First Bite/Chew: distinguish different types of food by first biting/chewing and the corresponding hand movement

Junyu Chen*
chenjy@g.ecc.u-tokyo.ac.jp
The University of Tokyo
Japan

Xiongqi Wang*
aidan108wang@gmail.com
Keio University
Japan

Juling Li*
lijuling@mail.ustc.edu.cn
Keio University
Japan

Thad Starner
thadstarner@gmail.com
Georgia Institute of Technology
United States

George Chernyshov
send.letters.for.me.here@gmail.com
Keio University
Japan

Jing Huang
hkoukenj@gmail.com
Keio University
Japan

Yifei Huang
hyf015@gmail.com
The University of Tokyo
Japan

Kai Kunze
kai@kmd.keio.ac.jp
Keio University
Japan

Qing Zhang
qzkiyoshi@gmail.com
Keio University
Japan

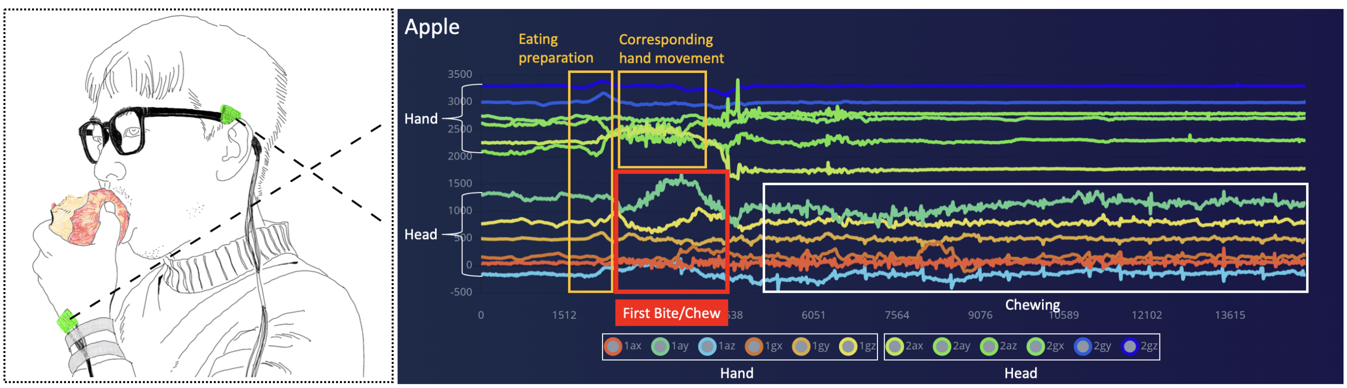


Figure 1: When a participant eats an apple, we can obtain 12 axes IMU data from his eating movements, like the Waveform plot on the right side. The red rectangle is the time window we define as the first bite/chew period, while the thin yellow rectangle corresponds to hand movement during the first bite/chew.

ABSTRACT

Imbalanced food intake contributes to various diseases, such as obesity, diabetes, high blood pressure, high cholesterol, heart disease, and type-2 diabetes. At the same time, food intake monitoring systems play a significant role in the treatment. Most current food intake tracking methods are camera-based, on-body sensor-based, microphone based, and self-reported. The challenges that remain are social acceptance, lightweight, easy to use, and inexpensive. Our method leverages two 6-axe Inertial Measurement Units (IMU) on the glasses' leg and the wrist to detect the user's food intake activities using a machine learning capable Micro Controller Unit (MCU). We introduced the concept of the first bite/chew, which is a stable and reliable indicator to distinguish food types. Our implementation results show that our method can distinguish seven kinds of food at an accuracy of 93.26% (average) over all four participants.

*Three authors contributed equally to this research.

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '23, April 23–28, 2023, Hamburg, Germany

© 2023 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585845

CCS CONCEPTS

• **Human-centered computing → Interaction devices**; **User interface toolkits**; • **Hardware**;

KEYWORDS

smart eyewear, food intake, diet monitoring

ACM Reference Format:
Junyu Chen, Xiongqi Wang, Juling Li, Thad Starner, George Chernyshov, Jing Huang, Yifei Huang, Kai Kunze, and Qing Zhang. 2023. First Bite/Chew:

distinguish different types of food by first biting/chewing and the corresponding hand movement. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585845

1 INTRODUCTION

Nutritionists and physicians often rely on food journalling to help diagnose diabetes, high cholesterol, high blood pressure, heart disease, and obesity. However, manual food intake estimation methods can easily be incorrect by 50% [14]. Paper-based journalling methods [4] require significant time and effort and are prone to be forgotten or abandoned [10]. Especially the accuracy of food intake is highly dependent on the effort of the users [14]. Therefore, several automatic food intake monitoring systems have been proposed, most of them either utilizing microphones to detect our chewing and swallowing sounds while eating[10, 17] or using cameras and computer vision algorithms to recognize food type and estimate amount [4, 13]. However, all these methods raise more concerns, like privacy problems, and are computationally expensive, which decreases the likelihood that the interface will be used. Another challenge is that many users do not want to wear devices that call unnecessary attention to themselves or might cause onlookers to believe the user has a disability [12].

This paper presents a wearable system that can distinguish various foods while maintaining a potentially socially acceptable appearance (standard optical glasses' based design). The system contains two IMU[1] sensors; one is used to obtain the vibration generated by biting and chewing, and the other is used to receive the corresponding hand movements during the eating activities. Our initial user study confirmed that our approach could distinguish seven types of food at an accuracy of 93.26% (n=4).

Our key contribution is the concept of First Bite/Chew-based food type detection concerning wearable design and socially acceptable appearance. Our first bite/chew-based approach has several benefits: (1) exact food intake: unlike related works that can only give an estimation of how many calories a certain food may contain, our approach can detect how many bites a user actually eats, which contribute to a more accurate food journaling (2) computational simplicity: our system only contains two IMUs and an inexpensive machine-learning capable MCU[2] as the main parts (3) easy-to-use: to monitoring the food intake, our approach does not require extra practice and learning (4) replicable: our approach can be easily replicated concerning cost and design (5) potentially socially acceptable appearance: our method does not require camera nor bulky battery (i.e., huge glasses' legs) and maintains an ordinary optical glasses-like appearance.

2 RELATED WORKS

Current food intake monitoring methods can be roughly categorized as IMU-based, image-based, sound-based, wearable on-body sensor based and self-reported methods.

[1]Inertial Measurement Unit
[2]Micro Controller Unit

2.1 IMU sensor based

IMU is widely used in detecting food intake. And its usage is commonly focused on detecting the wearer's hand movements.[3, 11, 16]. By putting IMU in the wristband and detecting the user's hand gestures, we can determine whether the user is intaking food.

Although all the above methods can detect food intake, they become limited when detecting food types and calories. Kim et al. [6] provided a smartwatch-based method to address different eating patterns and food types. Their method only handled rice and noodle in their tests. One challenge for hand movement-based methods is the significant variations of an individual and among groups, such as eating by hand(s), chopsticks, fork, knife, or other types of tableware [16].

Studies also combine IMU and Piezoelectric sensors on the eyeglasses to track the user's chewing action by detecting Jaw Elevation, and temporalis muscle contraction [15].

2.2 Sound based

Sound-based food intake detection has two main methods. The most studied one is using microphones from hearing aids, earphones, or headsets to capture users' chewing noise and use it as an indicator of the food intake action[10].

The other method uses the Throat microphones attached to the user's neck to detect the wearer's swallowing sound[17]. It also has IMU to capture throat vibration for further clarification.

Although sound-based detection can detect users chewing and swallowing, it can also be challenged when detecting the type of food and its Calories. Furthermore, with sound-based detection, the environment is also an important factor. It still needs time to prepare for real-life usage. On the other hand, hearing aid package-based methods or wired-looking devices further bring the concern of being disabled [10].

2.3 Image based

Regarding image analysis-based methods, processes like image segmentation, food recognition, and portion size estimation are required for a whole process of food intake estimation [4]. Those processes are generally computationally heavy, and some may even require users to follow strict guidelines to prepare the image source files [5, 20].

To monitor food intake, computer vision-based studies [4, 13] can offer estimated nutrition or calories for certain specific food types. However, they needed to provide how much the user actually ate food. Images that contain a full meal may also bring that image-based extra processes [4]. When it comes to food intake detection, the usage of the camera is relatively well studied. However, the usage of the camera is significantly different. Image recognition uses either a smartphone's camera or cameras attached to the wearable device to recognize the food through computer vision [7, 18].

2.4 Glasses based chewing detection

Mertes et al. [8, 9] developed a glasses-based method that detects the chewing motion of elderly people. Chung et al. [2] designed a pair of smart glasses that can classify food intake motions from physical activities. Bedri et al. [1] offered multi-modal sensor-based glasses with a camera detecting food intake events in noise areas and recording videos to help users remember today's intake. However, to the best of our knowledge, none of those existing smart glasses-based studies could distinguish different types of food.

2.5 Self-reported methods

Self-reported methods [4] require a lot of effort and time, like the tedious operation of smart devices and other types of recording medium from the user. They are prone to be forgotten or given up [10].

3 OUR APPROACH: FIRST BITE/CHEW BASED FOODS INTAKE MONITORING

Our approach leverages the different food textures, i.e., the potentially various feedback the food responds to the biting/chewing activities and the potentially different corresponding hand movements while eating to distinguish the different types of food.

To obtain intensive enough data from biting/chewing and hand movements while eating, we designed a device consisting of a pair of ordinary glasses and a wristband. As shown in figure 2, an IMU[3] was attached on the inner side (near the head) of the right glasses leg, which is close to the area of superior auricular muscle and temporalis muscle when glasses are put on. An IMU[4]-embedded MCU[5] was placed on the wristband and connected to the IMU on the glasses with four wires via IIC-Bus.

Since different types of food may share the similar kind of texture, we also add the corresponding hand movement during the first bite/chew to collaborate with the first bite/chew signals. See figure 3. Therefore, a first bite/chew time window contains two parts, the upper part in the thin yellow rectangle is the corresponding movement data of the dominant hand generated during the first biting/chewing. The lower part in the bold red rectangle is the vibration from teeth as well as the head and muscles' activities responding to the food during the first biting or chewing.

4 INITIAL FEASIBILITY STUDY

To prove first bite/chew is a stable indicator for food classification, 4 participants (2 male and 2 female), from 23 to 32 years old, were recorded one by one while having meals on campus. In total, we recorded five meals, including two lunches and three dinners for each participant over three days in a row.

Participants were required to confirm and sign a consent form and an allergic food form. The checking lists contain peanuts, milk, pork, beef, and typical kinds of seafood. Participants were also asked both *food and drink prohibitions* to make sure no one mistakenly takes inappropriate or even fatal foods. Considering participants' preferences and real-life scenarios, as figure 4 shows, seven types of daily foods are chosen to examine over different textures and potentially different hand movements. The seven types of food are instant noodles, apples, nuggets, hamburgers, peanuts, edamame, and fried rice.

[3]MPU6050 based IMU modular module.
[4]LSM6DS3
[5]Seeeduino Xiao BLE sense. https://www.seeedstudio.com/Seeed-XIAO-BLE-Sense-nRF52840-p-5253.html

4.1 Data Collection

Before data recording started, each participant was helped to put on the device. The wristband was put on the participant's dominant hand in order to obtain the corresponding eating hand gesture. All participants were provided with a typical meal consisting of one main course and two starters. Participants were free to have their meal in their most natural way after data recording started. We recorded the time of the beginning of the preparation movement and the first bite/chew as ground truth annotations. For further analysis, the accelerometer and gyroscope data were cut into 15 seconds pieces, a duration that covers a whole eating cycle of one bite.

We successfully recorded a total of 916 valid samples as our dataset. In detail, 125 for apples, 76 for hamburgers, 179 for edamame, 139 for instant noodles, 123 for nuggets, 169 for peanuts, and 105 for fried rice.

4.1.1 Observation. From our observation, we found that the four participants' eating motions and gestures have similarities and personal characteristics. Similarities include: 1) all participants open their mouth when the food is halfway moved to their mouth 2) for food like burgers and apples, participants tend to hold their food next to their mouths in preparation for the next bite, so the motion of hand-mouth approaching each other is not obvious 3) the eating preparation period of instant noodle is longer than other food, participants often blow on noodle for few seconds before they take their bites. Regarding personal characteristics of each participant's eating patterns, participant #1 has the tendency to look and spin the apple before eating, participant #2 has the habit of holding up the phone with left hand and browsing while eating, participant #3 looks down at the phone on the table while eating, participant #4 couldn't eat pork due to religious background. Since we couldn't find any fried rice with no pork product, participant #4 ate only the rest six types of food for the data collection.

5 EVALUATION OF FIRST BITE/CHEW-BASED APPROACH

5.1 Segmentation and Annotation

One previous research addressed two major features during an eating cycle, movements like hand and mouth approaching each other and the biting/chewing motion. Ye et al. point out that hand-to-mouth motions can be further divided into hand ascending period, biting period, and hand descending period [19]. To dig deeper, we found that for some specific types of food, eating preparation motion is the mouth reaching to the hand for food rather than moving food to mouth by hand.

Combining previous research and our observation, we address three steps in an eating cycle: 1) the preparation period during which hand and mouth approach each other, 2) the first biting or chewing motion 3) the regular chewing period. For the seven types of food we recorded, subjects' first bite follows closely after hand and mouth have approached each other.

First, we labeled the ground truth of the eating preparation period based on our recorded timestamp during observation. Then we labeled the beginning of the first bite segment right after the eating preparation period. For some types of food, the first bite is

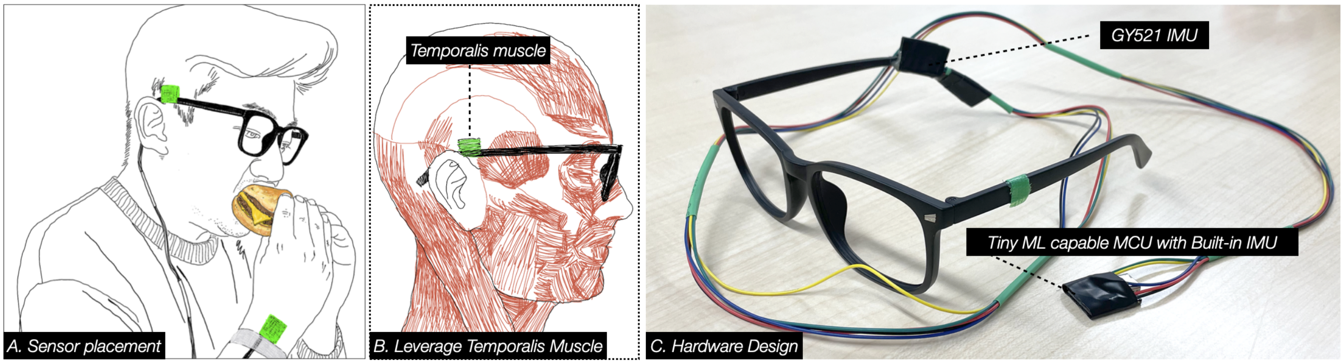


Figure 2: Hardware design and the sensors' placement. A. indicate how we place the two IMU sensors, one is attached to the right leg of the glasses, the other is fastened to the wrist of the dominant hand. B. shows how we leverage the temporalis muscle as one of our data resource. C. is the hardware design.

Figure 3: First bite/chew are like the red rectangles indicate, they are significantly different from the following chews which are relatively the same among different food types.

Figure 4: Seven selected food with different texture and potentially different eating patterns

unavailable in the second step because participants can put one bite of food straight into their mouths without biting motion, and we replaced it with the first chew. The chewing segment is extracted randomly from the following regular chewing period. All segments were exactly 2 seconds, a time period sufficient enough for covering all desired data.

5.2 Pre-Processing

A sliding window at a length of 1600 ms with 12.5% overlap was adopted for feature extraction and training. For each window, 132 features are extracted from both the time domain and frequency. In more detail, for each axis of the accelerometer and the gyroscope: 1) root mean square, 2) skewness, 3) kurtosis, 4)the first eight discrete Fourier transform coefficients.

5.3 Initial Result

For the food category recognition, a single-layer Neural Network was adopted. We trained and tested the classifier over the subject's own data by using 80% samples as training data and the remaining 20% as testing data. As the figure 5 shows, the classifier achieves an average accuracy of 93.3% and an average F1 score of 0.92 for all food across all participants. In detail, an average of 91.8% accuracy for Apple, 95.5% for Burger, 95.2% for Edamame, 96.1% for Egg fried rice, 91.2% for Instant noodle, 89.3% for Nuggets, and 94.7% for Peanuts.

6 DISCUSSION & LIMITATION

Though our initial study showed a promising result that it is feasible to detect different types of food using a combination of the first bite/chew and the corresponding hand movement from two IMUs, there are challenges remain.

Exact Food Types. Although our initial result suggests that the use of a combination of the first bite/chew and its corresponding hand movement could be a simple and reliable indicator to distinguish food types. However, since this current prototype was designed concerning socially acceptable appearance and did not use a camera. There is a chance that different types of food share a similar texture and the corresponding hand movement. It meets the limitation of our current approach.

Extending to New Users. The current model is trained on a per-user basis, making it difficult to adapt to a new user since different users may have different patterns even eating the same food. Take the noodle as an example. In our observation, participant #1 tends to bow the head to reach the bowl, while other participants prefer to pick up noodles and bring them to their mouth. Such differences will cause totally different patterns, including amplitude and frequencies. In future work, we plan to address this problem by collecting more data within a larger user group as our base model, such that it will contain more scenarios, and it may take less effort when using techniques like transfer learning to adjust to a new user.

Extending to New Food. Presently, our method can only distinguish seven kinds of different food. Although we could collect more categories in future work, in the training phase, it is impossible to include all food types the user would have in real life. One possible solution is applying techniques like few shots learning in the ongoing research, which is training a model to tell the similarity of two input signals instead of directly mapping signals to categories. As a result, the user only needs to collect a few samples of the new category as a reference. Each time a new real-time signal comes to the system, it will be compared with all references and will be categorized as the one with the most similarity. No extra training is required, and the user's burden is much relieved as they only need to collect a few samples.

7 POTENTIAL APPLICATIONS

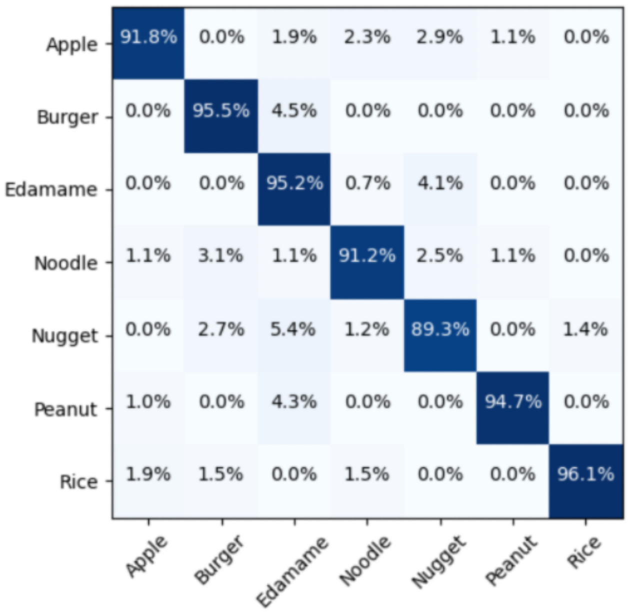
With the pursuit of food-related applications rather than only focusing on food intake detection, we expanded our focus on food classification to help people suffering from health problems and memory loss. Therefore we introduce the following two potential user scenarios supported by our proposed approach.

Precise Calories Intake Estimation

During our initial feasibility test, we found the bites that we need to eat up a particular food is relatively stable. Take participant #1 as an example, he ate a total of 4 double cheeseburgers in 3 days, all ending up with 8 bites, and a cup of instant noodles always takes 12 bites (SD = 0.47). The conclusion also holds for other people with a small bias. Therefore, our approach can be extended to calculate how much food we actually take. In contrast, current methods require either complex operation or manual monitoring.

Remote Elderly Food Intake Monitoring

Elderly people experience changes in eating habits and appetite for various reasons, especially the elderly, who suffer from age-related diseases. Since our system can detect what type of food and how many bites the user takes, we believe our approach can be used to monitor the elder's eating behavior remotely. What we can improve from existing methods for remote monitoring of elders' eating behavior is that our system is wearable and camera-free.



Food	F1 Score
Apple	0.94
Burger	0.91
Edamame	0.90
Noodle	0.93
Nugget	0.85
Peanut	0.97
Rice	0.96

Figure 5: Average Confusion Matrix and F1 score, each entry of which is the average value across 4 participants

8 CONCLUSION

In this work, we proposed a concept that aims to distinguish different food types by leveraging the first Bite/Chew and the corresponding hand movements during the first bite/chew. Then we validate the feasibility of our system by collecting real-time eating data from 4 participants and doing a series of analyses. The result shows though only 2 IMUs are adopted to capture hand and mouth movement, our approach still has a good performance by both metrics of accuracy and F1 score. Finally, we introduced some possible user scenarios enabled by the system and hope to inspire more future studies in this field.

REFERENCES

[1] 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.

[2] Jungman Chung, Wonjoon Oh, Dongyoub Baek, Sunwoong Ryu, Won Gu Lee, and Hyunwoo Bang. 2018. Design and evaluation of smart glasses for food intake and physical activity classification. *JoVE (Journal of Visualized Experiments)* 132 (2018), e56633.

[3] Yujie Dong, Adam Hoover, Jenna Scisco, and Eric Muth. 2012. A new method for measuring meal intake in humans via automated wrist motion tracking. *Applied psychophysiology and biofeedback* 37, 3 (2012), 205–215.

[4] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni. 2017. Automatic diet monitoring: a review of computer vision and wearable sensor-based methods. *International journal of food sciences and nutrition* 68, 6 (2017), 656–670.

[5] Hamid Hassannejad, Guido Matrella, Monica Mordonini, and Stefano Cagnoni. 2015. ID 45 - A Mobile App for Food Detection: new approach to interactive segmentation.

[6] Hyun-Jun Kim, Mira Kim, Sun-Jae Lee, and Young Sang Choi. 2012. An analysis of eating activities for automatic food type recognition. In *Proceedings of the 2012 asia pacific signal and information processing association annual summit and conference*. IEEE, 1–5.

[7] Corby K Martin, Hongmei Han, Sandra M Coulon, H Raymond Allen, Catherine M Champagne, and Stephen D Anton. 2008. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. *British Journal of Nutrition* 101, 3 (2008), 446–456.

[8] Gert Mertes, Hans Hallez, Tom Croonenborghs, and Bart Vanrumste. 2015. Detection of chewing motion using a glasses mounted accelerometer towards monitoring of food intake events in the elderly. In *International Conference on Biomedical and Health Informatics*. Springer, 73–77.

[9] Gert Mertes, Hans Hallez, Bart Vanrumste, and Tom Croonenborghs. 2017. Detection of chewing motion in the elderly using a glasses mounted accelerometer in a real-life environment. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4521–4524.

[10] Sebastian Päßler, Matthias Wolff, and Wolf-Joachim Fischer. 2011. Food intake recognition conception for wearable devices. In *Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare*. 1–4.

[11] M Popa. 2011. Hand gesture recognition based on accelerometer sensors. In *The 7th International Conference on Networked Computing and Advanced Information Management*. IEEE, 115–120.

[12] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K Kane. 2016. The AT effect: how disability affects the perceived social acceptability of head-mounted display use. In *proceedings of the 2016 CHI conference on human factors in computing systems*. 4884–4895.

[13] Manika Puri, Zhiwei Zhu, Qian Yu, Ajay Divakaran, and Harpreet Sawhney. 2009. Recognition and volume estimation of food intake using a mobile device. In *2009 Workshop on Applications of Computer Vision (WACV)*. IEEE, 1–8.

[14] Dale A Schoeller. 1995. Limitations in the assessment of dietary energy intake by self-report. *Metabolism* 44 (1995), 18–22.

[15] Jaemin Shin, Seungjoo Lee, Taesik Gong, Hyungjun Yoon, Hyunchul Roh, Andrea Bianchi, and Sung-Ju Lee. 2022. MyDJ: Sensing Food Intakes with an Attachable on Your Eyeglass Frame. In *CHI Conference on Human Factors in Computing Systems*. 1–17.

[16] Edison Thomaz, Abdelkareem Bedri, Temiloluwa Prioleau, Irfan Essa, and Gregory D Abowd. 2017. Exploring symmetric and asymmetric bimanual eating detection with inertial sensors on the wrist. In *Proceedings of the 1st Workshop on Digital Biomarkers*. 21–26.

[17] MA Tuğtekin Turan and Engin Erzin. 2017. Empirical mode decomposition of throat microphone recordings for intake classification. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*. 45–52.

[18] Donald A Williamson, H Raymond Allen, Pamela Davis Martin, Anthony J Alfonso, Bonnie Gerald, and Alice Hunt. 2003. Comparison of digital photography

to weighed and visual estimation of portion sizes. *Journal of the American Dietetic Association* 103, 9 (2003), 1139–1145.

[19] Xu Ye, Guanling Chen, and Yu Cao. 2015. Automatic eating detection using head-mount and wrist-worn accelerometers. In *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*. IEEE, 578–581.

[20] Fengqing Zhu, Marc Bosch, Insoo Woo, SungYe Kim, Carol J Boushey, David S Ebert, and Edward J Delp. 2010. The use of mobile devices in aiding dietary assessment and evaluation. *IEEE journal of selected topics in signal processing* 4, 4 (2010), 756–766.