The Impact of Walking and Resting on Wrist Motion for Automated Detection of Meals

SURYA SHARMA, PHILLIP JASPER, ERIC MUTH, and ADAM HOOVER, Clemson University

This article considers detecting eating in free-living humans by tracking wrist motion. We are specifically interested in the effect of secondary activities that people conduct while simultaneously eating, such as walking, watching television, or working. These secondary activities cause wrist motions that obfuscate those associated with eating, increasing the difficulty of detecting periods of eating. We collected a large dataset of 4,680 hours of wrist motion from 351 participants during free living. Participants reported secondary activities in 72% of meals. Analysis of wrist motion data revealed that the wrist was resting 12.8% of the time during self-reported meals compared to only 6.8% of the time in a cafeteria dataset, whereas walking motion was found 5.5% of the time during meals in free living compared to 0% in a cafeteria. Augmenting an eating detection classifier to include walking and resting detection improved accuracy from 74% to 77% on our free-living dataset ($t[353] = 7.86$, $p < 0.001$). Although eating detection could be improved using more sophisticated machine learning methods or sensor modalities, all approaches would be affected by secondary activities, as they affect the labeling of data itself. Our work suggests that future work should collect detailed ground truth on secondary activities being conducted during eating, as these activities could hold insights into when an eating activity starts or stops in the absence of video-based ground truth.

CCS Concepts: • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; • **Applied computing** → *Health informatics*; • **Computing methodologies** → Supervised learning by classification;

Additional Key Words and Phrases: Automated dietary monitoring, eating detecting, walking detection, resting detection, gesture recognition, m-health, wearables, obesity

ACM Reference format:
Surya Sharma, Phillip Jasper, Eric Muth, and Adam Hoover. 2020. The Impact of Walking and Resting on Wrist Motion for Automated Detection of Meals. *ACM Trans. Comput. Healthcare* 1, 4, Article 24 (September 2020), 19 pages.
https://doi.org/10.1145/3407623

1 INTRODUCTION

This work is motivated by the problem of detecting eating activities all day during free living. Eating activities refer to meals, snacks, and other contiguous periods of consumption. Detection of eating activities can be achieved by various sensing modalities, such as a microphone near the ear or throat [5], jawbone movements

This work was supported by National Institutes of Health (NIH) grant #R01HL118181-01A1.
Authors' addresses: S. Sharma, Department of Electrical and Computer Engineering, Clemson University, Riggs Hall, Clemson, SC 29634; email: meals@suryasharma.com; P. Jasper, 23445 North 19th Avenue, Phoenix, AZ 85027; email: phillipwjasper@gmail.com; E. Muth, 1601 East Market Street, Fort IRC Building, Greensboro, NC 27411; email: muth@clemson.edu; A. Hoover, Department of Electrical and Computer Engineering, Clemson University, Riggs Hall, Clemson, SC 29634; email: ahoover@clemson.edu.

2637-8051/2020/09-ART24 $15.00
https://doi.org/10.1145/3407623

Table 1. Change in Precision and Recall of Meal Detection When Transitioning from a Controlled Environment to Free Living

Previous Work	Controlled		Free Living	
	Precision	Recall	Precision	Recall
Thomaz et al. [28]	67	89	65 (−2)	79 (−10)
Mirtchouk et al. [17]	88	87	45 (−43)	85 (−2)
Chun et al. [8]	95	82	78 (−17)	73 (−9)
Zhang and Amft [33]	94	90	79 (−15)	77 (−13)

[8], throat movements [2], apnea detection [10], wrist motion [19, 23], or a combination of them [4, 22]. This work considers wrist motion that is tracked using a watch-like device that contains accelerometers, gyroscopes, and magnetometers. Our methods search for motion patterns of the wrist on the dominant hand that happen during eating, and use those patterns to segment and identify periods of time that look like eating activities. For example, consumption generally involves using the hand to pick up food and deliver it to the mouth, sometimes called a *hand-to-mouth gesture* [14], or more simply, a *bite* [18]. This motion tends to occur in clusters for what is commonly called a *meal* or *snack*. The specific problem considered in this article is that people can conduct secondary activities while eating, such as watching television, talking with other people, walking to retrieve more food, or resting for a few minutes before resuming consumption. These secondary activities impact the wrist motion pattern of bites in ways that are unknown. The overlap of these activities creates challenges for methods designed to detect and classify periods of time as eating activities.

Recent work in automated eating activity recognition has shown that classifier accuracy decreases when transitioning from the laboratory to free living (Table 1). However, there is ongoing debate as to its cause. Thomaz et al. [28] showed that in a laboratory setting, activities like chatting, using the phone, and brushing hair with a comb can be confused with eating. Zhang et al. [34] learned that most meals are consumed when a human is stationary, and thus excluded periods of walking showed that walking often looks like eating gestures in free living [34]. Bi et al. [6] discussed how walking while eating caused misclassification in their eating activity approach. We hypothesize that another important factor may be secondary activities conducted concurrently with eating. Laboratory tests have advantages in that the data can be collected under direct observation, which makes the annotation of ground truth behaviors more simple, and because a scripted list of activities can be given as instructions to participants. However, while being directly observed, participants are unlikely to conduct secondary activities while eating unless specifically instructed to do so. In contrast, collecting data from free-living participants is more complicated [8, 17, 28] because tools that can be used in the laboratory like a video camera cannot be easily implemented in a free-living setting, and when they are used, they bring concerns about the privacy of the participants and people they interact with [29]. Bedri et al. [5] discussed how it is difficult to collect data that is precise, generalizable, and real. The authors state how data collected in the laboratory is precise but not real, whereas that collected in free living is imprecise due to the lack of proper instrumentation that can provide accurate ground truth. Participants may also exhibit behaviors in free living that were not captured in the laboratory, as they are often limited by tracking devices utilized in the laboratory or when under observation. Alharbi et. al [1] provided examples of how participants wearing cameras during the collection of eating activity data experience social and surveillance discomfort, whereas Mirtchouk et al. [17] noted that their classifier performed poorly on 1 participant out of 11 (9%), as that participant had extended conversations and multi-tasked (did homework) while eating a meal. Another recent study performed statistical tests on data collected from participants in a laboratory versus data collected from participants in a free-living facility and reported numerous differences, including changes in the number of bites, the time spent eating, and the time and number of pauses between ingestion events [12].

Table 2. Complexity and Size of Datasets in Previous Works

Authors	Hours	Participants	Ratio
Dong et al. [11]	449	43	20:1
Thomaz et al. [28]	32	7	14:1
Bedri et al. [5]	45	15	8:1
Bedri et al. [5]	12	10	3:1
Mirtchouk et al. [17]	245	11	12:1
Zhang and Amft [33]	122	10	17:1
Farooq et al. [13]	10	40	16:1
This work	**4,680**	**351**	**18:1**

Behavior variability may also be broadened by increasing the size of the dataset used for testing, which in turn may decrease classifier accuracy. Previous works in eating activity recognition tested 4 to 43 participants (Table 2). It is questionable if this limited size of participants exhibits the total variability expected in eating behaviors across a larger population. In addition, the ratio of non-eating data to eating data in previous work ranges from 3:1 to 20:1. An average person spends 1.17 hours in a day on eating activities [30], a non-eating to eating ratio of approximately 20:1. This ratio is important, as it affects metrics like precision and the F1 score.

1.1 Contribution

The contribution of this work is as follows. We provide a dataset one to two orders of magnitude larger (351 participants) than all works cited previously (4–43 participants). All of the data was collected during free living. We collected information on secondary activities during eating and quantify how often they occur. We developed a classifier to detect two common secondary activities, walking and resting, and show that eating activities can be detected more accurately using this new classifier compared to a previous classifier [11]. We recommend that any future classifiers intended to detect eating episodes in free living be trained on data that contains secondary activities so that the classifier has a better chance of operating reliably during free living.

2 OVERVIEW

We first demonstrate how wrist motion patterns associated with eating can be obfuscated by wrist motions associated with secondary activities. Figure 1 shows 60 seconds of wrist motion data from a person eating a banana. Top to bottom are accelerometer x, y, and z, and gyroscope yaw, pitch, and roll. Modulations in the signals are caused by wrist motions moving food to the mouth and then moving the wrist back to a neutral position, whereas periods of no motion indicate when the wrist was at rest. This type of motion is typical during consumption. Figure 2 shows 60 seconds of wrist motion data from a person walking down a hallway and then entering a room and searching around inside it. Swinging the arms during walking causes sensor patterns that have regular oscillations and look clearly different than the sensor patterns during eating.

One can imagine several scenarios in which a person might walk or rest during a meal or snack. For example, a person might walk around a kitchen to prepare a second serving of food before sitting down and resuming consumption. A person might rest for several minutes while watching TV before resuming consumption. Figure 3 shows an example of this type of behavior toward the end of a meal in which a person rested for 15 seconds, engaged in consumption for 35 seconds, briefly rested, and then walked for 15 seconds before indicating the end of the meal. These secondary activities occurred intermittently within the period of eating, but depending on their frequency and duration, they can greatly increase the difficulty of detecting the period of eating. Figure 4 shows 1 minute of wrist motion data from a person eating a banana while walking. In this case, the secondary activity is occurring concurrently during eating, which obfuscates all of the wrist motions associated with eating.

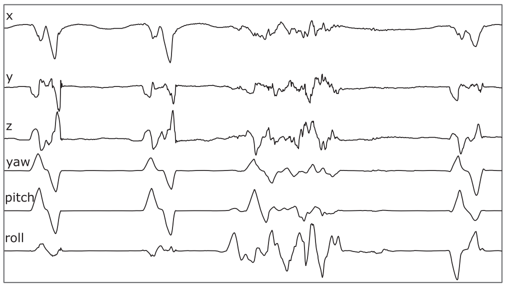


Fig. 1. Example of 1 minute of wrist data (linear accelerations x, y, z and gyroscope yaw, pitch, roll) from a person eating a banana with rest between bites.

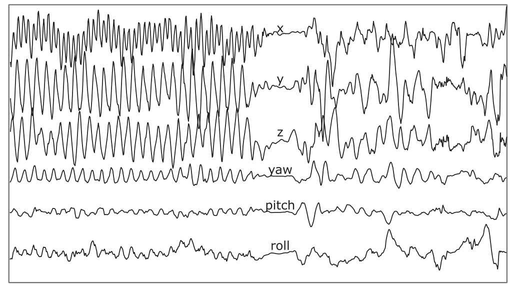


Fig. 2. Example of 1 minute of wrist data from a person walking regularly for 30 seconds, followed by a short stop, then looking around a room.

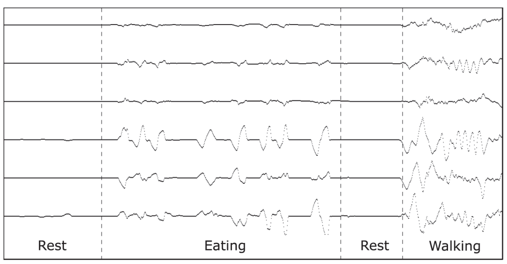


Fig. 3. Wrist data from the end of a meal, showing intermittent eating, walking, and resting. In this segment, the participant rested for a period of 15 seconds, consumed food for 35 seconds, rested briefly, then walked for 15 seconds before ending the meal.

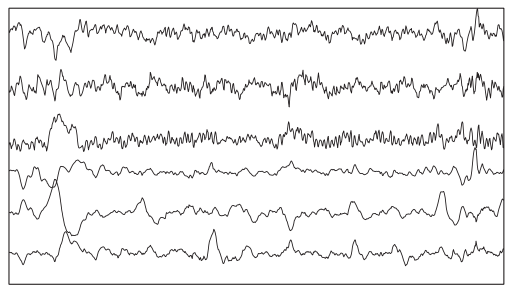


Fig. 4. Example of 1 minute of wrist data (linear accelerations from a person eating a banana while walking). The secondary activity of walking obfuscates the wrist motion signals indicative of food consumption.

Classification of free-living meals containing secondary activities could be performed by modeling mixtures of activities as different classes. However, a secondary activity may not be conducted continuously for the entire duration of the eating activity. We therefore take an approach where subsegments of a self-reported meal are analyzed and classified independently. We specifically consider two secondary activities: walking and resting. We present two experiments. In our first experiment, we develop detectors for periods of walking and for periods of resting. We test these detectors on two datasets for which video ground truth of activities is available. The first dataset was collected for a pedometer experiment and is known via visual confirmation of video to contain 100% walking. The second dataset was collected in a cafeteria. Through visual confirmation of video, human raters labeled 5.8% of the time during meals as rest. In our second experiment, we use the walking and resting detectors to measure how frequently these secondary activities occur during periods of eating in free living. No video-based ground truth is available for the second experiment, so we rely upon the results of the first experiment to provide confidence in the measures found in the second experiment. Finally, we show how walking and resting detectors can be added to a previously existing eating activity detection algorithm [11] to improve its performance.

3 WALKING AND RESTING DETECTION

The purpose of this experiment is to develop classifiers for detecting walking and resting by tracking wrist motion. We test them on two datasets containing acceleration and gyroscope data from wrist-mounted

Fig. 5. The Shimmer3 sensor mounted on the wrist. This sensor was used for data collection in the pedometer [16] and free-living eating activity datasets (this work). The custom IMU sensor in the cafeteria dataset [25] is mounted in the same manner.

sensors that have video-based ground truth of participant activities. Both of the datasets were collected using wrist-mounted sensors in the same configuration. The goal is to provide confidence that they work reliably enough to detect walking and resting on additional datasets also containing acceleration and gyroscope data for which video-based ground truth is not available.

3.1 Datasets

A dataset containing wrist activity data (from accelerometers, gyroscopes, and magnetometers) during walking was collected by Mattfeld et al. [16] for pedometer algorithm evaluation. A total of 30 participants were recorded. Each participant was instrumented with three Shimmer3 sensors (wrist, hip, and foot) and was followed by an experimenter using a smartphone to record synchronized video of their lower body. Figure 5 shows an example of the Shimmer3 mounted on the wrist. The Shimmer3 houses MEMS accelerometers, gyroscopes, and magnetomters. Participants walked an outside path, inside a building, and inside a room, collectively taking more than 60,000 steps. For this work, we use the data collected from the wrist while the participant walked an outside path. The dataset is publicly available at http://www.cecas.clemson.edu/~ahoover/pedometer.

A second dataset was collected in a cafeteria setting [21, 25]. A total of 271 participants were recruited from Clemson University and its neighboring areas. Volunteers were asked to pick a date and a meal (lunch or dinner) to consume in the university cafeteria in the company of three other volunteers. Each participant consumed a single meal during data collection. Participants sat at an instrumented table that had video cameras installed above it in the ceiling to record each participant and their food while they ate. Each participant wore a custom device on the wrist housing MEMS accelerometers and gyroscopes. Ground truth of bites and other eating-related gestures like resting and manipulating food was provided by trained reviewers watching the synchronized video. Full details and the dataset are publicly available at http://www.cecas.clemson.edu/~ahoover/cafeteria.

3.2 Preprocessing and Segmentation

All datasets used in this work were recorded at 15 Hz. To reduce sampling noise, we filter raw acceleration and gyroscope signals $R_t = \{a_x, a_y, a_z, \omega_\phi, \omega_\theta, \omega_\psi\}$ at time index t to smoothed signals $S_t = \{S_{x,t}, S_{y,t}, S_{z,t}, S_{\phi,t}, S_{\theta,t}, S_{\psi,t}\}$ using a standard Gaussian filter [9] operated independently on each axis. The filter operates on a window of past data 1 second long using a Gaussian σ of 10 seconds. For classification and evaluation, we segmented the data into fixed 1-minute windows, starting 10 seconds before the first step (pedometer dataset) or bite (cafeteria dataset), and ending 10 seconds after the last step or bite. Segments smaller than 1 minute were discarded.

3.3 Detection of Secondary Activities

3.3.1 Walking. A zero-crossing-based algorithm is employed to detect walking. For a given segment, a datum is identified as a zero-crossing $z(t) = 1$ if any axis in the gyroscope signal ($S_{gyro,t} \in \{S_{\phi,t}, S_{\theta,t}, S_{\psi,t}\}$) crossed zero (from negative to positive or vice versa). To suppress spurious detection of zero crossings from noise, the signal must also surpass ±5 deg/sec in the direction of the zero crossing. The threshold of ±5 deg/sec was decided heuristically to avoid spurious detection of zero crossings caused by noise in the signal during rest but is small enough to detect the oscillations of the wrist caused by walking.

$$z(t) = \begin{cases} 1 & \text{if } sgn(S_{gyro,t-1}) \neq sgn(S_{gyro,t}) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The feature $f_{1,w}$ provides the rate of zero crossings for a segment w of length W and is calculated using Equation (2). It defines the percentage of data in the segment identified as zero crossings and ranges from 0 to 1. Larger values (large amounts of zero crossings) tend to occur during walking compared to other activities. A segment is considered walking if $f_{1,w}$ is greater than threshold T_1.

$$f_{1,w} = \frac{1}{W} \int_W z(t) \tag{2}$$

3.3.2 Resting. Resting can be detected by looking for low variance in accelerometer and gyroscope signals. By using variance in the signal, we are able to detect rest regardless of the orientation of the device. For robustness, our classifier uses two steps. Each datum is first classified as rest $r(t) = 1$ or motion $r(t) = 0$. To do this, the standard deviation σ_t is calculated over a window of $M = 1$ second for each of the six signals $S_{x,t}$, $S_{y,t}$, $S_{z,t}$, $S_{\phi,t}$, $S_{\theta,t}$, $S_{\psi,t}$. If the sums of standard deviations for the acceleration ($\sigma_{A,t} = \sigma_{x,t} + \sigma_{y,t} + \sigma_{z,t}$) and gyroscope ($\sigma_{\omega,t} = \sigma_{\phi,t} + \sigma_{\theta,t} + \sigma_{\psi,t}$) signals are less than T_A and T_ω respectively, the datum is considered to be at rest:

$$r(t) = \begin{cases} 1 & \text{if } \sigma_{A,M} < T_A \textit{ and } \sigma_{\omega,M} < T_\omega \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

A segment is then classified as rest if the percentage of data detected rest in the segment is greater than the threshold T_2. The feature $f_{2,w}$ (Equation (4)) ranges from 0 to 1. Larger values tend to occur during resting compared to other activities.

$$f_{2,w} = \frac{1}{W} \int_W r(t) \tag{4}$$

3.4 Parameter Tuning

Parameters for the walking and resting detectors were tuned using histogram analysis of the pedometer and cafeteria datasets. Figure 6 shows the values of $f_{1,w}$ for data from each participant in the datasets. We set $T_1 = 0.15$ in the middle of the two histograms to label all walking segments as walking, and all cafeteria meals as not walking, since the cafeteria dataset did not contain any periods of walking.

For the rest detector, T_A and T_ω were set by calculating the maximum value of standard deviation in the acceleration and gyroscope signals in segments visually identified as rest. To identify threshold for the amount of rest in a segment (T_2), we plotted a histogram of the amount of rest in cafeteria meals (Figure 7). T_2 was set to 0.65 based on this histogram. There were some meals in the cafeteria where the amount of rest was greater than 65%; however, video evidence showed the participant was eating with the non-instrumented hand, with the instrumented hand largely at rest.

4 FREE-LIVING EATING ACTIVITIES

In this section, we discuss free-living eating activities and their classification. We describe the collection of a large new dataset and then augment a previously described eating activity classifier [11] with the proposed walking and resting detectors to learn how secondary activities can impact detection performance.

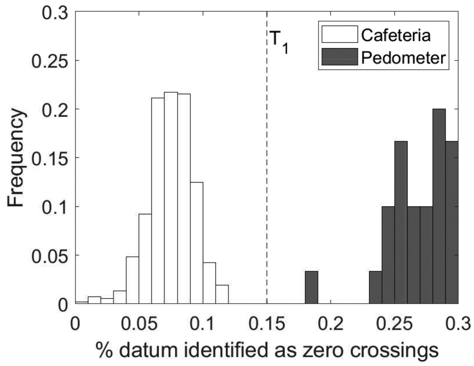


Fig. 6. Histogram showing amount of zero-crossings per participant in the pedometer and cafeteria datasets.

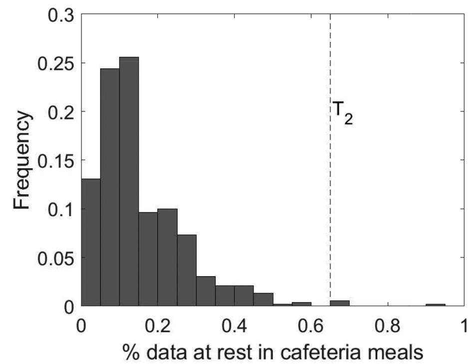


Fig. 7. Histogram showing the amount of rest per meal in the cafeteria dataset.

4.1 Free-Living Eating Activity Dataset

We programmed 30 Shimmer3 devices (see Figure 5) for the data collection, with a goal to collect data from approximately 20 participants each week. The Shimmer3 was programmed to record accelerometer, gyroscope, magnetometer, orientation, and button press data at 15 Hz. A team of eight research assistants helped with data collection, and a 3-day protocol was designed. Participants would first interview with a researcher on day 1, collect data during their free-living day on day 2, and return to the laboratory on day 3 to return the Shimmer3 and download the data. Each research assistant was trained on the use of the Shimmer3 devices and provided with an instruction manual. The manual contained details on the data collection protocol, and instructions on how to configure, recharge, and maintain the Shimmer3 devices.

The Clemson University Institutional Review Board approved the collection of this dataset. Participants were recruited from the student body, faculty, staff, and residents of the surrounding area using mailing lists, fliers, emails, and word of mouth. Temporary laboratory spaces were set up in buildings in the university and nearby cities for up to 2 weeks to assist with data collection. The recruitment strategy focused on balancing the age, gender, and ethnicity in the participant pool, as well as the day the data was collected on (weekday vs weekend). Each participant was provided informed consent and an incentive of $25 to collect wrist motion tracking data for 1 day during free living. An initial pilot of eight participants was conducted, and the data collection was revised based on participant feedback. Based on the feedback, we created a video explaining the data collection for the participants to watch. This video contained specific examples of what an eating episode is. We added a process where participants demonstrated the ability to correctly use the Shimmer3 as intended before data collection. We also designed a business card with summary instructions and a contact phone number to call if the participant needed any help.

Each participant was first screened using a questionnaire for past history of eating disorders. Any participant with a history of eating disorder was excluded from the study. Participants were then provided with dates and times for appointments and data collection. On the first day of the interaction, participants met with a research assistant who explained the 3-day study and provided informed consent. The research assistant collected measurements such as height, weight, BMI, body fat percentage, and hip and weight circumference. Each participant was assigned an anonymous participant ID number that was used to link the Shimmer3 data to this data, and it was also used on day 3 to download the data from the device. Participants were provided with their Shimmer3 device and watched a 4-minute video that defined eating activities and demonstrated how the Shimmer3 should be used the following day for data collection.

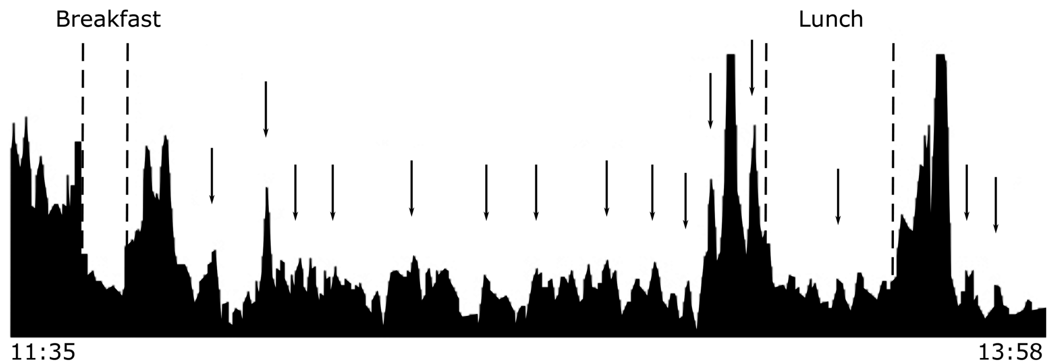


Fig. 8. Wrist motion activity during a day. Automatically detected peaks used to segment the data for classification are indicated with arrows. Self-reported eating activities are marked with dashed lines.

The video states the following: ‘[A]n eating activity occurs when consuming food or a beverage is the main activity. For example, if you prepare a cup of coffee or a bowl of cereal and dedicate time to finish that cup of coffee or a bowl of cereal, that is an eating activity. Going to a restaurant and ordering a meal that is finished is considered an eating activity. Grabbing a handful of pretzels or another snack with the intention of finishing them in one go is considered an eating activity. On the other hand, if you prepare coffee in the morning and drink it while preparing food for the day or helping family members during the day, do not mark it as an eating activity. Similarly, casually grazing on snacks while working on other activities is not considered an eating activity to be marked. Lastly, sipping on a beverage while focusing on other activities over a long period of time is not considered an eating activity.” On closing, the video recommends that participants record an activity if they doubt whether it is an eating activity or not.

Participants were asked to wear the device upon waking up the next day and start recording data. They were asked to tap a button on the Shimmer3 at the beginning and end of an eating activity. The video then guided the participants through a simulated meal. Each participant was asked to wear the Shimmer3, turn the device on, and press the button to provide self-reported meal start and stop times while a research assistant observed to verify correct operation. Before leaving, participants were provided with a business card that contained their participant ID, an emergency contact phone number, and summary instructions on device operation.

Subsequent to data collection on the second day, participants met with a research assistant for an exit interview on day 3. The research assistant downloaded motion tracking and button press data from the Shimmer device. Button press timestamps were reviewed with participants to help identify erroneous button presses and to pair start and stop times of eating. The research assistant also collected secondary information on eating activities such as location (e.g., home or restaurant), type (e.g., lunch or dinner), if the eating activity consisted of multiple servings of food (yes or no), if the participant was eating in company (yes or no), what was consumed (open response), and if any secondary activities were being performed while eating (open response). Research assistants visualized the collected data (as shown in Figure 8) and confirmed with participants that the meal activities were seen in the expected times and were of expected lengths. Participants and research assistants were not allowed to erase or change any wrist motion data. Research assistants verified that all expected data files (interview, labels, and wrist motion data) were present by comparing against a provided checklist. All collected data was uploaded to a central repository using file synchronization software (Box [7]).

A total of 408 participants were recorded (61% female, BMI $25.8 \pm 5.8 \frac{kg}{m^2}$, age 28 ± 12 years). Data was collected on all 7 days of the week, and took a total of 7 months to complete due to pauses in data collection caused by holidays and other events. Of these, 351 (86%) completed a recording that was usable. Demographic details of the 351 participants are reported in Table 3. Recordings were unusable for the following reasons: 19 device failures, 4 people forgot to turn it on, 11 people took it off partway through the day, 2 people carried it in

Table 3. Demographics and BMI for Participants in the Free-Living Eating Activity Data Collection

n	
351	
Age	
mean (±SD)	28 ±12 years
Gender	
Female	214 (61%)
Male	137 (39%)
Ethnicity	
Black	69 (20%)
White	205 (58%)
Other	77 (22%)
BMI	
mean (±SD)	25.7 ±5.73 $\frac{kg}{m^2}$

a pocket instead of wearing it, 2 devices were lost or damaged, and 13 people failed to follow instructions. Eight participants failed to turn off their device after taking it off at night, thus showing multiple hours of rest. Data from these participants was manually clipped after visually detecting rest at night to save storage memory requirements. One participant recorded for 3 days and another participant recorded for 2 days, yielding a total of 354 days of usable data. Of the usable data, 18 days did not contain labels or secondary information data due to file sync issues when uploading data to the central repository or error by a research assistant. We paired available button press information to identify labels in these cases and labeled all context information as unknown. In three meals, participants forgot to press the button at the end of a meal. The authors guessed meal end times from the visualized wrist motion signals for these three meals. The total duration recorded was 4,680 hours, containing 265 total hours of self-reported eating across 1,133 separate eating activities (meals, snacks). The average duration recorded per participant was 13.2 hours. The dataset is publicly available at http://www.cecas.clemson.edu/~ahoover/eat-detect.

4.2 Preprocessing

The eating detection algorithm uses features based on linear acceleration. Estimates of linear acceleration are provided by sensor fusion algorithms that track device orientation; however, these algorithms are developed for applications where the acceleration is moderate ($\pm 2g$) [15, 24]. For the Shimmer3, the estimates are provided by the embedded Invensense MPU-9150 chip [27]. Previous work has shown that wrist acceleration is much lower than other more common applications ($\pm 0.2g$) and disproportionately affected by residual errors that remain in linear acceleration estimates [24]. To mitigate these errors, we apply a high-pass filter by subtracting the average linear acceleration value over a sliding 1-minute window from each datum, as suggested in the work of Sharma and Hoover [24].

The resulting linear acceleration and angular velocity signals $R_t = \{l_x, l_y, l_z, \omega_\phi, \omega_\theta, \omega_\psi\}$ at time index t are filtered using a standard Gaussian kernel with the same parameters as those described in Section 3.2. The result is smoothed signals $S_t = \{S_{x,t}, S_{y,t}, S_{z,t}, S_{\phi,t}, S_{\theta,t}, S_{\psi,t}\}$.

4.3 Segmentation

Previous work by Dong et al. [11] showed that wrist motion activity peaks before and after meals with peaks being in temporal proximity to self-reported meal start and end times [11]. It is assumed that these peaks are caused by actions before an eating activity commonly related to meal preparation, and those after eating related to the cleaning up of left over food, utensils, or the area where food was consumed. Figure 8 shows an example of the same pattern seen in our new data collection. Wrist motion activity (magnitude of acceleration) is plotted on the Y axis versus time on the X axis.

Data is segmented as periods between peaks. We use the same method for peak detection described in the work of Dong et al. [11], which uses a hysteresis approach to find local maxima in the sum of linear acceleration. Peaks detected by this algorithm are marked by arrows in Figure 8.

4.4 Classification Features

Peak-to-peak segments are classified as walking, resting, eating, or other. Six features are calculated for each peak-to-peak segment. Two features characterize resting and walking, whereas four features characterize eating and were first introduced by Dong et al. [11]. The first feature f_1, the rate of zero crossings, and f_2, the amount of rest in a segment, are described in Section 3.

The third feature f_3 is called manipulation and measures the ratio of wrist rotation to linear motion:

$$f_{3,w} = \frac{1}{W} \sum^{W} \frac{|S_{\phi,t}| + |S_{\theta,t}| + |S_{\psi,t}|}{|S_{x,t}| + |S_{y,t}| + |S_{z,t}|}, \tag{5}$$

where $f_{3,w}$ is the value of the manipulation feature for the segment with time span W (number of samples), $S_{x,t}$, $S_{y,t}$, $S_{z,t}$ are the smoothed linear acceleration values for the respective axes, and $S_{\phi,t}$, $S_{\theta,t}$, $S_{\psi,t}$ are the smoothed angular velocities (yaw, pitch, roll) from the gyroscope. The fourth feature, linear acceleration, is calculated as

$$f_{4,w} = \frac{1}{W} \sum^{W} |S_{x,t}| + |S_{y,t}| + |S_{z,t}|. \tag{6}$$

Wrist roll motion is calculated as

$$f_{5,w} = \frac{1}{W} \sum^{W} |S_{\phi,t} - \frac{1}{W} \sum^{W} S_{\phi,t}|. \tag{7}$$

Regularity of wrist roll motion is calculated as

$$f_{6,w} = \frac{1}{W} \int_{W} 1 \; \forall \; t \in [\, |S_{\phi,t}| > 10^{\circ} \ldots \; t + 8\text{sec} \,]. \tag{8}$$

This feature represents the percentage of time the wrist was rolling, and takes a value between 0 and 1. This time is calculated as the amount of time the wrist roll was at least 10 deg/sec, plus the next 8 seconds after the wrist roll reduces to less than 10 deg/sec. The values 8 sec and 10 deg/sec were tuned in the work of Dong et al. [11].

4.5 Classification

We use a two-stage classifier. In the first stage, thresholds determine if a segment is walking or resting. In the second stage, remaining segments are classified using a Bayesian classifier as eating or other. A segment is considered as walking if $f_{1,w} \geq T_1$. Similarly, a segment is considered to be resting if $f_{2,w} \geq T_2$. The nature of peak-to-peak segmentation allows this method to only label segments as walking or resting if they are sufficiently long and largely walking or resting.

Segments not considered walking or resting are then labeled as eating or other by a naive Bayesian classifier that assumes independence of features. This classifier assigns a class $c_i \in C$ to a segment given feature values f_j

as shown in Equation (9).

$$c_i = \arg\max_c P(c_i) \prod_j P(f_j|c_i) \tag{9}$$

We have only two classes for the Bayesian classifier: eating (c_3) and other (c_4). We tested different values of prior probabilities $P(c_3)$ and $P(c_4)$, and found the best balance between eating and non-eating detection at $P(c_3) = P(c_4) = 0.5$. A normal distribution was used to calculate the probabilities of segment belonging to these classes given their feature values as shown in Equation (10):

$$P(f_j|c_i) = \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left(-\frac{(f_j - \mu_{i,j})^2}{2\sigma_{i,j}^2}\right) \tag{10}$$

where $\mu_{i,j}$ is the mean of feature j in class i and $\sigma_{i,j}^2$ is the variance. Collectively, the two-stage classifier can be stated as Equation (11).

$$c_i = \begin{cases} 1\ if\ f_{1,W} \geq T_1 \\ 2\ if\ f_{2,W} \geq T_2 \\ \underset{c}{argmax}\ P(c_i) \prod_{j=3}^{6} P(f_j|c_i)\ otherwise \end{cases}, \tag{11}$$

4.6 Parameter Tuning

Eating activity classification uses normal distributions trained from the free-living datasets. We use standard leave one out cross validation for training and testing. For feature values of the other class, data labeled other was split into 5-minute segments and feature values were calculated. For the eating class, feature values were calculated for time periods between the self-reported meal start and end times.

4.7 Evaluation Metrics

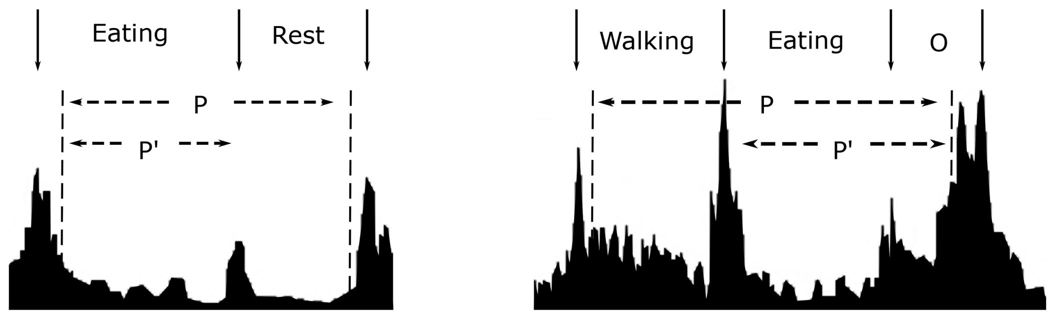
We use two metrics to evaluate our methods: an activity level recall and a per-second metric. The activity level recall evaluates how many self-reported eating activities intersected with a machine detected meal and answers the question “Can eating activities be detected?” The per-second metrics evaluate how many seconds of eating were correctly classified, answering the question “Can we detect the time period during which someone is eating?”

Many metrics, such as precision, the F1 score, Cohen’s kappa, and Mathews correlation, are known to be affected by class imbalance [26, 31, 32]. Eating occurs far less frequently than not eating, so the detection of periods of eating is by default an imbalanced problem [11], and these metrics result in different values for different datasets. We therefore evaluate our classifier using weighted accuracy [11, 17] to accommodate the imbalance in occurrence of eating vs other:

$$WACC = \frac{TP \times 20 + TN}{P \times 20 + N}, \tag{12}$$

where $WACC$ is the weighted accuracy, P (positives) represents the total number of seconds in self-reported meals, N (negatives) represents the total seconds detected non-eating, TP (true positives) represents the number of seconds classified as eating inside self-reported meal times, and TN (true negatives) represents the number of seconds classified as other outside self-reported meal times.

In the preceding equation, P (positives) includes secondary activities that may have occurred during a self-reported meal. When these secondary activities are detected correctly as non-eating, they are considered false negatives. Consider the two example meals shown in Figure 9. Although walking and resting are correctly detected, Equation (12) penalizes the classifier for not classifying segments as eating that were in fact not eating.



(a) (b)

Fig. 9. Wrist motion data examples. Peaks are marked with arrows, and self-reported meal boundaries are marked as dashes. Machine-detected labels are indicated between segments, with O indicating other. P contains all data in a self-reported segment, whereas P′ excludes segments identified as walking or resting.

Table 4. Amount of Walking Per Subject in the Pedometer Dataset and Per Meal in the Cafeteria Dataset

Dataset	Walking Labeled	Walking Detected
Pedometer (per subject)	100%	100%
Cafeteria (per meal)	0%	0.3%

We adjust for this by using P' instead of P:

$$WACC' = \frac{TP \times 20 + TN}{P' \times 20 + N}, \tag{13}$$

where P' is the number of seconds in self-reported meals with walking and resting removed—for instance, $P' = P - (\text{Walking} + \text{Resting})$ (Figure 9), and $WACC'$ is the adjusted weighted accuracy.

5 RESULTS

In this section, we first provide evidence on the prevalence of secondary activities in free-living meals. We then show the performance of the walking and resting detectors on the pedometer (walking) and cafeteria (eating) datasets for which ground truth video is available. Having confidence on the performance of these detectors, we show how much walking and resting is seen during free-living eating (Table 6). We discuss the performance of the classification algorithm, and how adjusting for secondary activities like walking and resting affects its accuracy.

5.1 Walking and Resting

We evaluated the walking and resting classifiers on 1-minute segments in the pedometer and cafeteria datasets. These datasets contain 4.7 hours of walking (pedometer) and 96 hours (cafeteria) of data collected from 30 (pedometer) and 271 (cafeteria) participants. In the pedometer data, all 1-minute segments (100%) were correctly classified as walking. In the cafeteria data, 0.3% of 1-minute segments were classified as walking (there was no walking in this dataset). This has been reported in Table 4. The rest detector detected no 1-minute segments as rest in the pedometer datasetbut detected 7% of the 1-minute segments in the cafeteria dataset as rest. On average, human raters labeled 6.8% of time in cafeteria meals as resting. Both detectors have a slight false-positive rate. This is shown in Table 5. We conclude that the classifiers are reliable enough to be used on the free-living dataset in our second experiment.

Table 5. Amount of Resting Per Subject in the Pedometer Dataset and Per Meal in the Cafeteria Dataset

Dataset	Resting Labeled	Resting Detected
Pedometer (per subject)	0%	0%
Cafeteria (per meal)	6.8%	7.0%

Table 6. Walking and Resting Occurred Much More Often in Free-Living Meals Compared to the Cafeteria

Dataset	Eating Hours	Meals	Walking Detected/Meal	Resting Detected/Meal
Cafeteria	96	518	0.3%	7%
Free living	249	1,133	5.5%	12.8%

Table 7. Secondary Activity Groups for the Dataset and the Number of Meals in Each Group

Activity	Just eating	Talking	Watching	Working	Walking	In a car	Unknown
Meals	317 (28%)	281 (25%)	223 (20%)	210 (19%)	20 (2%)	19 (2%)	63 (6%)

We evaluated the 1,133 self-reported meals in the free-living dataset using our walking and resting detectors to assist with cleaning the dataset before analysis. Of the 1,133 self-reported eating activities, 21 (1.6%, 3.5 hours) were classified as walking and 38 (3.4%, 12 hours) were classified as resting by the walking and resting detectors. After examining each of these meals individually, we concluded that secondary activities dominated the wrist motions so much that consumption could not be seen. We removed the labels of eating from these periods of time and instead marked them as periods of non-eating. We also marked 9 meals shorter than 1 minute and 2 meals where the Shimmer3 failed to record valid data for more than half of the self-reported duration of the meal as non-eating. A total of 70 of 1,133 meals (6.2%, 16 hours) were marked as non-eating.

In the free-living dataset, we found that participants walked 5.5% of the time during self-reported meals (Table 6). Walking virtually never happens in controlled laboratory experiments involving eating. If a classifier was trained on laboratory data and then deployed to free living, it could be expected that the presence of walking would be new and likely to reduce the accuracy of detecting eating activities. We detected participants resting an average of 12.8% of the time during free-living meals, whereas in cafeteria meals, resting was detected 7% of the time (average). It is likely that participants take longer to eat as they rest more during free living than in a controlled setting. This might happen because they are conducting passive secondary activities such as watching television. This again could help explain why a classifier trained on laboratory data might have lower accuracy when tested on free-living data.

5.2 Secondary Activities in Free-Living Eating

Open responses on the secondary activities for each meal were coded into seven groups: just eating, talking, watching, working, walking, in a car, and unknown. Of the 1,133 self-reported meals, we labeled 317 (28%) as just eating (no secondary activity was performed), 281 (25%) as talking (participants were talking to someone else), 223 (20%) meals as watching (participants were watching television, in a movie theater, or listening to a lecture), 210 as working (participants state they were working, reading, or using a device), 19 as in a car (participants were driving or riding in a car), 20 as walking (participants were walking while eating), and 63 meals as unknown (activity information was not collected during an interview or was lost). These numbers are reported in Table 7. A short list of some of the activities is shown in Table 8. Of note are activities during

Table 8. Some of the Activities Reported During a Meal

Putting away groceries	Playing with niece	Walking dog
Watching media on computer	Sitting in meeting	House chores
Checking email and internet	Working on papers	Reading book
Talking and playing pool	Sitting in car	Talking in car
Talking to team in class	Getting ready	Doing work
Talking to friend in a fair	Playing trivia	Working
Talking and texting	Watching TV	Reading

Table 9. Confusion Matrix Showing Time Classified by the Eating Activity Classifier in Hours

		Labeled Class	
		Eating	**Non-Eating**
Predicted Class	**Eating**	190 hours	1186 hours
	Non-Eating	47 hours	3245 hours

eating previously unseen in the literature like putting away groceries, playing with a niece, walking a dog, or house chores. This list demonstrates the breadth of complexity of the free-living dataset in representing free-living behavior.

5.3 Eating Activity Classification

Of the 1,063 remaining meals, 946 triggered a positive detection (89%) and 117 meals were missed. A total of 4,966 false positives were triggered (5 false positives for every true positive). Although this number seems high, it is important to note that the classes are imbalanced, as humans eat only 5% of the time. A 20:1 class imbalance causes challenges in balancing false positives and false negatives, in that if false positives occur equally with false negatives, the method could have a low false-positive rate while completely missing all actual meals. For example, suppose that there were only 3 false-positive meals per day and 3 false-negatives meals; assuming only 3 meals were consumed that day, they would all have been missed. With a high imbalance in data, accuracy must be balanced.

Although the method of Dong et al. [11] performs with a weighted accuracy of 74% on our free-living dataset, our new method that detects secondary activities like walking and resting improves weighted accuracy to 77%. Table 9 shows the confusion matrix for the new method (in hours). Weighted accuracy had a mean of 75% per participant and a median of 78%. Approximately 72% of the participants had weighted accuracy of 70% or higher. Paired samples t-test results show that the change in accuracy per participant due to walking and rest detection is significant ($t[353] = 7.86$, $p < 0.001$), largely due to a 23% reduction in the number of false negatives.

To evaluate the impact of secondary activities on the classification of meals, we split all meals into two categories of secondary activity: yes or no. Secondary activities included anything that could affect wrist motion for significant periods of time such as working or driving. Talking occurs frequently during free-living meals and may involve some wrist motions caused by gesturing, but we hypothesize that these wrist motions are relatively infrequent compared to those used for consumption. Therefore, meals with descriptions of secondary activities that included only talking were grouped with meals with no secondary activities. Table 10 shows the results. Meals with no secondary activity were detected more frequently (91% vs 87%), and the total duration of eating conducted with no secondary activity was detected better than the total duration with secondary activities (83% vs 78%).

Table 10. Effect of Secondary Activities on Meal Detection

Secondary Activity	No. of Meals	Duration Detected	Meals Detected
No	565	83%	91%
Yes	498	78%	87%

Table 11. Average Time Difference Between Logged Meal Start and End Times, and Nearest Peaks in Minutes

Dataset	Start Time Difference (min)	End Time Difference (min)
Dong et al. [11]	−0.6	1.5
Free living (this work)	−4.5 ± 14	7.3 ± 11

The presence of secondary activities also affects the presence of wrist activity peaks before and after a meal [11]. Peaks are used to segment the data before classification, and thus an important part of the method to detect eating activities. Although Dong et al. [11] reported average differences of −0.6 minutes and 1.5 minutes between peaks and the start and end of self-reported meals, respectively, we see average differences of −4.5 ± 14 and 7 ± 11 minutes (Table 11) between self-reported start and stop times and the detected peaks. The reason for this change is likely the presence of intermittent secondary activities during free-living meals, enabled by the low weight of the Shimmer3 (24g) compared to the iPhone used by Dong et al. [11] that weighs 140g.

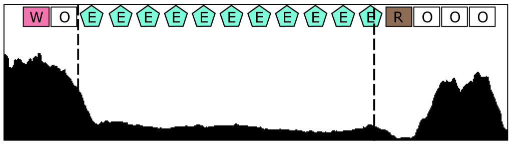
5.4 The Anatomy of a Free-living Meal

We visualized the effect of secondary activities on meals with the help of a 1-minute classifier. Figure 10(a) shows a 10-minute meal recognized as all eating. This is similar to what could be expected in controlled experiments where participants consume food without conducting secondary activities. Figure 10(b) shows an 11-minute meal in which the majority of the time is recognized as eating, but a few minutes are recognized as other. It is likely that the participant engaged in brief secondary activities. Figure 10(c) shows a 19-minute meal in which less than half the duration is recognized as eating. It is likely that the participant engaged in passive secondary activities such as watching television, as is evident from the long periods of rest. Figure 10(d) shows a 30-minute meal in which only 3 minutes are recognized as eating. It is likely that the participant engaged in multiple secondary activities, including passive (as is evident from the long periods of rest) and active (as is evident from the long periods of other). Figure 10(e) shows a 34-minute meal in which long periods of time were recognized as walking. It is likely that the participant was multi-tasking, such as walking around and doing chores while simultaneously eating. It is important to note that all of these are conjecture and that we do not have minute-level ground truth of secondary activities.

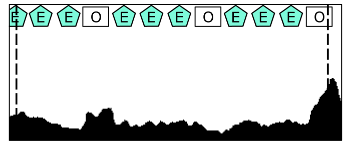
6 DISCUSSION

To the best of our knowledge, the Shimmer3 dataset collected for this work is the largest of its kind, consisting of 4,680 hours of free living, containing 1,133 meals/snacks, from 351 participants. Analysis of self-reported descriptions of activities during eating indicated that 72% of all meals were consumed while performing a secondary activity (25% talking, 47% other). By detecting two common secondary activities, walking and resting, we were able to more accurately detect eating events, demonstrating the need to account for secondary activities in free-living experiments on eating detection.

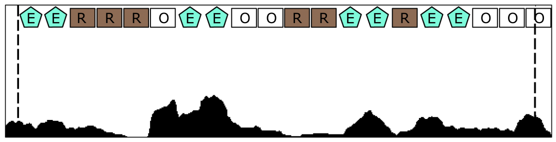
Some previous works (see Table 1) have seen a drop in performance when transitioning from the laboratory to free living. We detected resting 7% of the time in cafeteria meals while detecting resting 12.8% of the time in



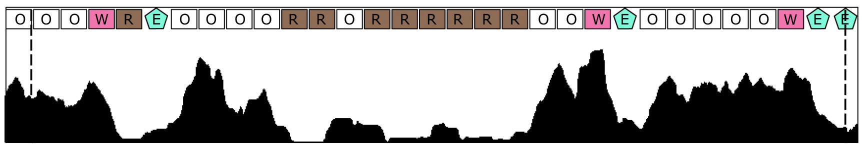
(a) 10 minute meal with no secondary activity.



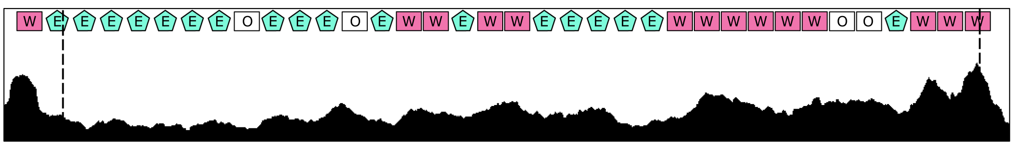
(b) 11 minute meal recognized as mostly eating.



(c) 19 minute meal with <50% duration recognizied as eating.



(d) 30 minute meal showing large resting and walking, suggesting secondary activities.



(e) 34 minute meal showing large amounts of walking, suggesting secondary activities.

Fig. 10. Five example meals from our free-living dataset, in increasing order of complexity. The bar on top of each figure shows majority activity detected for 1-minute segments (walking (W), resting (R), eating (E), or other secondary activity (O)). Meal start and end times are indicated by vertical dashed lines.

self-reported free-living meals. Similarly, there was no walking during cafeteria meals, whereas walking was detected 5.5% of the time during self-reported free-living meals. We show that our new classifier that uses walking and resting detectors improves the performance of detecting eating activities (weighted accuracy) by 3%. The practical implication of this improvement is that eating episodes in free living look different from eating episodes in a cafeteria. Besides the growing evidence that laboratory meals may not be representative of free-living meals, the “in-between” environment of a cafeteria may also not be representative of free-living meals due to the differences in secondary activity distributions. Another practical implication is that it is possible to detect subsets of an eating episode that are secondary activities. Although we only modeled two types of secondary activity, our experiment demonstrates that is possible to improve accuracy in this manner.

It may be possible to develop a taxonomy of classes using secondary activities such as eating+reading, eating+texting, eating+watching, and so on, and then develop a classifier for each of these classes. However, this would need more data, even more than our current dataset contains. Secondary activities do not necessarily occur steadily through an entire eating episode; instead, they occur for subsets of time within an eating episode. Given this knowledge, we believe that a better approach to detecting eating activities might be to design a classifier that models an eating episode as sequence of varying combinations of activities. This is a topic for future work.

In our work, of 408 participants recruited, only 351 participants (86%) provided useable data because the participants failed to comply with instructions or the device failed. This provides some insight into how often a wrist-worn device might be expected to fail or be used improperly.

The activities considered secondary in this article are specific to wrist-based sensing modalities. However, secondary activities would likely impact experiments based on other modalities. For example, methods based on sound detection at the throat [3] could consider long periods of talking a secondary activity conducted during what a person considers part of a meal. Driving and walking could affect many sensor modalities via background noise and rhythmic motions. Meals eaten while conducting house chores or simultaneously preparing to leave the home during the morning could also affect other modalities.

As reported in Table 11, the difference between peaks and meal boundaries was much larger in our new dataset compared to that in the work of Dong et al. [11]. Some participants would rest for extended periods of time after eating or worked on a computer. Since the segmentation method used is based on high-intensity motion before and after a meal, it sometimes misses activities surrounded by rest. We hypothesize that this segmentation method impacts the classification performance of the algorithm, and we plan to investigate other segmentation methods in future work. This issue, and the analysis of free-living meals, also shows that metrics that relate to time (e.g., weighted accuracy, or error in detected meal start or start times compared to self-reported start and stop times) might not perform well for the free-living eating activity dataset, and researchers should consider what metrics are appropriate for the specific task they are modeling, such as meal detection or meal segmentation.

One limitation of this work is that only coarse descriptions of secondary activities during meals were captured (e.g., "standing," "talking to friends in a classroom," "watching netflix"). Ground truth on *when* walking or resting happened during a meal was not collected due to the difficulty of collecting such information during free living. However, our tests on the cafeteria and pedometer datasets showed very high accuracy in detecting walking and resting compared to video-based ground truth. We also specifically chose walking and resting as secondary activities due to the distinctiveness of their motion patterns. Resting can be detected by a lack of sensor motion, and walking can be detected by its rhythmic motion. Although the resting and walking detectors have not been validated on free-living datasets, we believe that the combination of large datasets and readily discernible differences in the motion patterns of these activities provides confidence in the translation of our walking and resting detectors from the semi-controlled datasets to the free-living dataset. For future work, we believe that it will be important to collect not only the types of secondary activities conducted during eating in free living but also exactly when they happen.

Another limitation of this work is the use of a previously established eating activity classifier [20] and only one modality of sensing—wrist motion. More sophisticated approaches may help improve the detection of eating; however, we believe that the issue of secondary activities would affect all types of classifiers. This is supported by previous work that detected eating using audio from a contact microphone. The authors learned that talking while walking or moving excessively can be misclassified as eating [6]. In addition, eating may be affected by other factors such as culture or personal behavior. Eating behavior may be socially driven or individualistic and may also change over time. Our current dataset only contains 1 day of data from each participant, limiting the analysis of these factors. In future work, we plan to use the lessons learned from this work to collect multiple days of data from multiple participants so that these factors can be considered.

To conclude, by analyzing a very large free-living dataset, we learned that secondary activities during meals are common and might not be captured in controlled or semi-controlled environments due to the lab coat effect when participants know they are being observed or video recorded [1]. In the future, we recommend that researchers design laboratory or semi-controlled studies keeping secondary activities in mind. We show that one way to address the phenomena of secondary activities is to augment a classifier to recognize these activities. An alternative may be to model eating as a multi-class activity rather than a single class while using classifiers like a neural network or a support vector machine. These are topics for future work.

REFERENCES

[1] Rawan Alharbi, Tammy Stump, Nilofar Vafaie, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2018. I can't be myself: Effects of wearable cameras on the capture of authentic behavior in the wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 90.

[2] Nabil Alshurafa, Haik Kalantarian, Mohammad Pourhomayoun, Jason J. Liu, Shruti Sarin, Behnam Shahbazi, and Majid Sarrafzadeh. 2015. Recognition of nutrition intake using time-frequency decomposition in a wearable necklace using a piezoelectric sensor. *IEEE Sensors Journal* 15, 7 (2015), 3909–3916.

[3] Oliver Amft, Martin Kusserow, and Gerhard Troster. 2007. Automatic identification of temporal sequences in chewing sounds. In *Proceedings of the 2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM'07).* IEEE, Los Alamitos, CA, 194–201.

[4] Oliver Amft and Gerhard Tröster. 2008. Recognition of dietary activity events using on-body sensors. *Artificial Intelligence in Medicine* 42, 2 (2008), 121–136.

[5] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 37.

[6] Shengie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, et al. 2018. Auracle: Detecting eating episodes with an ear-mounted sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 92.

[7] Box Inc.2019. Box: Content Management, Workflow, and Collaboration. Retrieved December 2, 2019 from https://www.box.com/home [Online; accessed 2-December-2019].

[8] Keum San Chun, Sarnab Bhattacharya, and Edison Thomaz. 2018. Detecting eating episodes by tracking jawbone movements with a non-contact wearable sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 4.

[9] Guang Deng and L. W. Cahill. 1993. An adaptive Gaussian filter for noise reduction and edge detection. In *Proceedings of the 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*. IEEE, Los Alamitos, CA, 1615–1619.

[10] Bo Dong, Subir Biswas, Robert Gernhardt, and Janik Schlemminger. 2013. A mobile food intake monitoring system based on breathing signal analysis. In *Proceedings of the 8th International Conference on Body Area Networks*. 165–168.

[11] Yujie Dong, Jenna Scisco, Mike Wilson, Eric Muth, and Adam Hoover. 2014. Detecting periods of eating during free-living by tracking wrist motion. *IEEE Journal of Biomedical and Health Informatics* 18, 4 (2014), 1253–1260.

[12] Abul Doulah, Xin Yang, Jason Parton, Janine A. Higgins, Megan A. McCrory, and Edward Sazonov. 2018. The importance of field experiments in testing of sensors for dietary assessment and eating behavior monitoring. In *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'18).* IEEE, Los Alamitos, CA, 5759–5762.

[13] Muhammad Farooq, Abul Doulah, Jason Parton, Megan A. McCrory, Janine A. Higgins, and Edward Sazonov. 2019. Validation of sensor-based food intake detection by multicamera video observation in an unconstrained environment. *Nutrients* 11, 3 (2019), 609.

[14] Juan M. Fontana, Muhammad Farooq, and Edward Sazonov. 2014. Automatic ingestion monitor: A novel wearable device for monitoring of ingestive behavior. *IEEE Transactions on Biomedical Engineering* 61, 6 (2014), 1772–1779.

[15] Sebastian O. H. Madgwick, Andrew J. L. Harrison, and Ravi Vaidyanathan. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *Proceedings of the 2011 IEEE International Conference on Rehabilitation Robotics.* IEEE, Los Alamitos, CA, 1–7.

[16] Ryan Mattfeld, Elliot Jesch, and Adam Hoover. 2017. A new dataset for evaluating pedometer performance. In *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM'17)*. IEEE, Los Alamitos, CA, 865–869.

[17] Mark Mirtchouk, Drew Lustig, Alexandra Smith, Ivan Ching, Min Zheng, and Samantha Kleinberg. 2017. Recognizing eating from body-worn sensors: Combining free-living and laboratory data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 85.

[18] Eric Muth, Ryan Mattfeld, and Adam Hoover. 2017. Bites as a unit of measurement. In *Advances in the Assessment of Dietary Intake.*CRC Press, Boca Raton, FL, 149–161.

[19] Eric R. Muth and Adam Hoover. 2017. Designing mobile health technologies for self-monitoring: The bite counter as a case study. In *Mobile Health*. Springer, 101–120.

[20] Raul I. Ramos-Garcia, Eric R. Muth, John N. Gowdy, and Adam W. Hoover. 2015. Improving the recognition of eating gestures using intergesture sequential dependencies. *IEEE Journal of Biomedical and Health Informatics* 19, 3 (2015), 825–831.

[21] James N. Salley, Adam W. Hoover, Michael L. Wilson, and Eric R. Muth. 2016. Comparison between human and bite-based methods of estimating caloric intake. *Journal of the Academy of Nutrition and Dietetics* 116, 10 (2016), 1568–1577.

[22] Edward Sazonov, Stephanie Schuckers, Paulo Lopez-Meyer, Oleksandr Makeyev, Nadezhda Sazonova, Edward L. Melanson, and Michael Neuman. 2008. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. *Physiological Measurement* 29, 5 (2008), 525.

[23] Giovanni Schiboni and Oliver Amft. 2018. Sparse natural gesture spotting in free living to monitor drinking with wrist-worn inertial sensors. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers.* ACM, New York, NY, 140–147.

[24] Surya Sharma and Adam Hoover. 2019. A study on linear acceleration of the wrist during free-living. In *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM'19)*. IEEE, Los Alamitos, CA.

[25] Yiru Shen, James Salley, Eric Muth, and Adam Hoover. 2017. Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables. *IEEE Journal of Biomedical and Health Informatics* 21, 3 (2017), 599–606.

[26] Julius Sim and Chris C. Wright. 2005. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 85, 3 (2005), 257–268.

[27] TDK Invensense. 2018. MPU-9150. Retrieved August 31, 2020 from https://www.invensense.com/products/motion-tracking/9-axis/mpu-9150.

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

[29] Edison Thomaz, Aman Parnami, Jonathan Bidwell, Irfan Essa, and Gregory D. Abowd. 2013. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, NY, 739–748.

[30] U.S. Department of Labor. 2016. Bureau of Labor Statistics Data for Activity: Eating and Drinking. Retrieved August 31, 2020 from https://data.bls.gov/timeseries/TUU10101AA01013237.

[31] Wikipedia Contributors. 2019. Confusion Matrix—Table of Confusion. Retrieved August 6, 2019 from https://en.wikipedia.org/wiki/Confusion_matrix#Table_of_confusion [Online; accessed 6-August-2019].

[32] Wikipedia Contributors. 2019. Precision and Recall—Imbalanced Data. Retrieved August 6, 2019 from https://en.wikipedia.org/wiki/Precision_and_recall#Imbalanced_Data [Online; accessed 6-August-2019].

[33] Rui Zhang and Oliver Amft. 2017. Monitoring chewing and eating in free-living using smart eyeglasses. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (2017), 23–32.

[34] Shibo Zhang, Rawan Alharbi, Matthew Nicholson, and Nabil Alshurafa. 2017. When generalized eating detection machine learning models fail in the field. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2017 ACM International Symposium on Wearable Computers*. ACM, New York, NY, 613–622.

Received August 2019; revised February 2020; accepted April 2020