MyDJ: Sensing Food Intakes with an Attachable on Your Eyeglass Frame

Jaemin Shin KAIST Daejeon, Republic of Korea jaemin.shin@kaist.ac.kr
Seungjoo Lee KAIST Daejeon, Republic of Korea seungjoo.lee@kaist.ac.kr
Taesik Gong KAIST Daejeon, Republic of Korea taesik.gong@kaist.ac.kr
Hyunchul Roh DYPHI Inc. Seoul, Republic of Korea roh@dyphi.com
Andrea Bianchi KAIST Daejeon, Republic of Korea andrea@kaist.ac.kr
Sung-Ju Lee KAIST Daejeon, Republic of Korea profsj@kaist.ac.kr

ABSTRACT
Various automated eating detection wearables have been proposed to monitor food intakes. While these systems overcome the forgetfulness of manual user journaling, they typically show low accuracy at outside-the-lab environments or have intrusive form-factors (e.g., headgear). Eyeglasses are emerging as a socially-acceptable eating detection wearable, but existing approaches require custom-built frames and consume large power. We propose MyDJ, an eating detection system that could be attached to any eyeglass frame. MyDJ achieves accurate and energy-efficient eating detection by capturing complementary chewing signals on a piezoelectric sensor and an accelerometer. We evaluated the accuracy and wearability of MyDJ with 30 subjects in uncontrolled environments, where six subjects attached MyDJ on their own eyeglasses for a week. Our study shows that MyDJ achieves 0.919 F1-score in eating episode coverage, with 4.03× battery time over the state-of-the-art systems. In addition, participants reported wearing MyDJ was almost as comfortable (94.95%) as wearing regular eyeglasses.

1 INTRODUCTION
Food journaling is an effective method that is widely recommended by clinicians and dietitians for maintaining healthy eating habits. Writing a food journal brings awareness of the food intake and leads to a healthy choice of food and effective weight and chronic diseases management [33, 51, 52]. While there are tools that facilitate interactive food journaling, such as web or mobile food logging apps [23, 38, 60], the manual effort involved in the journaling process often results in losing habit in the long-term [19, 20].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA.
© 2022 Copyright held by the owner/authors. Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9157-3/22/04...
https://doi.org/10.1145/3491102.3502041
MyDJ achieves an average accuracy of 0.984 and an F1-score of 0.919 in eating episodes detection while achieving 4.03× battery life improvement over a previous eyeglass eating detection system [7]. To evaluate the long-term accuracy and wearability of MyDJ, we collected 477 hours of data from six participants, who attached MyDJ on their own eyeglasses for a week in uncontrolled environments. Throughout the week-long study, MyDJ detected 111 out of 120 meals or snacks. Furthermore, our user survey shows that the comfort level of wearing glasses with MyDJ attachable is 94.95% compared with wearing own eyeglasses.

This paper contributes to the field of HCI as follows: We present (1) a design and implementation of an attachable eating detection system that easily integrates on any eyeglass frames; (2) an evaluation based on a week-long in the wild data collection in which users attach MyDJ on their eyeglasses. Our data collection is quite extensive as the longest data collection from previous eating detection eyeglasses was for two days [7].

2 BACKGROUND AND RELATED WORK

We first define eating and eating episode, then survey eating detection systems in the form of eyeglasses and other wearable devices.

2.1 Definition of Eating and Eating Episode

We use the definition of eating and eating episode from Bi et al. [13] throughout this paper. Eating is defined as “an activity involving the chewing of food that is eventually swallowed”. We thus exclude the detection of drinking or chewing gums. Eating episode is defined as “a period of time beginning and ending with eating activity, with no internal long gaps, but separated from each adjacent eating episode by a long gap”, where a gap is a period in which no eating activity occurs, and where long means a duration greater than a specified parameter. In this paper, we used 15 minutes as a parameter to specify the long gap between the eating episodes, as in [13]. Note that the shortest eating episode on our system is 15 seconds, which allows the detection of most eating episodes, including short snacking events.

2.2 Eating Detection on Wearable Form Factors Other Than Eyeglasses

Wearable eating detection systems have been proposed for various form factors. Bi et al. proposed a form of headgear [13] and headband [12] that both utilize piezoelectric sensing on a mastoid bone for capturing chewing signals. While accurate and energy-efficient, wearing headgear or headband is socially unacceptable in various situations. Neckbands [25, 68, 82] that use acoustic sensing to capture chewing and swallowing sounds have limited usability in warm weather due to the sweat between the band and the neck [75]. Moreover, constant audio sensing could lead to privacy concerns. Necklaces [2, 17, 39, 74, 84] utilizing proximity sensing to track jaw movements or piezoelectric sensing to capture the throat vibrations are also popular. While necklaces are common and socially acceptable to wear, these form factors are not widely accepted as a survey reports that 45% of people would never wear such form factors [5], whereas 64% of the US wear eyeglasses on a daily basis [29]. Moreover, proximity sensing on the necklace is prone to error under direct sunlight or user movements while eating [17, 62]. Other systems that leverage in-ear proximity sensing on an earpiece [9, 10], acoustic sensing on a Bluetooth headset [31], or surface pressure sensing on a cap [87] could be unacceptable in a social dining situation. Lastly, smartwatch-based eating detection systems [22, 43, 72, 78] perform hand-to-mouth gesture recognition and require users to wear it on their dominant hand for eating detection. It is also shown that such a system suffers from high false positives in uncontrolled settings [17].

In summary, the performance of the surveyed eating detection wearable form factors degrades in uncontrolled, real-life settings. Moreover, certain wearables could be inappropriate to wear in social dining situations. We believe eating detection on eyeglasses could be a viable option as it could be easily worn during any dining experience and could also achieve high accuracy due to close proximity to the chewing location [41]. We now review previous research on eating detection using eyewear.

2.3 Eating Detection on Eyeglasses

Eating detection on eyeglasses could offer accurate eating detection thanks to the sensor placements close to the mouth and jaws (i.e., where chewing and swallowing happen) while simultaneously providing a socially acceptable wearable form factor [41].

Previous approaches on eyeglasses, however, mostly require custom-built eyeglass frames for specific sensor placement, which limits their adoption to users who wear non-instrumented commodity eyeglasses. Some of the proposals were uncomfortable to wear or failed to achieve accurate or energy-efficient eating detection in real-world deployments. The system by Zhang et al. [83], for instance, requires personalized frames to place electrodes to the human skin for Electromyography (EMG) sensing. Its accuracy suffers when sweat or hair get in between the electrodes and the user’s skin [1]. Some methods require placing piezoelectric sensors [27, 28] directly in contact with the skin using medical tape, which hinders comfort. Systems solely based on accelerometer [29] are inaccurate for users who are not aggressive chewers (as we discuss in Section 3 and 5). FitByte [7] utilizes sensor-fusion with gyroscopes, an accelerometer, and a proximity sensor placed on the eyeglasses frame. However, it drains a battery in less than a day (with the same battery as recent commodity smart eyeglasses [80]). It also requires the eyeglasses temple to be built with flexible materials to ensure a snug fit to improve IMU readings. Rahman et al. [67] and Chung et al. [18] proposed using inertial sensors and load cells respectively, but the accuracy of both of their design was evaluated only in controlled lab settings. Mirtchouk et al. [57, 58] used a combination of inertial and acoustic sensing on eyeglasses, smartwatches, and earbuds. However, wearing earbuds is typically not an acceptable social behavior and their reported eating detection F1-score was lower than those by other approaches using only eyeglasses.

Unlike previous approaches, MyDJ overcomes the aforementioned limitations by achieving both accurate and energy-efficient eating detection with a new sensing design on eyeglasses. We also believe the design of MyDJ as an attachable to eyeglass frames enables comfortable usage and deployability.
3 MYDJ DESIGN

We present the design overview of MyDJ (Figure 2). We first illustrate the hardware and sensing design of MyDJ that reports the raw data signals from sensors when a user is chewing. We then describe each component of the eating detection framework that processes the sensor data in real-time. Finally, we demonstrate the implementation of our prototype, built on a custom printed circuit board (PCB).

3.1 Design Goals

Our design has the following goals:

- **Design a sensing system that could easily integrate with existing eyeglass frames:** As users select different eyeglass frames based on their own style and need [15, 35], a desired eating detection system should be available for most existing frames. It should not require specific materials or the shape of an eyeglass frame to sense food intake.
- **Use low-power sensors and optimized data processing pipeline for energy-efficiency:** Energy efficiency of a wearable eating detection system is a critical factor for its usability [13]. The system should utilize low-power sensors and effectively minimize its computational overhead in sensor data processing (e.g., feature selection).
- **Capture complementary eating-related signals for robust and accurate sensing:** Eating detection systems should be robust to different users or environmental changes [66]. Multimodal sensing with different signals enables the system to work even when one source of the signal is weak [61].
- **Place sensors for accurate sensing without sacrificing user comfort:** An eating detection wearable should be comfortable to wear [79]. Sensor placements should not cause user discomfort to achieve accurate sensing (e.g., in-ear canal [3]). We aim to place sensors that achieve both high accuracy and user comfort.

3.2 Overview of Hardware and Sensing

We use a piezoelectric sensor and an accelerometer on MyDJ, which operates in relatively low-power than other transducers [50, 69]. We designed each sensor to capture two complementary chewing signals, as shown in Figure 3. Note that these are not the only eating-related signals available on eyeglasses; other signals such as the chewing sound of mastication muscle activation could also be leveraged [3, 83]. However, we do not consider other signals as robust sensing of such signals is limited in the presence of loud background noises or requires more power-consuming sensors [6].

3.2.1 Piezoelectric Sensor. We use a piezoelectric sensor to capture the temporalis muscle contraction (Figure 3a) that elevates the mandible (the lower jaw) on chewing. This muscle contraction generates huge mechanical dynamics on its skin, which is easily noticeable even with our fingers.

Piezoelectric sensors require firm contact with human skin for better sensing quality. Previous studies used form factors such as headgear [13] and headband [12] or used medical tape [28] to attach the sensor. We chose a novel design of piezoelectric sensor placement on eyeglasses that is comfortable and achieves accurate sensing. We place the sensor on the inner side of the eyeglass frame near the ear, where the sensor’s contact with human skin could be naturally provided. Figure 1b visualizes the placement of the piezoelectric sensor on our prototype. This location near the ear is where the temporalis muscle is located beneath the skin. Our sensor placement could monitor chewing activity without causing user discomfort. Our design is also readily applicable at most eyeglass frames, as eyeglasses are commonly designed to be placed on ears.

3.2.2 Accelerometer. We use an accelerometer to capture the propagation of mechanical vibrations (Figure 3b), which occurs when chewing food — primarily caused by the crunched food and by clenching the teeth. These mechanical waves propagate onto the eyeglass frame via the locations near the nose and the ear, where the eyeglasses are in contact with the human skin. We placed the sensor on the eyeglasses temple, which is shown in Figures 1a and 1b. As we place the small sensor on an eyeglasses temple, it is
MyDJ
Eating Detection Framework
Piezoelectric
Sensor Data Pre-
processing
Feature 
Extraction
Data Pre-
processing
Feature 
Extraction
DNN Classification
Accel.
Features
Piezo.
Features
Model
Output
Accelerometer

Figure 2: System Overview of MyDJ.

Figure 3: FL on two datasets with different deadline configu-
ration methods: (a) Temporalis contraction. (b) Mechanical 
waves propagation.

comfortable and easy to wear and could be integrated into various 
eyeglass frames through simple adjustments.

3.2.3 Hardware Implementation. Our prototype is powered by 
nRF52840, a 32-bit ARM Cortex-M4 MicroController Unit (MCU) by 
Nordic Semiconductor, with a floating-point unit running at 64MHz, 
1MB flash memory, and 256kB RAM. The MCU is connected to 
LDT0-028K piezoelectric sensor and ADXL313 3-axis accelerom-
eter via a 12-bit Analog-to-Digital Conversion (ADC) interface 
and I2C serial communication interface, respectively. We used an 
MDBT50Q-P1M module that encapsulates the MCU connected to 
a trace antenna for Bluetooth Low Energy (BLE) communication 
and mounted it on our custom PCB (1.6 cm × 4.1 cm), alongside 
with ADP3301 3.3V regulator, micro-USB connector, and micro SD 
card connector. Figure 1 shows the example of a custom PCB being 
attached to a commodity eyeglass frame with the 220mAh battery 
inside a plastic housing.

3.3 Combining Raw Signals from Two Sensors

Using our MyDJ prototype, we capture the raw sensor responses 
from both sensors and illustrate how each sensor captures unique as-
pects of the chewing signals when eating. We also explore whether 
the sensor responses of eating could be distinguished from other 
human activities. The sensor responses discussed in this subsection 
used 256Hz and 400Hz sampling rates for a piezoelectric sensor and 
an accelerometer sensor, respectively. We also use these sampling 

rates in our eating detection framework in Section 3.4, which are 
chosen to be minimal but sufficient to capture chewing signals after 
multiple iterations of testing on different configurations.

First, we observe the raw time-domain signals on both sensors 
when a user is chewing. We provide the raw data in two cases where 
1) a user is chewing without head motion and 2) a user is chewing 
and moving his head. The purpose of performing the second case 
is to assess both sensors’ stability on MyDJ while a user is freely 
moving one’s head while eating in a natural setting. In this scenario, 
the person with MyDJ horizontally shakes the head at 0.5Hz while 
chewing. Figure 4 shows the raw sensor responses of both sensors 
for both cases.

We observe that each sensor captures unique signals distin-
guished from the other. The sensor response in Figure 4a shows 
clear peaks at chewing by both sensors. The boxes in the figure 
indicate the sensor-specific patterns of peaks observed with each 
chewing activity. The piezoelectric sensor response shows a pattern 
of low-frequency peaks of relatively long duration that starts with 
the Jaw Elevation Start (JES) and ends after the Jaw Elevation End 
(JEE). Such response pattern around JES and JEE matches with 
the temporalis activity at the chewing cycle. On the other hand, 
the accelerometer response shows high-frequency peaks of short 
duration after the JEE. These high-frequency peaks are generated 
from the interference of multiple mechanical waves propagated 
from the chewing impulse.

When a user is chewing while moving his head (Figure 4b), 
similar patterns are visible for both sensors, which indicates that 
our design with these two sensors is robust to the motion noise. 
Note that the y-axis range of Figure 4b is wider than in Figure 4a.

In Figure 5, we visualize both sensors’ responses with a sequence 
of various human activities; walking, being stationary, eating, and 
talking. It is shown as a time-domain raw data (top) and as a spectro-
gram that shows the frequency domain response over time (bottom). 
From both sensors’ responses, eating is distinguished from other 
activities in both the time and frequency domains. While walking 
and eating seem to have similar high-frequency peaks at the time-
domain response of an accelerometer, their spectrogram response 
shows distinct patterns, especially at the frequency range over 
50Hz. The spectrogram of eating and talking in the piezoelectric 
sensor shows distinct patterns, even though both activities include 
jaw movements. This distinction stems from the fact that eating 
involves more regular and intensive jaw movements than talking.
Figure 4: Raw data signals from two sensors when chewing. The bottom images in Figure 4a are from the video recorded during the experiment for demonstration. JPS, JES, JEE, and HMS stands for Jaw Protraction Start, Jaw Elevation Start, Jaw Elevation End, and Head Moving Start. Boxes indicate the distinctive signal patterns of each sensor that appears with chewing. The piezoelectric sensor shows low-frequency peaks of longer duration while the accelerometer shows short high-frequency peaks after JEE, which indicates that these two sensors are sensing the different sources of signals from chewing activities.

Figure 5: Time/frequency domain sensor responses at the sequence of different human activities.

3.4 Eating Detection Framework
3.4.1 Data Preprocessing. A piezoelectric sensor and an accelerometer equipped on MyDJ continuously generate raw data stream at 256Hz and 400Hz sampling rates, respectively. As we use a 3-axis accelerometer that outputs X, Y, and Z values per sample, we calculate the root sum squared of each value and use it as an aggregated acceleration of the sample. Both streams of input data are segmented into non-overlapping windows of three seconds, which was chosen based on the previous work [12] that experimented with varying window sizes for eating detection. We use the same size of windows for both sensors to extract features from the same window and determine whether a user is eating at the window, which is further aggregated to detect eating episodes as illustrated in Figure 3.4.3.

3.4.2 Feature Extraction and Selection. We apply feature extraction and use extracted features as an input to the classification model. While recent sensor-based applications use complex neural layers (e.g., CNN, autoencoders) for artificial feature engineering [59, 70], we use extracted features to minimize the power and memory consumption. Note that we run our eating detection framework on board of MyDJ to preserve user privacy without transmitting the raw data externally.

By applying reflection padding at both ends, a three-second window is divided into 24 frames each, with 75% overlapping and a one-second duration. On each frame, we extract three types of frame-level features: Short-time Fourier transform (STFT), Mel-Frequency Cepstral Coefficients (MFCCs), and Root-Mean-Square (RMS). We chose STFT since the spectrogram response of eating is visually differentiated from other human activities, as shown in Figure 5. The MFCCs, which are widely used in automatic speech recognition systems, were selected as they apply a discrete cosine transform at mel-scale filter banks that mimic the human ear perception of sound [32, 71]. STFT and MFCCs use the sampling rate of each sensor as the number of FFT points. Lastly, we used RMS to capture the mean power magnitude of the input signal.

Once we extract the features, we further statistically aggregate frame-level features on each frequency to generate window-level features as in BodyBeat [68]. In total, we extracted 1,500 features (1,290 for STFT, 200 for MFCCs, and 10 for RMS) from each sensor.
We used Joint Mutual Information Maximisation (JMIM) [11] for feature selection, which finds the optimal set of features that has maximal Mutual Information (MI) with the label. According to the benchmark study of various feature selection methods [14], JMIM achieves the best median accuracy among all filter methods with the selected set of features when it is tested on 16 large classification datasets.

### 3.4.3 Window Classification and Eating Episode Detection

For the classification of each three-second window, we used a fully-connected Deep Neural network (DNN) classifier as in previous approaches with the same type of sensors [26, 46]. While some studies [13, 83] used lightweight linear classifiers such as Logistic Regression (LR) or Linear Discriminant Analysis (LDA), we used DNN as it shows superior performance in various domains [4], including eating detection [24, 31], by capturing nonlinear separation within data. Other approaches [7, 8] used ensemble classifiers such as Random Forests (RF), but their configuration with 100 trees might result in large memory footprint (∼700KB) on wearable microcontrollers. We used only one hidden layer with 50 hidden nodes for our DNN, which consumes only 10.61KB of memory.

As the window-level classifier could output false-positive results, MyDJ determines a user is eating based on the detection of eating episodes. MyDJ detects eating episodes with a similar strategy from a previous work [13]. Five consecutive three-second windows are used to detect Eating Episode Frame (EEF) of 15 seconds, as shown in Figure 6a. If there are more than two windows classified as eating, the five windows are aggregated as EEF. Once the EEF is detected, MyDJ determines that a user is eating.

We use 15 seconds for EEF to capture short instances of eating, such as snacking, instead of 1 minute used by the previous work [13]. The rationale of using 15 seconds is to use multiple three-second windows and to detect eating episodes that contain only one chewing episode, which has 13 seconds of mean duration [65].

EEF becomes a building block of longer eating episodes, and Figure 6b shows how multiple EEFs are aggregated into longer eating episodes. Aggregated EEFs with an interval shorter than 15 minutes are considered as a single eating episode, as defined in Section 2.1.

### 3.4.4 Software Implementation

We used Python toolkit librosa [55] for the extraction of all features and analysis on the server and used R package praznik [42] for JMIM feature selection. We implemented the DNN classifier with PyTorch [64] for the evaluation in Sections 5.2 and 5.3. For the evaluation of MyDJ’s power consumption in Section 5.4, we implemented the extraction of selected features on our prototype in embedded C with nRF5 SDK provided by the Nordic Semiconductor. We implemented DNN inference on our prototype using matrix multiplication from CMSIS DSP Software Library [49] that comes with the ARM Cortex processors.

### 4 DATA COLLECTION

We conducted two IRB-approved data collection studies with different lengths and user constraints. From the first study, we collect our training data where users’ behavior is precisely captured with the camera. Since the primary goal of MyDJ is long-term real-world usability, we collect the training data from the outside-the-lab environments for a day-long period. With this data, we train the classification models for MyDJ to evaluate the accuracy of MyDJ for a longer duration. To this end, we perform the second study where we collect week-long data where more realistic and diverse user behavior is captured without the camera. We interviewed the participants of both studies on the experiences of wearing MyDJ to assess the usability of the device.

#### 4.1 Day-long Data Collection with Ground-truth Collection Camera

We recruited 24 participants (13 males; 11 females; aged 20-46). Twenty-one were university students and the rest were a nurse, a homemaker, and an office worker. Nine users wear eyeglasses daily, 13 had worn eyeglasses in the past but were no longer wearing (i.e., got Lasik operations), and two had no experience of daily wearing eyeglasses. Each participant participated for a day and got compensated $50. Eighteen participants collected data on a weekday, while 6 participants collected on a weekend. On the day of the study, each participant visited the laboratory in the morning to get equipped with MyDJ-attached eyeglasses and left to collect data throughout the day. We used one type of commodity eyeglass frames to attach MyDJ for this study, which is shown in Figure 1. Note that we asked the participants who wear eyeglasses on their daily lives to instead wear contact lenses and then wear the MyDJ-attached eyeglasses that we provide during the study. Participants were encouraged to do any activity of their choice, including their regular daily routine. Participants were allowed to take off eyeglasses when they were in a situation that required it (e.g., swimming), but we asked them to limit such time to a maximum of two hours. They returned to the lab in the evening to return the device and be interviewed for the experience of wearing MyDJ.

A total of 237 hours of data (9.88 hours on average per participant) with 94 eating episodes (48 meals and 46 snacks) were collected. Participants consumed various types of food, including meat (pork, beef, and chicken), sandwiches, fried rice, hamburgers, noodles, tonkatsu (pork cutlet), pizza, salad, cake, chocolate chips, etc. The data also include diverse non-eating activities such as brushing their teeth, riding a bicycle, driving, cooking, washing dishes, attending a conference, playing drums, exercising in a gym, walking with a dog, knitting, etc.

**Ground-truth collection & annotation:** To collect ground-truths, participants were asked to carry a smartphone throughout
the day and record themselves with a front-facing camera. We additionally provided a supplementary phone battery and a portable smartphone flip stand for day-long recording. Figure 7 shows how the camera and the experiment setup were used in the experiment, along with example screenshots. We recorded the video without sound. We asked the participants to record themselves as much as possible, including the eating moments. We allowed them to cover the lens when they were not eating and wanted to avoid the video recording; however, with the participant’s approval, such data were not erased and were simply annotated as non-eating. To assess our camera system’s impact on participants’ data collection during the study, we conducted a survey that asks how participants perceived the camera system and report it in Section 6.4.

To synchronize MyDJ sensors and the video, we asked the participants to tap the temple of the eyeglasses nine times in front of the camera at the beginning of the study. This created a unique sensor signal pattern that allowed us to identify the exact synchronization moment on both the video and the sensor data.

The annotation from the video was manually done by three of the authors. One annotated the entire data, while the remaining two divided the data into halves and annotated each, making two sets of annotated labels. For the labels that conflict between the two sets, each set’s annotators had a discussion session to determine the final label for the data. We annotated the data in every second of the data as one of the following: eating and non-eating. We determined a second as eating if a participant chews food at least once. Otherwise, a second was determined as non-eating. Thus, drinking was not labeled as eating. Furthermore, for each of the three-second windows, we determined a window as eating if any of the three seconds was annotated as eating. Otherwise, the three-second window was annotated as non-eating. We calculated the intercoder reliability using Cohen’s Kappa [45] based on previous study [8]. Our annotation resulted in Kappa (κ) = 0.846, where κ > 0.8 represents almost perfect agreement [56].

### 4.2 Week-long Data Collection with MyDJ Attached on Participants’ Eyeglasses

The goal of this data collection study was to evaluate the long-term accuracy and usability of MyDJ when it is attached to users’ own eyeglasses. We recruited six participants (four males; two females; aged 25-51) who have their own eyeglasses and did not participate in the prior study. Four were university students, and the rest were a homemaker and a lecturer. The specification of participants’ eyeglasses frame and head are shown in Table 1, and the metrics used for the measurement are illustrated in Figure 8. Four users always wear eyeglasses except when sleeping, and two users wear them for few hours a day for specific purposes (e.g., driving, blue-light protection). We asked participants to continuously wear eyeglasses during the study. Each participant participated for seven days and got compensated $150.

On the first morning of the study, each participant came to the laboratory and we attached MyDJ on the participant’s eyeglasses. Participants left to collect data and visited the lab on the evening of the seventh day of the study to return the device and be interviewed. The battery life of MyDJ performing data collection is longer than a day; with average power draw of 22.95mW, it lasts 35.43 hours on a 220mAh battery. Thus, we asked participants to charge MyDJ once a day with a given micro-USB charger before going to bed, to ensure that MyDJ is operating properly during the study. The battery was not replaced during the study.

A total of 477 hours of data (11.34 hours on average per day per participant) with 136 eating episodes (93 meals and 43 snacks)

---

**Table 1: Specification of participants’ eyeglass frame and head.**

<table>
<thead>
<tr>
<th>Participant Id</th>
<th>Image</th>
<th>P25</th>
<th>P26</th>
<th>P27</th>
<th>P28</th>
<th>P29</th>
<th>P30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>Material</td>
<td>Metal+Plastic</td>
<td>Metal</td>
<td>Metal</td>
<td>Metal</td>
<td>Plastic</td>
<td>Plastic</td>
</tr>
<tr>
<td>Weight</td>
<td>21.8 21.4</td>
<td>18.5 14.9</td>
<td>21.4 14.9</td>
<td>14.9  20</td>
<td>33.6 47/129/137</td>
<td>46/135/146</td>
<td>33/128/133</td>
</tr>
<tr>
<td>T2T/E2E</td>
<td>140/113</td>
<td>142/99</td>
<td>132/113</td>
<td>133/96</td>
<td>130/136</td>
<td>158/162</td>
<td></td>
</tr>
<tr>
<td>Head</td>
<td>O2O/HL</td>
<td>160/235</td>
<td>155/215</td>
<td>155/235</td>
<td>165/240</td>
<td>175/245</td>
<td></td>
</tr>
</tbody>
</table>

Units are given in g for weight and mm for length. The weight of MyDJ is 9.7g.

---

![Camera setup.](image1)

![Screenshots from the collected videos.](image2)

**Figure 7:** Example of camera usage for ground-truth label collection from the day-long study with various user activities; eating, working at a desk, conducting a chemical experiment, exercising at a gym, etc.

**Figure 8:** Illustration on the metrics of eyeglass frame and head of participants that are used in Table 1. Obs-to-obs is the straight-line distance between the left and right otobasion superius, which is the point of attachment of eyeglasses near the temporalis muscle [54].
were collected. The data includes multiple eating episodes and activities that are more diverse than the previous study. While participants ate most types of food from the previous study, they also ate new types such as dumplings, squid sashimi, peach, avocado, fish and chips, soba noodles, grilled duck, lotus root, and shrimp. Participants went to their own workplaces where they gave lectures, conducted chemical experiments, etc. They also went to multiple public places and social meetings, where they used various modes of transportation (e.g., driving, riding a bus, subway, or bicycle). One participant even went on a date with MyDJ-attached eyeglasses. We conducted a survey that asks how participants perceived the social acceptance of MyDJ at different places (Section 5.5).

Ground-truth collection & annotation: We removed the smartphone camera from this study to minimize the constraints on the participant’s behavior. Instead, we collected the ground-truth via a mobile messenger app (KakaoTalk [21]) as shown in Figure 9. During the study, the participants were asked to send information to the authors about their food intake in real-time. The information includes (1) start and end time of the food intake at a minute-level, (2) types of food that are being consumed and its image, (3) whether it is snack or meal, and (4) whether a participant is confident about the time of the food intake. The confidence was collected as participants occasionally forgot to send messages in time; participants were instructed to send their best guess on the time of the food intake. The confidence was collected as participants occasionally forgot to send messages in time; participants were instructed to send their best guess on the time of the food intake when they were not confident. Out of 136 eating episodes, 120 were replied as confident in this data collection study, and we excluded non-confident eating episodes from the evaluation.

5 EVALUATION
We evaluate MyDJ to answer the following key questions: 1) How accurate is MyDJ in eating detection? 2) How much power does MyDJ consume? 3) How is the user experience of wearing MyDJ?

5.1 Experiment Settings & Procedures
We preprocessed the collected data and trained the eating detection model to assess if MyDJ performs well in the wild. We evaluate the eating detection accuracy of MyDJ in both studies by comparing different types of input features on the metrics as follows.

5.1.1 Input Feature Types. To understand the effectiveness of fusing two different types of sensors on MyDJ, we show our evaluation results on each following input feature types: Piezo, Accel, and Combined. Piezo uses only the features from the piezoelectric sensor on MyDJ, Accel uses only the features from the accelerometer sensor on MyDJ, and Combined uses the input features from both sensors, representing the performance of our design with MyDJ.

5.1.2 Evaluation Metrics. We use the following metrics to evaluate the eating detection accuracy of MyDJ:

• **Undetected eating episodes / False alarms:** We count the number of ground-truth eating episodes without true positives as undetected eating episodes to evaluate how often eating episodes MyDJ would miss. Moreover, we count the number of detected eating episodes without true positives as false alarms to evaluate how often false alarms MyDJ would trigger. We also provide analysis on each occurrence of undetected eating episodes and false alarms to better understand under what circumstances MyDJ works and fails.

• **Coverage ratio / Duration difference / Delay:** We additionally measure the metrics that were widely adopted in other eating detection approaches [7, 8, 13]. The coverage ratio is defined as the percentage of the correctly recognized duration of an eating episode. Note that the recall and the coverage ratio are equivalent on an eating episode, but we report the recall by averaging it per person, while we report the coverage ratio by averaging it per episode. The duration difference is defined as the absolute duration difference between an eating episode and corresponding detected eating episode, and the delay is defined as the elapsed time from the beginning of an eating episode which the system starts to detect it.

5.1.3 Method. For each input feature type, we processed the data and extracted features following the method presented in Section 3.4. For the day-long study, we applied the Leave-One-User-Out (LOUO) methodology to study the performance of MyDJ when deployed to a new user. We divided the 24 users into one test user and 23 training users and performed feature selection on the 23 training users. From the feature selection result, we used top-K selected features for training a DNN for window-level classification. DNN was trained with 23 training users, and we explored the impact of K from the following list: 5, 10, 20, 50, 100, 500, and 1,500. The motivation for using different numbers of selected features is to assess MyDJ’s accuracy with a smaller number of features, as the number of input features is a crucial factor for the power and memory consumption of the system. We split the test user’s data into two chunks by dividing it in half without shuffling, which ensures at least one meal to be included in each chunk. We tested each test user twice and averaged the results, with configuring one chunk as the validation data and the other as the test data, and vice versa. After training the window-level classifier, we infer the window-level label of the test data chunk and perform eating episode detection based on the method described in Section 3.4.3. We repeat the above process 24 times with configuring each of the 24 users as a test user and report the average accuracy and per-user accuracy. For each training process, the model was trained for 50 epochs with a learning rate of 0.001.

For the week-long study, we utilize a pre-trained model that is trained while evaluating the day-long study to infer the window-level label and perform eating episode detection. We chose a model for each Piezo, Accel, and Combined, which have shown the highest F1-score on window-level classification with each input feature type on the day-long study dataset. Moreover, we perform fine-tuning on the pre-trained model of Combined to generate a personalized model on each user to validate if MyDJ performs better when trained with the target user data. From seven days of data on a
user, we use the data from a single day to fine-tune the pre-trained model and evaluate on the remaining six days, which we repeat the process seven times for each user with utilizing each day for fine-tuning. To generate the personalized model, the pre-trained model was additionally trained with smaller epochs (=5) and learning rate (=0.00001), as common practice for fine-tuning [47]. From the evaluation, we exclude eating episodes in which participants did not reply “confident” for its start and end times.

5.2 Day-long Study Results
Figure 11 shows the performance of MyDJ with a different number of selected input features from the day-long study. Figure 11a shows the average F1-score in eating episode coverage. The Combined outperforms both Piezo and Accel by up to 0.136 in the mean F1-score. The mean F1-score of Combined stays consistently high even with decreasing number of input features, with >0.890 mean F1-score in eating episode coverage at all number of selected features. In contrast, with a number of features less than 100, a mean F1-score of <0.840 could be achieved using only a single sensor. Compared with the 1,500 features case, Combined could achieve a 99.7% reduction of input feature space with five features, sacrificing only 0.037 in mean F1-score. The Accel shows a comparably high mean F1-score at 500 selected features, but the usage of both sensors in Combined produces high performance with a low number of selected features, which is critical in maintaining low computation and memory consumption of the system.

Figure 11b and 11c each depicts the averaged counts of undetected eating episodes and false alarms from 24 participants on a different number of selected features. Combined shows fewer undetected eating episodes than Piezo and Accel at all number of selected features. Combined maintains ≤3 over all number of selected features while Piezo and Accel yield as high as 22 and 11 undetected eating episodes. This result suggests that the fusion of both sensors reduces occurrences of undetected eating episodes. For false alarms, Accel shows the least count with features less than 20, while Combined shows the least count otherwise. We suspect that Accel with a small number of features reliably classifies non-eating data but also classifies some borderline eating cases as non-eating, as it has higher precision than recall (e.g., 0.901 vs. 0.831 with 15 features).

Based on our experiments, we recommend to use 50 input features for MyDJ. While decreasing the number of input features would reduce computation overhead, we also aim to achieve high eating episode coverage with fewer undetected eating episodes and false alarms. We chose 50 input features as MyDJ achieves 0.919 F1-score in eating episode coverage while it drops to 0.917 and 0.916 with 20 and 100 input features, respectively. MyDJ achieves >0.920 F1-score with 500≥ features, but we use 50 features as it requires 10x feature processing for only 0.005 F1-score improvement. 50 input features also yield the lowest false alarms and comparably low undetected eating episodes (≤2). Our recommended model for MyDJ achieves 0.984 accuracy, 0.919 F1-score, 0.923 precision, and 0.925 recall in eating episodes coverage while detecting 92 out of 94 eating episodes only with 12 false alarms from the day-long experiment with 24 participants. It also achieves a 92.0% coverage ratio and 121.4 seconds of duration difference for each eating episode, with 11.9 seconds of delay in detecting the beginning of the episode. Note that it shows a 0.794 F1-score in window-level classification, which is comparable with a state-of-the-art system [13] that performs window-based eating detection.

Two undetected eating episodes were less than a minute long, where participants had a mini Oreo and tapioca balls in the bubble tea. We suspect MyDJ could not detect the short episodes with these small-sized snacks, as it is primarily trained on meal data. We found that false alarms mostly occurred in unusual circumstances, such as when a participant fiercely scratched her head or walked unsteadily with irregular steps. 10 out of 12 false alarms have a duration of 15 seconds, which is a length of an eating episode frame. Ones with longer durations (164 and 494 seconds) happened when a participant was exercising in a gym or walking outside. We expect such occurrences will be reduced when the MyDJ is trained with a larger and more diverse set of real-world data.

**Table 1.** The prediction results of MyDJ on 24 participants.

<table>
<thead>
<tr>
<th>ID</th>
<th>Day</th>
<th>Time</th>
<th>confidence</th>
<th>Type</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>P27</td>
<td>Aug 18th</td>
<td>22:07-22:40</td>
<td>False</td>
<td>Snack</td>
<td>ice cream, cookie</td>
</tr>
<tr>
<td>P29</td>
<td>Aug 21th</td>
<td>09:10-09:43</td>
<td>True</td>
<td>Meal</td>
<td>salmon, salad, onion, avocado, kimchi, rice, apple</td>
</tr>
</tbody>
</table>
In Figure 12, we report the per-participant F1-score of the eating episode coverage. In both cases, we observe that Combined consistently shows higher F1-score for every participant compared with Accel and Piezo. The lowest F1-score of Combined among all participants is 0.671, which outperforms the lowest F1-score of Accel (0.221, P23) and Piezo (0.0, P18). Our results suggest that eating detection systems with a single accelerometer on glasses could result in low accuracy for some users — such as P3, P15, and P23 in our experiment. Some users (P12, P18) show extremely low performance with Piezo; we suspect that their eating data on the piezoelectric sensor is distinct from others, as the classifier trained on other users classifies most of the eating data as non-eating. For participants P1, P7, and P23 who had low Combined F1-score (<0.8), we suspect that their chewing styles were different from the majority of users. We believe their performances could be further improved by model personalization as we introduce in Section 5.3. In summary, Combined (i.e., MyDJ) consistently provide the highest F1-score on each participant for most of the cases (14 out of 24), providing at least >0.671 F1-score on all users even if the single-sensor approach outperforms.

### 5.3 Week-long Study Results

Figure 13 shows the per-participant performance of MyDJ from the week-long study. For each participant, Figure 13a depicts the F1-score in eating episode coverage, while Figures 13b and 13c show the number of the undetected eating episodes and the false alarms, respectively. Between the pre-trained models, Combined achieves the highest mean F1-score of 0.777, while Piezo and Accel achieve 0.641 and 0.651, respectively. The lowest F1-score of Combined among all participants is 0.626 (P30), which outperforms the lowest F1-score of Piezo (P29, 0.381) and Accel (P26, 0.185). In addition, the mean coverage ratio of Combined is 0.894, which outperforms Piezo (0.703) and Accel (0.703). This result suggests that multimodal sensing of Combined provides a high F1-score and coverage ratio on most users with different types of eyeglass frames.

Moreover, the total count of undetected eating episodes of Combined is nine, which is less than Piezo (26) and Accel (27). Every undetected eating episode from the pre-trained Combined model were snacking episodes, except for one meal episode, which was only three minutes long. This does not mean that MyDJ cannot detect short eating episodes; 12 of 17 eating episodes that are less than or equal to three minutes long were detected. Participants were having eggs, banana, peach, ice cream, grapes, or fried onion when eating episodes were undetected; however, MyDJ detected other eating episodes with these foods (e.g., MyDJ detected a participant having ice cream for seven minutes). The total count of false alarms of Combined is 91, which is more than Accel (57) but less than Piezo (132). While Accel results in the least count of false alarms, it results in more false alarms than other input types on one user (P26). Combined consistently shows ≤23 false alarms on each user, with less undetected eating episodes than single-sensor approaches on all users. Among the false alarms on the pre-trained model of Combined, 63 out of 91 were less than a minute long. The longest false alarm was 33 minutes and 44 seconds long. As we did not use a camera at the week-long study, we asked participants what they were doing when the false alarm happened. For the long false alarms, participants replied that they were wearing a headphone, a VR headset, or a safety goggle, which might have physically adhered...
with MyDJ and affected the sensors. Participants also replied that they were working out at a gym when false alarms were detected. We expect these false alarms could be resolved when the model is trained with more diverse real-world data. Some participants replied that they were chewing straws with iced drinks at the false alarm moments, which indicates that MyDJ detected chewing as designed for such occasions.

The pre-trained model of Combined further improves with the fine-tuning process. The mean F1-score increases by 0.072 on the personalized model (0.777 to 0.849), with the lowest F1-score increasing by 0.154 (0.626 to 0.780). While there exist users (P26 and P27) whose F1-score decreases with the personalized model, their F1-score drop (0.002 and 0.025) is significantly lower compared with the user who gained the most F1-score after fine-tuning (P30, 0.229). Moreover, the personalized model results in 12.4 undetected eating episodes and 37.4 false alarms on average of seven iterations of fine-tuning. The count of false alarms decreased more than twice after fine-tuning on most users, while the number of undetected eating episodes only increased by 0.5 per user. The mean coverage ratio decreased after fine-tuning (89.4% to 84.5%), but it still shows ≥72.6% coverage ratio on all users. In a nutshell, the personalized model improves the F1-score and reduces false alarms, especially on users with low performance on the pre-trained model. As generating the personalized model requires a user to provide labels as in our week-long study, we interviewed the participants on their experience of labeling and further discuss it in Section 6.3.

5.4 Power Consumption

We used Monsoon Power Monitor (FTA22D) [36] to measure the power draw of our prototype on performing real-time eating detection. We divided the functionality of MyDJ into three categories: raw data sensing, feature extraction, and classification. Starting with the idle state of the device that does not perform any functionality, we added each functionality one by one to measure the power consumption of each. For each measurement, we measured the power draw by averaging the results from five minutes of execution. We used a 3.7V voltage supply of the power monitor.

The results reported in Table 2 show that the total power draw of MyDJ is 26.06mW, which results in 66 hours and 38 minutes of operating time with the same battery of recently released commodity smart eyeglasses (470mAh, Vuzix Blade Upgraded [80]). This is 4.03× longer battery time compared with the reported battery time of the state-of-the-art eating detection system on glasses [7]. We observe that most of the power was drawn by the raw data sensing with two sensors. We also note that the power consumption of the classification is noticeably small. We conjecture that this is due to the small size of the neural network (50×50×2 nodes) and the power-efficient matrix multiplication APIs of ARM.

5.5 User Experience Survey

To assess the comfort level of wearing MyDJ in uncontrolled environments, we first asked two questions to the 24 subjects (P1 - P24) who participated in the day-long study: (Q1) How convenient was it to put on the wearable device? (Q2) Would you wear this device in your daily life? Questions were answered with a 0-10 point scale, where 10 indicates the highest comfort. Moreover, to further understand the user experience of wearing MyDJ at a long term, we surveyed six subjects (P25 - P30) who participated in the week-long study. We asked how much do participants agree with the following statements, where each describes a different impression of wearing a device in their daily life: (Q3) I could wear this device for more than a week in my daily routine. (Q4) Wearing this device made it difficult to carry out my daily life. (Q5) I do not feel secure and safe wearing the device in my daily life. Participants answered each statement with a 0-10 point scale, where 0 indicates “strongly disagree” and 10 indicates “strongly agree”. In addition, we
interviewed the six users’ experience of wearing MyDJ in the long term, in terms of its social acceptance and frame weight unbalance. We additionally asked Q1, Q2, and Q4 on regular eyeglasses to understand the participants’ personal preference on eyeglasses-type wearables. For the questions on regular eyeglasses, participants answered based on their personal experiences or perceptions of wearing eyeglasses.

Comfort: Figure 14a shows the box plot graphs of the survey scores. For Q1, MyDJ scores 6.58 ± 2.15 and the regular eyeglasses score 7.13 ± 1.90 from 24 participants. For Q2, MyDJ scores 6.54 ± 2.45 and the regular eyeglasses score 6.70 ± 2.42. Moreover, we conducted a paired t-test (α=0.05) on Q1 and Q2, as both results are found to be following normal distribution on the Shapiro–Wilk test [73]. For Q1, the p-value of MyDJ is 0.02009, indicating that both results are statistically distinguished and the regular eyeglasses are found to be more comfortable than MyDJ for our participants. For Q2, however, MyDJ shows a p-value of 0.76, where participants are willing to wear MyDJ at a similar level compared with the regular eyeglasses. We noticed that there were users who were not willing to wear MyDJ, regardless of its functionality. Nine participants who wear eyeglasses daily gave a higher score to MyDJ (Q1: 6.78 ± 1.64, Q2: 6.89 ± 2.09) than others (Q1: 6.46 ± 2.44, Q2: 6.33 ± 2.69), which indicates that users who do not usually wear eyeglasses prefer less to adopt MyDJ. However, most users replied that the overall experience of wearing MyDJ was comfortable, that MyDJ scored 94.95% of the regular eyeglasses on average: “I found no difference of wearing MyDJ with wearing regular eyeglasses. As I’m wearing eyeglasses daily, I would be happy to wear them with some additional functionalities.” (P5, P8).

Long term wearability of MyDJ: Figure 14b shows the box plot graphs of the Q3-Q5 scores. For Q3, participants gave a score of 8.33 ± 2.25, where they mostly agreed to wear MyDJ for more than a week. We additionally asked the maximum duration that each user could wear MyDJ. Four out of six participants replied that they could continuously wear MyDJ, and one other participant replied “1 year”. A remaining participant who replied “1 week” explained that the difficulties with the current MyDJ prototype: “As the device externally exposes the circuit board and the wires, I was worried that the device could be broken when I played sports or got caught in the rain.” (P29). We believe that this problem could be easily handled by encapsulating and protecting each component of MyDJ in future prototypes. While our study lasted for only a week, this result suggests that MyDJ with durable prototype could be accepted on most users for more longitudinal study.

For Q4, MyDJ (3.5 ± 2.43) and the regular eyeglasses (1.67 ± 1.03) score below 5 on average, indicating that both wears do not significantly disturb users in their daily routine. A participant who gave the largest score difference (MyDJ 8, regular eyeglasses 3) implied that it is due to the pain from the prototype’s piezoelectric sensor film: “I could continuously feel the sharp edges of the film sensor on my skin” (P30). Again, we expect this would be resolved in the next prototypes of MyDJ by switching the sensor with a soft and stretchable design [76] or packaging it with soft silicon-based protection [16]. We also asked whether the participants changed their schedule because of the fact that they are wearing MyDJ, and all the participants replied they were able to go through a week as scheduled: “The eyeglasses feel mostly the same with and without the attachable device, and there was nothing that I couldn’t do because of the device.” (P25, P27).

For Q5, participants gave a score of 1.00 ± 1.10, where most of them felt safe and secure while wearing MyDJ. One participant mentioned that the narrowed-down vision could become a problem: “The battery part of the device blocked my sight, and that felt slightly risky when I was working out or jogging.” (P25). We believe that this could be resolved by placing the battery at different locations (e.g., on the circuit board) or using a narrower battery that fits the width of the eyeglass frame.

Social acceptability: To understand how the physical appearance of MyDJ impacted the users in the long term, we asked participants if they were worried about how they look with MyDJ. Four out of six participants replied that they were completely fine with their appearance while wearing MyDJ. Other two participants replied that they felt unnatural to put an additional device on the eyeglasses: “I felt that the device attachment on my eyeglasses made them look different from ordinary eyeglasses, and this made me worry about what others would think.” (P28, P29). However, both participants reported that the main problem arose from the appearance of the current prototype, where a circuit board and wires are exposed. They both agreed that the problem would be resolved if the next prototype of MyDJ has electronics housing. Moreover, a participant mentioned that wearing MyDJ would be much easier if it becomes a mainstream wearable: “We all laughed at Airpods when it first came out, but we all wear it now. Just like that, I think wearing this device will no longer make me nervous when many others are wearing it.” (P29).

Additionally, we asked participants how the others reacted when they wore MyDJ. Most of the participants replied that they were frequently asked multiple times about what the device was, but they haven’t received any negative comments about the device.

<table>
<thead>
<tr>
<th>Functionalities</th>
<th>Average Power Draw(mW)</th>
<th>Battery Life1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle</td>
<td>10.76</td>
<td>161 hrs 20 min</td>
</tr>
<tr>
<td>Raw data sensing with two sensors</td>
<td>+8.73</td>
<td>89 hrs 11 min</td>
</tr>
<tr>
<td>Feature extraction</td>
<td>+6.38</td>
<td>67 hrs 7 min</td>
</tr>
<tr>
<td>Classification</td>
<td>+0.19</td>
<td>66 hrs 38 min</td>
</tr>
<tr>
<td>Total</td>
<td>26.06</td>
<td>66 hrs 38 min</td>
</tr>
</tbody>
</table>

1 Battery life is calculated based on the 470mAh battery of recent smart eyeglasses, Vuzix Blade Upgraded [80].
Table 3: A comparison with the previous eating detection studies. We compare the following metrics: F1-score, Undetected Eating Episodes (UEE), False Alarms (FA), Power Draw (PD) in mW, Battery Capacity (BC) in mAh, and RunTime (RT) in hours.

<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Wearable</th>
<th>Sensors</th>
<th>F1</th>
<th>UEE</th>
<th>FA</th>
<th>PD</th>
<th>BC</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>Thomaz et al. [78]</td>
<td>smartwatch</td>
<td>S1</td>
<td>0.76</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2016</td>
<td>Farooq et al. [27]</td>
<td>eyeglasses</td>
<td>S4</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2017</td>
<td>Bedri et al. [8]</td>
<td>outer-ear flap</td>
<td>S1-S3</td>
<td>0.80</td>
<td>1 out of 16 (0.06)</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2017</td>
<td>Chung et al. [18]</td>
<td>eyeglasses</td>
<td>S6</td>
<td>0.94</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2018</td>
<td>Bi et al. [13]</td>
<td>headgear</td>
<td>S4</td>
<td>0.78</td>
<td>2 out of 26 (0.08)</td>
<td>12</td>
<td>14.47</td>
<td>110</td>
<td>28.10</td>
</tr>
<tr>
<td>2018</td>
<td>Chun et al. [17]</td>
<td>necklace</td>
<td>S5</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2018</td>
<td>Farooq et al. [29]</td>
<td>eyeglasses</td>
<td>S1</td>
<td>0.86</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2018</td>
<td>Zhang et al. [83]</td>
<td>eyeglasses</td>
<td>S7</td>
<td>&gt; 0.77</td>
<td>1 out of 44 (0.02)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2019</td>
<td>Zhang et al. [84]</td>
<td>necklace</td>
<td>S1-S3, S5, S8</td>
<td>0.77</td>
<td>13 out of 76 (0.17)</td>
<td>-</td>
<td>81.96</td>
<td>350</td>
<td>15.80</td>
</tr>
<tr>
<td>2020</td>
<td>Bedri et al. [72]</td>
<td>eyeglasses</td>
<td>S1, S2, S5</td>
<td>0.89</td>
<td>6 out of 28 (0.21)</td>
<td>4</td>
<td>105.08</td>
<td>900</td>
<td>31.68</td>
</tr>
<tr>
<td>2022</td>
<td><strong>MyDJ</strong></td>
<td>eyeglass attachable</td>
<td>S1, S4</td>
<td>0.92</td>
<td>2 out of 94 (0.02)</td>
<td>12</td>
<td>26.06</td>
<td>220</td>
<td>27.83</td>
</tr>
</tbody>
</table>

1 S1-accelerometer, S2-gyroscope, S3-magnetometer, S4-piezo, S5-proximity, S6-load cell, S7-EMG, S8-light
2 Power draw values are calculated assuming the 3.7V powered system.
3 These studies tried multiple combination of wearables and sensors, and here we report what they recommended for the real-world usage.
4 These studies report the F1-score from the in-lab study.

Some participants even replied that the people they meet daily (e.g., family, friends, colleagues, etc.) mostly haven’t noticed a difference while MyDJ was attached: “My family did not know the presence of the device for days until I explain it first.” (P27). “Most of my colleagues did not notice it at first glance. Later, some of them asked me if I have got a new pair of eyeglasses.” (P25, P30). One participant mentioned that others questioned if MyDJ contains a camera: “Some of my colleagues asked if the device is recording video. After I explained the type of sensors on the device and how it worked, they were fine with it.” (P28). This indicates that the device could be initially viewed privacy invasive, but our MyDJ design without such sensors (e.g., camera) make it less concerning.

We asked if it was difficult to go to public space or social meetings with MyDJ. Five of six participants found it not difficult to wear MyDJ in such contexts. It is also shown in what participants reported during the week-long study that they have been to malls, gyms, restaurants, lectures (as a lecturer), and dates while wearing MyDJ. One participant who opposed others replied that it is mainly due to others asking frequently: “I become more nervous as others ask what the device is, and that made me avoid going to the public places.” (P29). We envision this problem to be resolved with smart eyewear being more widespread and common.

**Battery management:** Five of six participants reported that they had no problem with charging MyDJ once a day. A participant mentioned “I always charge my phone and smartwatch before going to bed, and it was no hassle to add one more device.” (P25). A participant with a different opinion from others replied “I usually charge my electronics during daytime, but I had to additionally charge this device while I was asleep, because I cannot see anything while it’s charging.” (P26). The participant further expressed that the problem would be resolved if MyDJ could be easily attached and detached from eyeglasses, which would allow it to be charged during daytime. This is part of our plan in developing the next prototype of MyDJ.

6 DISCUSSION

6.1 Comparison with Previous Methods

MyDJ achieves the highest F1-score among the studies with in-the-wild experiments, with the least ratio of undetected eating episodes to the total eating episodes. Compared with Farooq et al. [29] that used one accelerometer placed on the eyeglasses, MyDJ reports higher F1-score due to the multimodal sensing of MyDJ with additional piezoelectric sensor. Note that a comparison of results reported by each different paper is not ideal, as each study used the data collected from different group of people and environments. Nevertheless, we believe that such comparison gives insights in understanding the performance of MyDJ over prior studies.

MyDJ also shows 4.03× less power consumption than the state-of-the-art system on eyeglasses (Bedri et al. [7]), due to the use of less number of sensors with lower sampling rate (400Hz vs 4kHz on an accelerometer). While Bi et al. [13] achieves the lowest power.
draw, MyDJ achieves less ratio of false alarms to the total eating episodes with a runtime over a day on a 220mAh battery.

### 6.2 Performance on Different Eyeglass Frames

While we have demonstrated the performance of MyDJ on various eyeglass frames in Section 5.3, one might wonder how different eyeglass frames could affect the performance of one person. Three of our lab colleagues performed a pilot study where they wore three types of eyeglass frames with MyDJ for three days, wearing each frame for a day. We chose a frame from three major categories of eyeglass frames [48], which are rimless, semi-rimless, and full-rimmed. None of the three participants were authors. We used the model trained in Section 5.2 to infer the collected data. The results show that MyDJ performs $\geq 0.886$ F1-score in eating episode coverage regardless of the frame type on the participants. The largest F1-score difference between eyeglass frames on each of the three participants were 0.084, 0.095, and 0.038. The semi-rimless eyeglass frame shows the highest F1-score (0.987, 0.981, 0.954) on all participants, while the rimless eyeglass frame shows the lowest F1-score (0.903, 0.886, 0.916) on all participants. We believe the low performance of certain frames could be improved with fine-tuning as in Section 5.3, or using a model that is trained on data from various frames.

### 6.3 User Experience on Providing Labels for the Model Fine-tuning

We observed that the personalized model shows improved performance for a portion of participants in Section 5.3. As it requires the user to manually label their eating episodes for a day, we asked participants if conducting such a task is feasible for better performance. All of the six participants replied that they could provide such labels for one or two days: “It was not really a burden for me, as I usually take a picture of everything I eat.” (P26). However, participants mostly agreed that the labeling process should be eased, rather than entering plain text on a mobile messenger app: “It would be much easier with a fixed-form entry, as it was hard to recall the things that I should record each time.” (P50), and “I expect this process to be challenging for elderly, as they are not all familiar with using keyboards on smartphones.” (P27). Based on the feedback, we believe the model personalization on new MyDJ users could be done, but it requires a simpler method of label entry. We leave this as our next step of research.

### 6.4 Impact of the Ground-Truth Collection on Smartphone Camera

We conducted a qualitative study on participants to assess the impact of our ground-truth collection system on the day-long study. We asked the participants if the smartphone camera system generally affected their movement or eating activity. Few participants replied that their behavior has changed due to the presence of a video-recording smartphone in proximity: “I unusually wiped out my mouth multiple times while eating, as I could see myself from the smartphone screen.” (P24) and “At the beginning of the study, I felt weird because of the feeling of being watched. However, I soon got used to it.” (P12). Nevertheless, most participants replied that they did not feel any change: “I did not care the camera at all while eating.” (P17). We asked the participants if they had changed their daily schedule because of the smartphone camera system, and all participants replied that there was no change in their schedule.

### 6.5 Limitations

**Eyeglass form factor:** One clear limitation of MyDJ is that it cannot support eating detection on people who do not desire to wear eyeglasses. As shown in Section 5.5, there are users who do not want to wear eyeglasses, where one of the primary reasons is the societal perceptions that wearing eyeglasses is unattractive [40]. However, as discussed in Section 2.2, eyeglasses are familiar to more users than other eating detection form factors (e.g., necklaces). We believe that eyeglasses could gain popularity with various smart glasses and AR glasses appearing with numerous functionalities [34, 37, 53, 77, 81, 85].

**Data collection on multiple glass frames:** While MyDJ is designed to be attached to any eyeglass frame, we collected our training data on a single eyeglass frame. While our model successfully detected most of the eating episodes on the participants’ own eyeglasses, we expect our model could be more robust to different eyeglass frames when it is trained on the data from various frames. This is part of our future work.

### 6.6 Use Cases of MyDJ

We expect user application of MyDJ could be helpful for real users in the following ways. First, when MyDJ detects the eating moment of a user in real-time, the system could provide Just-In-Time Adaptive Interventions (JITAI) to provide feedback based on the user’s eating activity. For example, a system could be designed to prevent overeating, providing real-time interventions when a user is spending too much time eating [84]. Second, when a user wears MyDJ long-term, aggregated result of detected eating episodes could be used to provide personalized feedback on the user’s eating activity. For example, if there are a number of detected eating episodes around nighttime, the system could suggest users to have fewer midnight snacks. We believe that designing the appropriate user application with MyDJ could potentially help 1.9 billion and 650 million overweight and obese people worldwide, respectively [63], and 70 million patients with eating disorders [30].

### 7 CONCLUSION

We propose MyDJ, an accurate and energy-efficient eating detection system that could be attached to any eyeglass frame. Our sensing fusion of a piezoelectric sensor and an accelerometer on an eyeglass temple achieves accurate sensing in uncontrolled environments, as each low-power sensor captures the distinct source of chewing signals. We collected a total of 714 hours of data with 30 participants from uncontrolled environments, where six of them attached MyDJ on their eyeglasses for a week. MyDJ reaches 0.984 accuracy and 0.919 F1-score in eating episode detection outside-the-lab, with a 4.03× battery time improvement over the state-of-the-art eating detection glass system. Our survey on the comfort level of wearing MyDJ shows a 94.95% score compared with wearing regular eyeglasses. We believe realizing high accuracy, energy efficiency, and user comfort is the right step toward developing automated eating detection systems in practice.
ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.NRF-2020R1A2C1010462). Moreover, this work was also supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-00990).

REFERENCES


[79] Janet van den Boer, Annemiek van der Lee, Lingchuan Zhou, Vassileios Papa-
panagiotou, Christos Diou, Anastassios Delopoulos, and Monica Mars. 2018. The
SPLENDID eating detection sensor: Development and feasibility study. JMIR
mHealth and uHealth 6, 9 (2018), e170.


[81] Peter Washington, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar
Fazel, Titas De, Carl Feinstein, Terry Winograd, and Dennis Wall. 2017. Super-
powerGlass: a wearable aid for the at-home therapy of children with autism.
Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies
1, 3 (2017), 1–22.

Computing (Pittsburgh, Pennsylvania) (Ubicomp ’12). Association for Computing
2370269

using smart eyeglasses. IEEE journal of biomedical and health informatics
22, 1 (2017), 23–32.

[84] Shibo Zhang, Yuqi Zhao, Dzung Tri Nguyen, Runsheng Xu, Sougata Sen, Josiah
Detecting Eating Activities in Free-Living Conditions. Proceedings of the ACM on

[85] Yuhang Zhao, Elizabeth Kupferstein, Hathairom Rojnirun, Leah Findlater, and
Shiri Azenkot. 2020. The Effectiveness of Visual and Audio Wayfinding Guidance
on Smartglasses for People with Low Vision. In Proceedings of the 2020 CHI
Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI
’20). Association for Computing Machinery, New York, NY, USA, 1–14. https:
//doi.org/10.1145/3313831.3376516

[86] Ya-Li Zheng, Xiao-Rong Ding, Carmen Chung Yan Poon, Benny Ping Lai Lo,
Heye Zhang, Xiao-Lin Zhou, Guang-Zhong Yang, Ni Zhao, and Yuan-Ting Zhang.
2014. Unobtrusive sensing and wearable devices for health informatics. IEEE
Transactions on Biomedical Engineering 61, 5 (2014), 1538–1554.

with Smart Fabric Mechanomyography on the Temporalis. In Proceedings of the
2020 International Symposium on Wearable Computers (Virtual Event, Mexico)
(ISWC ’20). Association for Computing Machinery, New York, NY, USA, 96–100.
https://doi.org/10.1145/3410531.3414305