that our wearable could significantly reduce the number of missed events.

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 achieved on identifying food contents automatically [22, 26, 28]. ADM systems use sensors that detect one or more eating actions like chewing, swallowing, and repetitive hand-to-mouth gestures to identify eating moments. This section reviews a wide range of ADM systems from the literature.

2.2.1 Detecting Hand-to-mouth Gestures. Observing hand movements as a proxy to detect eating has been a well-studied approach [1, 14]. Sen *et al.* [27] used the accelerometer and gyroscope embedded in an off-the-shelf smartwatch to detect eating events, and the smartwatch they used has a camera that captures food images. Lin and Hoover [19] also used a smartwatch inertial sensor to monitor the number of bites taken during a meal. Thomaz *et al.* [29] 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 and classified between them.

2.2.2 Detecting Chewing. In the literature, several sensing approaches have been employed to detect chewing. GlassSense [10] 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 [16] used a piezoelectric strain sensor placed on the temporalis muscle to detect chewing bouts. Bedri *et al.* [6, 7] 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.* [9] 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 detect chewing [23]. Bedri *et al.* [5] also used inertial sensors on an earpiece to detect chewing. Zhang and Amft built custom 3D printed eyeglasses with EMG sensors [33, 34]. The EMG dry electrode is placed on the eyeglass's temples to monitor the mastication muscle movement.

2.2.3 Detecting Swallowing. To detect consuming liquids and solids, one of the most promising approaches is to listen to throat sounds to detect swallowing. Rahman *et al.* [24] have used a piezoelectric microphone on the neck to detect sounds of drinking, eating, and other activities. In [21] 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.

2.3 Challenges with field deployments

In their review paper, Schiboni and Amft [26] 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. Some researchers have addressed these problems in recent years and provided solutions to mitigate these challenges. For example, Zhang and Amft built custom 3D printed eyeglasses with EMG sensors to detect mastication from the Temporalis muscle movement [33, 34]. The system achieved 95% accuracy in detecting eating episodes in unconstrained environments when tested. Another example is FitByte [4] that is an automatic diet monitoring system also based on an eyeglass form factor that utilizes a set of inertial sensors to detect chewing and swallowing and a proximity sensor to detect hand-to-mouth gestures. FitByte achieved 92.7% F1-score in detecting both eating and drinking episodes that lasted for more than 10 seconds 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.

To our knowledge, only a few ADM systems have been evaluated in field deployments of a week or more. These deployments have used an off-the-shelf smartwatch to detect eating from hand-to-mouth gestures. Thomaz *et.al.* [29] 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.* [30] have deployed another watch-based ADM for 4 weeks with 12 participants. The study's goal 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 bites detected. Participants had to remember to turn on the byte counting App every time they ate 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.* [20] 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). Still, the authors didn't thoroughly investigate the system's usability nor the impact it had on the user experience with food journaling.

This paper presents an in-depth analysis of the user experience with ADM and its impact on compliance with the food journaling process. Our FitNibble system was capable of detecting meals and small snacking events in real-time. This feature greatly impacted the overall experience and significantly improved compliance with the food journaling process.

3 METHOD

The main goal of this evaluation is to assess the value ADM can provide to the food journaling experience and understand the system's influence on journaling compliance and other usability and utility factors.

In this evaluation, we aim to use *FitNibble* to send just-in-time notifications to remind the users to do their logs every time it detects they are eating. We hypothesize that using this approach can lower the cognitive load required by self-report journaling methods and reduce the number of logging errors, especially when it comes to missed events. This tool should significantly improve the user experience and help them adhere to the food journaling task.

In our analysis, we used several metrics to assess the user experience. These metrics include utility, usability, social acceptability, and any privacy concerns raised from using this system. We used a combination of quantitative and qualitative methods to assess these metrics. The results of this evaluation were then used to inform new design recommendations for ADM systems.

3.1 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 interested in understanding their dietary behavior in general and not focused on specific goals like weight loss. We had the second criteria only to recruit individuals who wear eyeglasses regularly because our ADM setup is based on that form factor. We didn't want the participants' experience to be influenced by their unfamiliarity with wearing eyeglasses all day.

The study has two phases; each phase should last for nine days (18 days total). This study will also introduce participants to a traditional photo-based food journaling method. This method requires the user to take photos of their meals and snacks throughout the day and review them before going to sleep. This approach delivers sufficient information to help users understand their dietary patterns without focusing on minute details like calorie count. We chose this approach because it requires minimal effort from the users [11].

This study was designed to target users already familiar with journaling but struggling with compliance. Therefore, we intentionally did not counterbalance the order of interventions because our research question focuses on studying the impact on the user experience after the transition from the status-quo of journaling without ADM to journaling with ADM. Since all of our participants did not have prior experience with photo-based journaling, we used this order to ensure that all participants are familiar with the journaling process before introducing them to the wearable.

3.1.1 *Phase1: Photo-based journaling without ADM.*

In the first phase, participants will be asked to use the *FitNibble* app (check the description in section 4.3) 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 remember to log if they know when they are most likely to eat. To help participants distinguish between snacks and meals, we defined snacks as any eating event outside breakfast, lunch, and dinner. Participants can also log drinking events, but it is not required.

We require participants to use this journaling method for nine days. In this period, they will have two days to get familiar with the journaling method and seven days for the actual data collection. At the end of each day, we asked participants to fill out 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, focusing on utility, usability, social acceptability, and impact on compliance. For the interview questions, check Appendix D.

3.1.2 *Phase2: Photo-based journaling with ADM.*

We introduce the participants to our wearable ADM setup in the second phase. At the beginning of this phase, we install the hardware setup on the user's glasses as described in section 4.2. We then explain to the participant how to put on the wearable setup and how to connect the wearable to the phone app via Bluetooth. This new setup should allow the app to send reminder notifications to users every time they think they are eating and ask them to log.

Before starting the study, we ensure the system is working correctly by performing a functionality test. We ask the user to simulate an eating event by chewing and performing several hand-to-mouth gestures in the test. We run this test multiple times to ensure that the system detects eating events reliably.

Like the first phase, participants will use this journaling method for nine days (2 days to get familiar with the method and seven 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 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. We used an emergent thematic coding method to analyze the interview data.

3.2 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 used the app and how often they used specific features, such as the time-based reminder. We also recorded their responses to the wearable notifications to help us track the number of true positives and false positives per day. We recorded none of the logs data and photos to preserve participants' privacy.

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 [25]. 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 (five female and eight male) with an average of 34 years, ranging between 21 and 54 years. We recruited these participants by posting a public ad on the Craigslist website. This helped us recruit from the general population of our city. We also advertised for the study through several university mailing lists.

Our study population had six university students, and the remaining seven were from diverse backgrounds, including high school teachers, artists, electrical technicians, and stay-at-home moms. All participants were screened to confirm they meet the study criteria and ensure they don't have any personal relationship with members of the research team.

Since we conducted this study during the COVID 19 pandemic and as the city lifted lockdown restrictions, we asked participants to report their expected activity level. They reported their activity for the 18 days of the study, including when they worked from home, left for the office, or ate outside (Appendix C). In our recruitment, we didn't require participants to have a certain activity level.

4 SYSTEM DESCRIPTION

4.1 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 2). The wearable handles 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 handles the features and sends it to a server that 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.

4.2 Wearable

The FitNibble wearable we developed is informed by the original FitByte design and field evaluation [4]. Appendix E details our wearable design process and explains all the design decisions made at every iteration.

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 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 3.right). Like 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 extends to a cloth pocket attached to the back of the user collar (Figure 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 eyeglass' weight is as light as possible and place all the heavy components on the back, which was inspired by the Earbit design [5]. We designed FitNibble to be attached to any pair of eyeglasses. In the study, we chose to use the participants' glasses to avoid any discomfort 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. Like the original FitByte pipeline, we calculate features 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. Like the original FitByte pipeline, we preprocessed the data. We 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 features).

The only significant 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 4 illustrates all the steps done in each component of the system. The overall power consumption of the system is 25 mAh.

4.3 FitNibble iOS App

On the mobile side, we developed the *FitNibbleApp*. 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 *FitNibbleApp* to a secure off-the-shelf journaling App, *Foodility*. The app allows users to do their logs and save the information away from the *FitNibbleApp*, so the research team doesn't access participants' private data. *Foodility* (Appendix E) is a simple food journaling app on the App Store that allows users to track their food consumption securely. With *Foodility*, users can select meal types, take short notes, and manually log their estimated calorie intake. Moreover, *Foodility* possesses the feature of taking a photo of the food, which most other journaling apps do not offer. 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 *FitNibbleApp*, 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 study at a specific time of the day (usually in the evening). The app also features setting daily journaling reminders at particular 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

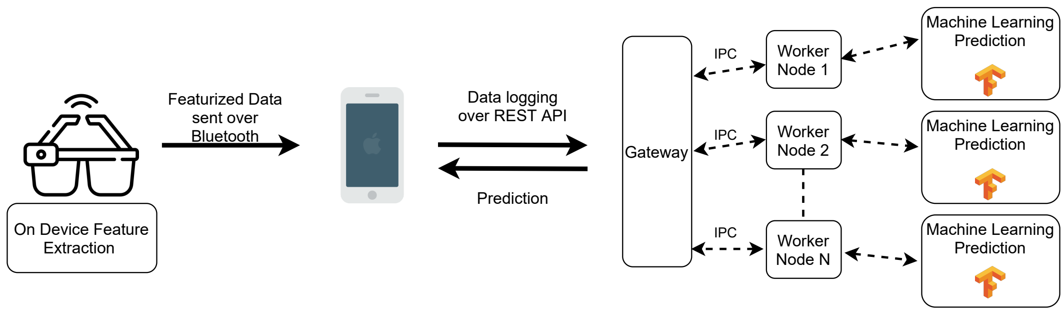

Figure 2: System architecture: FitNibble wearable sends the extracted features over Bluetooth to our iOS mobile App. The backend obtains this data over REST API and then forwards it for data logging and model prediction of eating.

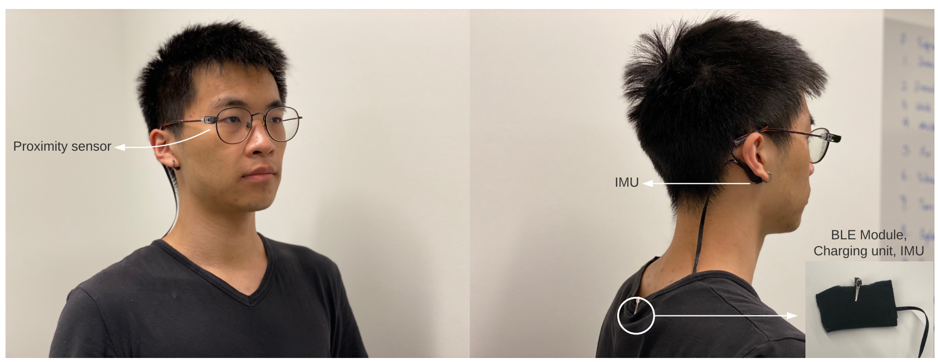

Figure 3: The wearable has a proximity sensor to detect hand-to-mouth gestures (left), an IMU in contact with the lower jaw bone to detect chewing (right), and small cloth pocket clipped to the back of the user's shirt containing the BLE module, battery, and the reference IMU (bottom right).

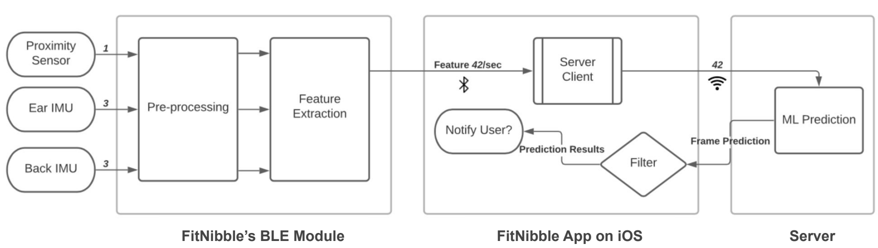

Figure 4: The flow of data and information across the the three main components of the end-to-end system.

the wearable after turning on Bluetooth pairing in the app, and they can connect to the wearable by tapping their device on 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 five consecutive responses infer the participant is eating, the app will 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 received a false positive. The "just a test" button is there for the user to check the device 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 study, *FitNibbleApp* 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. We then send these activity logs to the server.

4.4 FitNibble Backend Server

Our backend server is 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.

4.4.1 Rest APIs for Data Logging and Eating Detection. We implemented server REST interfaces to our *FitNibble* backend focused on logging the user interactions with the *FitNibbleApp*. This logged 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 4.3.

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 to process the prediction requests and send the prediction back to the app, which is then shown as a notification.

4.4.2 Machine Learning. We explored several machine learning classifiers to build reliable models suitable for detecting eating in real-time. We trained all machine learning models on the publicly available *FitByte* dataset [4]. The dataset consists of several activities such as eating, drinking, walking, talking, and silence. Using this data, we extracted seven features from the sensor channels corresponding to the available sensing modalities on *FitNibble*. These seven features are the same as those recommended in the *FitByte* paper(i.e., entropy, variance, absolute median, zero-crossing count, zero-crossing variance, and the RMS of each channel). To compare classifiers performance, we ran a leave-one-participant-out (LOPO) cross-validation to measure accuracy at the frame level using a sliding window of 5 seconds sliding by 1 second. In our comparison, we used ten different machine learning classifiers including, KNN, SVM, Random Forest, AdaBoost, and DNN. Our evaluation showed that an unoptimized DNN model outperforms all other classifiers with a frame-level accuracy of 80% (76% precision and 74% recall). Therefore, we decided to use this model for real-time eating detection.

Figure 6 shows our custom-built DNN model architecture. The model has 13695 hyper-parameters with 70 hidden layers. The input layer is a vector with 42 elements, a flattened representation of the six data dimensions. To feed the data into our neural network, we shape it so that each person has multiple two-dimensional records, which holds the data for each of the sensors from the *FitNibble* wearable device. Each record is also associated with one label, which feeds into the neural network during the training process. Prediction from DNN gets passed to the *FitNibble* App every second. The app would detect an eating event if it received five consecutive eating predictions. We defined the parameters of this aggregating step after experimenting with different values in a pilot with 5 participants. The chosen parameters resulted in the highest accuracy (precision and recall) for detecting eating episodes.

5 RESULTS

This section presents our findings from the data we collected, including the daily surveys, interviews, and App usage data. As we gave participants two days to get acquainted with the journaling method, we only included data from the last seven days of each phase in our analysis. 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 aspects: (1) compliance, (2) utility and usability, and (3) social acceptability and privacy concerns. We summarize our findings from the qualitative and quantitative data analysis for each aspect.

5.1 Compliance

The main challenge with food journaling is the low compliance rate. ADM systems are designed to improve compliance 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 compliance. This section presents our results on how our ADM system FitNibble impacted the overall compliance by comparing results from phase 1 (without FitNibble) and phase 2 (with FitNibble).

Figure 7 presents the percentages of days with missed logs as reported by participants in their daily survey. The pie charts represent the response to the question "Did you miss logging any events today? ". In Phase 2, the percentage of *No* responses (i.e., No missed events today) increased by 19.6%, which indicates that participants were less likely to miss logs. This is a clear sign of improved compliance 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 compliance was significant X^2 (2, N =163)=6.1478 $p = 0.013158$.

When investigating the reason behind this improvement in compliance, we wanted to examine the possibility of any carry-over effects. One assumption that can be made here is this improvement is due to the familiarity with the journaling method, as participants have been doing it for more than a week. Therefore, we examined the reported data over time and looked for any trends that could have been carried over from phase 1 to phase 2. Figure 8 shows the percentage of the *"No"* responses per day for both phases. It is clear from the data that compliance did not improve over time in

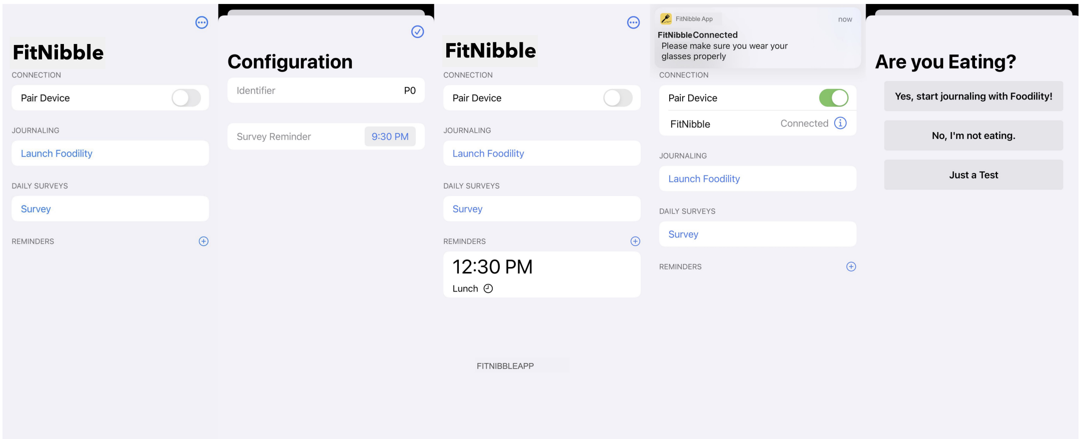

Figure 5: This figure shows the main page of the FitNibbleApp and explains the main functionalities available on it, including setting reminders and receiving notifications when eating is detected.

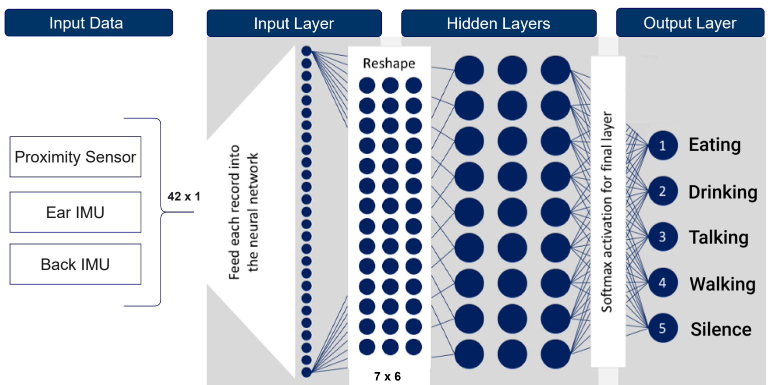

Figure 6: FitNibble's DNN model topology.

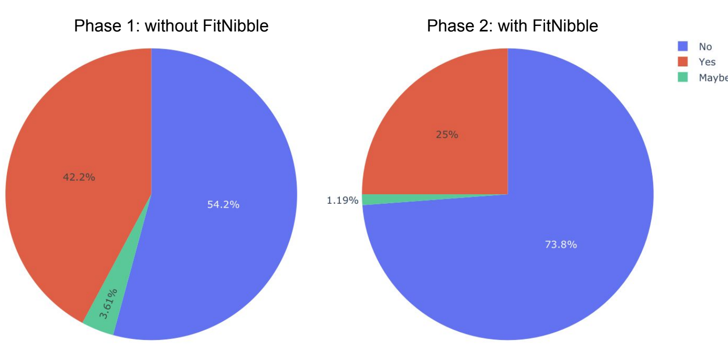

Figure 7: Percentage of days with missed logs before and after participants starts wearing FitNibble.

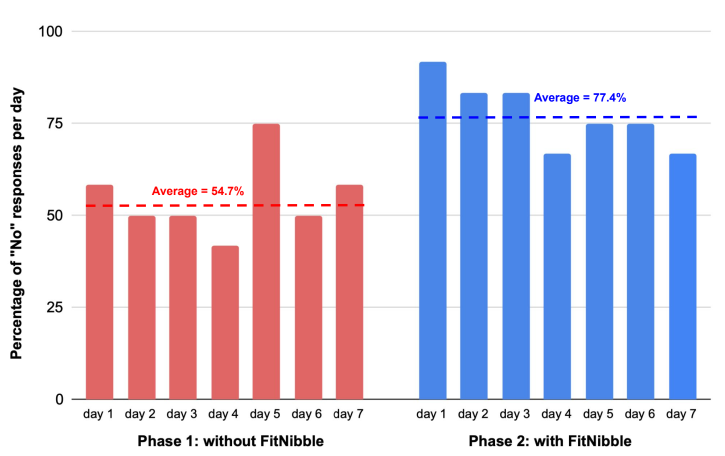

Figure 8: The daily percentage of No responses for phase 1 and phase 2.

Phase 1, and there is no visible learning effect. While in Phase 2, we witnessed a big jump in compliance from the get-go (Phase1-day 7: 58.3% ⟶ Phase2 - day 1: 91.6%). The average percentage per phase has also improved with FitNibble (54.7% ⟶ 77.4%), and it's worth noting that all the reported percentages in phase 2 are above the average of phase 1. All these findings indicate no carry-over effect between the two phases.

To investigate further the reason behind this sudden improvement in compliance, we asked participants in the exit interview why they started to report more *Nos* in the second phase. Most participants attributed this improvement to their experience with the wearable. *P*10'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's contribution to the improved compliance was examining the number of times participants used the wearable notifications to open the app *vs*. directly opening it from the home screen. This data would explain how often participants depended on the wearable notifications in doing their logs. After analyzing the logs, we found that nearly half of the recorded eating events were initiated by a wearable notification. This result indicates that FitNibble played a significant role in reminding participants to do their logs.

These results support the argument that FitNibble is the main factor behind the observed improvement in journaling compliance.

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

5.1.1 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.* [12] in manual food journaling studies. *P*9 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, *P*8 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 missed events. "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" (*P*1). In phase 2, participants' compliance 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 *P*13 "*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.*"

5.1.2 Missing snacks. Another common theme from phase 1 interviews is that most participants realized that they missed journaling small meals such as snacks more than main meals. *P*3 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 *P*12 said "*Yeah, I think I've missed pretty much all the snacks.*". In phase 2, we saw the complete opposite; participants became aware of their snacking habits. For example, *P*4 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 *P*11 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 a snack I am relying on the glasses 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 (1). To evaluate the difference between journaling with and without FitNibble wearable, we ran a repeated measure ANOVA [1] test for both meals and snacks for both phases and the differences were not significant ($F_meal(2, 83) = 0.39, p_meal = 0.844, F_snacks(2, 83) = 0.39, p_snacks = 0.99$).

5.1.3 Depending on the wearable's notifications. In the second phase, the participants depended on FitNibble in the journaling

[1]We chose repeated measures ANOVA because it is fairly "robust" to violations of normality assumption which is common in small data samples.

process. Note that we explicitly instructed participants not to rely on the device and continue to log events when they remember, but most participants soon after 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 compliance 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 in the average scores (beginning: 3.7, end: 3.85), but we didn't find any correlation between these scores and the participants' level of compliance.

5.2 Utility and Usability

This section discusses the usability of all the features introduced in the FitNibble 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.

5.2.1 Time reminders has low value. One of the features we introduced in the FitNibble App allows users to set logging reminders at a specific time if they know they are most likely to eat at a particular time. This feature was available in both phases of the study, but the general feedback we received was that participants didn't find it helpful 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 1. We noticed that more participants found the feature helpful in phase 2. Still, 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 challenging 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 mealtime is not fixed.*"

5.2.2 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 opened 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 recognize from the daily surveys that the journaling difficulty has dropped in the second phase 1. Figure 9 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 ($F(2, 83) = 5.524, p = 0.021$).

In the second phase, we asked participants to report how often the wearable notifications helped them today in the daily survey. The average rating for this feature was 3.3± 0.8 (above midpoint)1.

All these results point to the positive experience with the wearable, but a few 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. 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 forgetful and less punctual users. Still, for users who don't have these issues, journaling with ADM negatively affects their experience.

5.2.3 Variable perception of accuracy. When it came to how participants perceived the accuracy of the wearable notifications, there was a split between the good and bad responses. 61.9% of the daily response found the accuracy to be average or above average, while 38% of the responses found it below average. 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 logs, we found that half of the participants received 0 to 5 false positives per day, and the other half received 6 to 11 false positives per day. Figure 10 shows the distribution of participants according to their average false positives per day. After reexamining the rating data, we found that participants tend to give a lower rating if they receive six or more false positives per day, which explains the split in perceived accuracy.

We also investigated how perceived accuracy can influence the overall user experience. Despite the high false-positive rate for some participants, they still saw value in using the wearable. For example, $P11$ who was receiving on average 11 FPs/day said "*The device reminds falsely often, but that helped me remember to log the food. The device also helped quite a bit when it comes to snacks outside of the regular routine.*"

One reason that can explain the variable precision across participants can be the improper placement of the wearable or loose-fitting. This might explain why the accuracy is user-dependent. At the beginning of phase 2, we trained participants on how to put on and take off the wearable and ensure proper fit, but there is no way for us to know how well they complied with our instructions in the

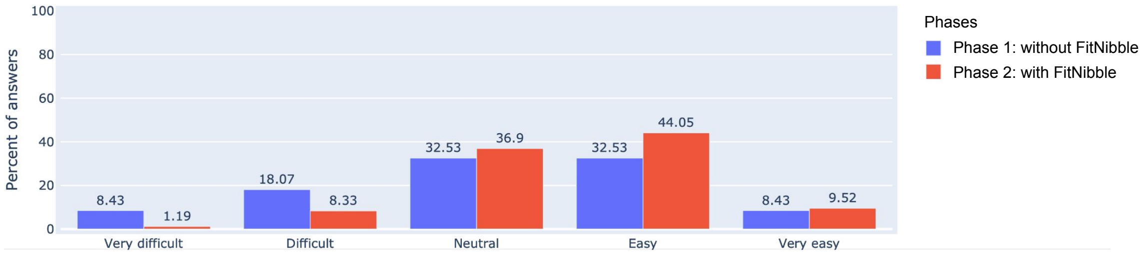

Figure 9: Journaling difficulty before and after participants start wearing FitNibble.

field. Further investigation is required to understand the factors influencing wearable precision.

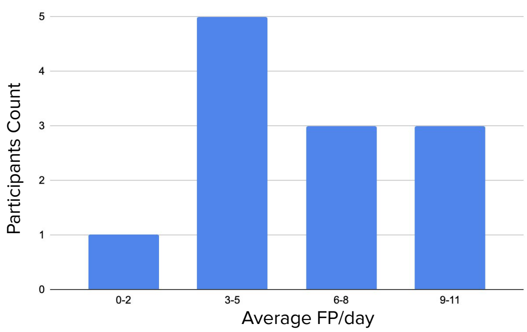

Figure 10: Histogram of study participants based on the average false positives they receive per day.

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 a connection or a fitting problem most of the time. *P*13 said, "*one thing I notice for the false negatives it's probably just because I'm not wearing my glasses properly.*"

5.2.4 Increased awareness of dietary patterns. One central 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, *P*3 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 *P*5 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, *P*12 said "*I realized that I eat irregularly at night which will range from 5:00 PM to 9:00 PM*", and *P*3 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. *P*5 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.*".

5.2.5 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 from being widely adopted. Most participants said they would use FitNibble if we improved the prototype to a finished product quality. Some suggested changes to the form factor to enhance comfort. For example, *P*1 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 *P*3 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 across multiple days. One other desired feature was integrating other health-related journals like fitness tracking, glucose level, calorie counting, and mood logs.

5.3 Social Acceptability and Privacy Concerns

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

5.3.1 Social acceptance. The central theme under this topic was the wide acceptance of the FitNibble 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 found it "cool". *P*1 mentioned "*Really cool, yeah. They all think it's really cool!*", and *P*9 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. *P*11 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*". *P*6 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 FitNibble design makes it invisible to others. For example, *P*7 said "*Most of the time I think people don't even notice*" and *P*5 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*".

5.3.2 Social collaboration. This is one of the exciting themes we found in the data. Some participants mentioned that they depended

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

on their friends and family to remind them to log meals in the first phase. $P3$ 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 remind 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"*.

5.3.3 Having complete control over the wearable camera. As we mentioned in Section 4.2, we were considering including an onboard 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 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 onboard camera and demanded it be removed from the wearable. However, after using FitNibble in phase 2, he changed him opinion and said, "*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 had complete 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 DISCUSSION

This paper tried to assess the value an ADM system can provide to the end-user. The results of this preliminary investigation highlighted the potential of ADM in increasing compliance to food journaling and improving the user experience with the process. In this section, we will go over the primary outcomes from the study, provide a list of design recommendations for the next generation of ADM systems, and discuss study limitations and future work.

6.1 Main outcomes

In a period of one week, our results showed around a 20% drop in days with missed events, and we saw a significant improvement in compliance with $p = .013158$. The main reason participants cited for this improved compliance 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 FitNibble, participants became more aware of their dietary behavior, especially when snacking, as most missed small eating events.

While the rate for false positives was variable across participants, the overall perception of accuracy leaned towards the positive side. Also, journaling difficulty dropped significantly after using FitNibble, and most participants saw value in using the wearable. Most participants said they would use FitNibble in their daily lives 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 no significant barriers to adopting this wearable ADM as a product for daily use.

Many participants felt that doing food journaling in public was socially acceptable, and even some participants relied on their friends and family to remind them to log. This finding is a sign that now 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 using a wearable camera to help users take photos of their food. While all participants felt this feature would raise many privacy concerns, after using FitNibble, 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 to accept the camera feature participants demanded complete 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.

6.2 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.

6.2.1 Targeted users. Our evaluation 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 device notifications annoying. Therefore, we recommend that designers keep these differences in mind when defining their targeted end-users.

6.2.2 *Acceptable range of error.*

One other question we tried to answer in this evaluation was *"What is acceptable number of false positives per day?"*. The feedback we received from participants was very similar despite the discrepancy in the false-positive rates. 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. ADM designers should keep in mind 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 chewing gum, biting their nails, or drinking. Our model design considered all these actions to count as eating events. These participants were not fully aware of that, which influenced their experience with the wearable.

6.2.3 *Improving interaction.*

Through the interviews we conducted, we received many recommendations on improving the interaction with the wearable. For example, some suggested using 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 that when they eat at home, they don't usually have their phone with them to miss the notifications. One participant liked that FitNibble 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 started eating and would mark it as a false positive, but when she eats, she doesn't get a notification because the device was snoozed. Worth noting that this type of false positive was still helpful 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 a 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 FitNibble, in improving the user experience with food journaling.

6.3 Limitations and Future work

The outcomes of this study highlight the values ADM has in improving the food journaling experience. That being said, this study has limitations and can only be described as a small-scale deployment intended to demonstrate how augmenting typical food journaling procedures with our system can help improve compliance. We understand that the current study design does not provide a counterbalanced evaluation, and the order effects can influence the presented results. Therefore, the next step after this preliminary investigation is to verify these findings in a fully-fledged between-subjects experiment to demonstrate the superiority of journaling with ADM versus self-report.

7 CONCLUSION

In this paper, we present an in-depth analysis of the utility and usability of *FitNibble*. Our long-term field deployment allowed participants to experience the difference between traditional self-monitoring methods and journaling with the aid of an ADM system. Our analysis indicated that participants relayed on *FitNibble* to remind them of logging whenever they start eating. Participants indicated that using textitFitNibble significantly reduced the cognitive load required to maintain their journals, as they don't need to pay attention to their activities constantly. This feature *FitNibble* helped improve adherence to food journaling by significantly reducing the number of missed events ($19.6\% improvement, p = .0132$). Participants believe that journaling with *FitNibble* made them more aware of their dietary behavior, especially when it comes to snacking, which has high health implications but is usually missed by self-monitoring methods. Our participants saw the value of journaling with ADM and are willing to use this technology in its finished product form. All these outcomes highlight the potential of ADM in improving the food journaling experience and making it wildly adopted at the same level of fitness tracking.

REFERENCES

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

[2] Edward Archer, Gregory A Hand, and Steven N Blair. 2013. Validity of US nutritional surveillance: National Health and Nutrition Examination Survey caloric energy intake data, 1971–2010. *PloS one* 8, 10 (2013), e76632.

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

[4] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. 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*. 1–12.

[5] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 37 (Sept. 2017), 20 pages. https://doi.org/10.1145/3130902

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

[7] Abdelkareem Bedri, Apoorva Verlekar, Edison Thomaz, Valerie Avva, and Thad Starner. 2015. 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* (Osaka, Japan) *(ISWC '15)*. Association for Computing Machinery, New York, NY, USA, 91–92. https://doi.org/10.1145/2802083.2808411

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

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

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

[11] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. 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*. 3207–3216.

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

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

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

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

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

[17] Elina Helander, Kirsikka Kaipainen, Ilkka Korhonen, and Brian Wansink. 2014. 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 (2014), e109.

[18] Frederick H Kanfer. 1970. Self-monitoring: Methodological limitations and clinical applications. (1970).

[19] B. Lin and A. Hoover. 2019. 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)*. 1–4. https://doi.org/10.1109/BSN.2019.8771066

[20] 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. 2020. 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 (2020), e20625.

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

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

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

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

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

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

[27] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Balan, and Youngki Lee. 2018. Annapurna: building a real-world smartwatch-based automated food journal. In *2018 IEEE 19th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 1–6.

[28] Thomas Stütz, Radomir Dinic, Michael Domhardt, and Simon Ginzinger. 2014. Can mobile augmented reality systems assist in portion estimation? A user study. In *2014 IEEE international symposium on mixed and augmented reality-media, art, social science, humanities and design (ISMAR-MASH'D)*. IEEE, 51–57.

[29] Edison Thomaz, Irfan Essa, and Gregory D Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 1029–1040.

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

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

[32] Walter Willett. 2012. *Nutritional epidemiology*. Oxford university press.

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

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