

# A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing

Edison Thomaz, Irfan Essa, Gregory D. Abowd

School of Interactive Computing  
Georgia Institute of Technology  
Atlanta, Georgia, USA

## ABSTRACT

Recognizing when eating activities take place is one of the key challenges in automated food intake monitoring. Despite progress over the years, most proposed approaches have been largely impractical for everyday usage, requiring multiple on-body sensors or specialized devices such as neck collars for swallow detection. In this paper, we describe the implementation and evaluation of an approach for inferring eating moments based on 3-axis accelerometry collected with a popular off-the-shelf smartwatch. Trained with data collected in a semi-controlled laboratory setting with 20 subjects, our system recognized eating moments in two free-living condition studies (7 participants, 1 day; 1 participant, 31 days), with F-scores of 76.1% (66.7% Precision, 88.8% Recall), and 71.3% (65.2% Precision, 78.6% Recall). This work represents a contribution towards the implementation of a practical, automated system for everyday food intake monitoring, with applicability in areas ranging from health research and food journaling.

## Author Keywords

Activity recognition; Food Journaling; Dietary Intake; Automated Dietary Assessment; Inertial Sensors

## ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

## INTRODUCTION

Dietary habits have been studied by health researchers for many decades, and it is now well-understood that diet plays a critical role in overall human health [19]. To elucidate the mapping between diet and disease, nutritional epidemiologists have typically relied on validated dietary assessment instruments driven by self-reported data including food frequency questionnaires and meal recalls [33]. Unfortunately, these instruments suffer from several limitations, ranging from biases to memory recollection issues [15, 22]. For this

reason, over the last 15 years, a large body of research has aimed at fully automating the task of food intake monitoring [11, 12, 21, 34]. Although significant progress has been achieved, most proposed systems have required individuals to wear specialized devices such as neck collars for swallow detection [2], or microphones inside the ear canal to detect chewing [20]. These form-factor requirements have severely limited the immediate practicality of automated food intake monitoring in health research.

There are two key technical challenges in building a fully automated food intake monitoring system: (1) recognizing *when* an individual is performing an eating activity, and then (2) inferring *what* and *how much* the individual eats. In this paper we focus on recognizing when an *eating moment* is taking place, which includes having a sit-down meal with utensils, eating a sandwich, or having a snack.

Our aim with this work is to explore a *practical* solution for eating moment detection; we describe an approach leveraging the inertial sensor (3-axis accelerometer) contained in a popular off-the-shelf smartwatch. This approach contrasts with methods that require either multiple sensors or specialized forms of sensing.

Our eating moment recognition method consists of two steps. First, we perform food intake gesture spotting on the stream of inertial sensor data coming from the smartwatch, reflecting arm and hand movements. Secondly, we cluster these gestures across the time dimension to unearth eating moments. To evaluate our approach, we first ran a formative study with 20 participants to validate our experimental design protocol and instrumentation. Informed by this pilot, we conducted user studies that resulted in three datasets, (1) a laboratory semi-controlled study with 20 participants, (2) an in-the-wild study with 7 participants, and (3) 422 hours of in-the-wild data for one participant collected over the course of 31 days.

The contributions of this work are:

- A practical system for eating moment estimation leveraging the inertial sensor (3-axis accelerometer) of a popular off-the-shelf smartwatch.
- An evaluation of a lab-trained eating moment classification model in-the-wild with two datasets: 7 participants over one day (76.1% F-score, 66.7% Precision, 88.8% Recall), and one participant over 31 days (71.3% F-score, 65.2%

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

UbiComp '15, September 7–11, 2015, Osaka, Japan.

Copyright 2015 © ACM 978-1-4503-3574-4/15/09...\$15.00.

<http://dx.doi.org/10.1145/2750858.2807545>

Precision, 78.6% Recall). The model was tested on its ability to recognize eating moments every 60 minutes.

- An anonymized and annotated dataset of 3-axis accelerometer sensor data collected from a smartwatch. It comprises data gathered in the laboratory and in-the-wild studies.

## MOTIVATION

Today, dietary intake self-reporting is the gold standard when it comes to methods for studying the mapping between diet and disease, energy balance, and calorie intake. Although self-reports have been validated and used for decades, health researchers have long known that self-reported data is fraught with weaknesses, such as biases and memory recollection issues [15, 22]. Recently, there has been a stronger sentiment in the health research community that more resources need to be allocated towards the development of more objective and precise measures [9, 23]. Some have even questioned the validity of the National Health and Nutrition Examination Survey (NHANES) data throughout its 39-year history [4].

The need for improved dietary assessment is also shared by individuals interested in meeting health goals. Recently, health concerns linked to dietary behaviors such as obesity and diabetes have fueled demand for dietary self-monitoring, since it is one of the most effective methods for weight control [6]. However, adherence to dietary self-monitoring is poor and generally wanes over time [5], even with modern smartphone-based systems such as MealSnap<sup>1</sup> and MyFitnessPal<sup>2</sup> [8].

Semi-automated food journaling is a promising new approach where the food tracking task is split between individuals and an automated system. This method offers a reduction in the manual effort involved in food logging while keeping individuals aware of foods consumed. A critical requirement in semi-automated dietary monitoring is the *identification of when an eating moment is taking place*, which is exactly the focus of our work.

There are many scenarios that illustrate how a semi-automated food journaling system could be used. For instance, if individuals are wearing a camera such as the one in Google Glass, the recognition of an eating moment could automatically trigger a reminder to capture a relevant food photo. If an electronic food diary is being used, a new entry could be automatically created at the time and location of the recognized eating moment. Finally, individuals could be sent an SMS message at an opportune time later in the day prompting for details about an inferred eating moment.

A practical and reliable automated food intake monitoring system would represent a breakthrough for health researchers and individuals looking to improve dietary habits. This work addresses technical challenges towards realizing this goal.

<sup>1</sup><http://www.mealsnap.com>

<sup>2</sup><http://www.myfitnesspal.com>

## RELATED WORK

Research in the area of activity recognition around eating activities dates back to the 1980s when researchers tried to detect chews and swallows using oral sensors in order to measure the palatability and satiating value of foods [29]. Ongoing research work in this area ranges from the use of crowdsourcing techniques [24], instrumented objects [17], wearable cameras, acoustic sensing, and inertial sensing.

The key advantage of lightweight wearable sensors for food monitoring is that individuals are free to move amongst different locations and eat anywhere since they are carrying the system with them at all times. In other words, they are not restricted to the infrastructure in the built environment. On the other hand, to have practical value, wearable sensors must meet a number of requirements ranging from battery life, comfort, and social acceptability.

### Acoustic Sensing

Sazonov et al. proposed a system for monitoring swallowing and chewing through the combination of a piezoelectric strain gauge positioned below the ear and a small microphone located over the laryngopharynx [28]. More recently, Yatani and Truong presented BodyScope, a wearable acoustic sensor attached to the user's neck [34]. Their goal was to explore how accurately a large number of activities could be recognized with a single acoustic sensor. The system was able to recognize twelve activities at 79.5% F-measure accuracy in a lab study and four activities (eating, drinking, speaking, and laughing) in an in-the-wild study at 71.5% F-measure accuracy. Cheng et al. also explored the use of a neckband for nutrition monitoring [7].

Recently, Liu et al. developed a food logging application based on the capture of audio and first-person point-of-view images [20]. The system processes all incoming sounds in real time through a head-mounted microphone and a classifier identifies when chewing is taking place, prompting a wearable camera to capture a video of the eating activity. The authors validated the technical feasibility of their method with a small user study.

### Wearable Cameras

The method of observing individuals from first-person point-of-view cameras for overall lifestyle evaluation has been gaining appeal [10]. In this approach, individuals wear cameras that take first-person point-of-view photographs at regular intervals throughout the day (e.g., every 30 seconds), documenting one's everyday activities including dietary intake [25, 30].

Although first-person point-of-view images offer a viable alternative to direct observation, two fundamental problems remain, image analysis and privacy. With regards to image analysis, all captured images must be manually coded for salient content (e.g., evidence of eating activity), and even with supporting tools such as ImageScape [27] and ImageDiet Day [3], the process tends to be tedious and time-consuming. To address this limitation, Thomaz et al. explored crowdsourcing the task of identifying eating activities from first-person photos [31], and Platamate was built to

extract nutritional information from food photographs, also through human computation [24].

### Inertial Sensing

The widespread availability of small wearable accelerometers and gyroscopes has opened up a new avenue for detecting eating activities through *on-body inertial sensing*. Amft et al. have shown eating gesture spotting with a measurement system comprised of five inertial sensors placed on the body (wrists, upper arms and on the upper torso) [2, 16]. Recognition of four gesture types resulted in recall of 79% and precision of 73% in a study with four participants. A key difference between our work and Amft et al.’s is that our system is more practical; it requires only a smartwatch, as opposed to a body sensor array.

Zhang et al. investigated an approach for eating and drinking gesture recognition using a kinematic model of human forearm movements [35]. With accelerometers located on the wrists, features were extracted using an extended Kalman filter, and classification was done with a Hierarchical Temporal Memory network. Results showed a ‘successful rate’ around 87% for repetitive eating activities. The authors were not explicit about which performance measures they used in their evaluation (i.e., what they meant by ‘successful rate’), how many participants took part in the study, and whether the results reflected person-dependent or person-independent findings. Additionally, the study focused exclusively on eating and drinking activities so the system’s ability to differentiate between eating and drinking versus other activities is unclear.

Also with wrist-based inertial sensors, Kim et al. proposed an approach for recognizing “Asian-style” eating activities and food types by estimating 29 discrete sub-actions such as “Taking chopsticks”, “Stirring”, and “Putting in mouth” [18]. In a feasibility study with 4 subjects, the authors obtained an average F-measure of 21% for discriminating all sub-actions. The system performed better when considering only certain classes of sub-actions, but hand actions could not be identified at all. These measurements led the authors to state that the 29 pre-defined sub-actions may not be suitable for the recognition of meals. Our approach is different in two key ways: it is primarily focused on eating moment detection, and it does not require the estimation of any specific sub-actions to infer food intake gestures. Additionally, our system was evaluated in realistic conditions with 8 participants.

Recently Dong et al. put forth a method for detecting eating moments in real-world settings [12, 11]. Our work differs from Dong et al.’s in important ways. Firstly, our method revolves around modeling intake gestures and estimating eating moments from intake gesture temporal densities. In contrast, their strategy is based on a wrist-motion energy heuristic that might be susceptible to multitasking while eating. Secondly, our system collects inertial sensing data from a smartwatch, whereas Dong et al.’s system was evaluated with participants wearing a smartphone on the wrist; it is unclear how much the placement and weight of the phone influenced intake gesture movements. Lastly, from the reported metrics, we believe our system outperforms Dong et al.’s, particularly with regards to false positives in real-world settings. Having said this, it is

Dataset	# Participants	Avg Duration	% Eating
Lab-20	20	31m 21s	48%
Wild-7	7	5hrs 42m	6.7%
Wild-Long	1	31 days	3.7%

**Table 1.** To evaluate our system, we conducted laboratory and in-the-wild studies that resulted in three datasets. The duration for the Lab-20 and Wild-7 datasets above represent average duration across all participants.

difficult to compare results due to differences in evaluation techniques. For example, Dong et al. report accuracies while weighting true positives to true negatives at a ratio of 20:1. We report our results using non-weighted, and thus traditional precision and recall measurements.

Finally, Amft et al. proposed a system for spotting drinking gestures with one wrist-worn acceleration sensor. Based on a study with six users that resulted in 560 drinking instances, the system performed remarkably well, with average of 84% recall and 94% precision [1]. In this work, the authors also attempted to recognize container type and fluid level, and achieved recognition rates over 70% in both cases. Compared to our work, and beyond the clear eating versus drinking recognition distinction, Amft et al. used a more specialized wrist sensor, which was tethered to a laptop. The sensor provided acceleration and gyroscope data. Another important difference is that Amft et al. collected high-quality training data for each participant, and tested the model in a semi-controlled study. In our work we collected training data in a semi-controlled lab setting and evaluated it in completely naturalistic conditions and over multiple weeks for one participant.

### EVALUATION

Our approach for estimating eating moments was evaluated in two contexts, in the lab and in-the-wild. The questions we explored in our analysis were:

- How well does the model recognize food intake gestures and eating moments with data collected in a controlled setting?
- How does a model trained with lab data perform at recognizing eating moments in unseen in-the-wild data?
- What is the temporal stability of eating moment recognition in-the-wild using a model trained with laboratory data?

We conducted three user studies, a laboratory semi-controlled study with 20 participants (Lab-20), an in-the-wild study with 7 participants over the course of one day (Wild-7), and a naturalistic study with one participant where we collected 422 hours of in-the-wild data over a month (Wild-Long). More details about these details are available in Table 1.

### Pilot Study

To evaluate our approach to eating moment detection with wrist-mounted inertial sensors, we first ran a formative study

	P1-P12	P13-P21
<b>Eat (Fork &amp; Knife)</b>	Lasagna	-
<b>Eat (Hand)</b>	Popcorn	Popcorn, Sandwich
<b>Eat (Spoon)</b>	Breakfast Cereal	Rice & Beans
<b>Non-Eating</b>	Watch Trailer	Watch Trailer
	Conversation	Conversation
	Take a Walk	Take a Walk
	Place Phone Call	Brush Teeth Comb Hair

Table 2. In the laboratory study, participants were assigned to one of two activity groups. Some of the activities involved eating different types of food items while others required participants to perform non-eating tasks. The food eating activities were categorized according to eating style, and utensil type.

Activity	Avg Duration
<b>Eat (Fork &amp; Knife)</b>	5m 1s
<b>Eat (Fork/Spoon)</b>	5m 48s
<b>Eat (Hand)</b>	5m 54s
<b>Watch Movie Trailer</b>	3m 47s
<b>Chat</b>	5m 3s
<b>Take a Walk</b>	2m 18s
<b>Place Phone Call</b>	1m 28s
<b>Brush Teeth</b>	3m 54s
<b>Comb Hair</b>	39s

Table 3. This table is showing the average duration of each activity in our laboratory user study across all participants.

with 20 participants to validate our experimental design protocol and instrumentation for the semi-controlled laboratory study. Participants were asked to eat a variety of foods including fruits (e.g., apple), pizza, and snacks of varying sizes and shapes, such as cookies and M&M's. To test the feasibility of food intake gesture spotting from a wrist-mounted inertial sensor, we collected data from a smartphone attached to participants' arm, the same setup employed by Dong et al. [12]. A custom application logged all the sensor data on the phone, and all individuals were continuously video-recorded as they ate the food provided.

The pilot study helped us address a number of issues in our experimental procedures, such as the foods offered to participants, the types of non-eating activities we asked participants to perform, the amount of time in-between activities, and our data annotation process. In particular, after observing participants wearing a smartphone attached to their wrists, it became clear that the device's weight and size could affect participants' arm and hand movements, and thus influence our study results. As a result, we transitioned to a smartwatch platform for data collection.

### Laboratory Study

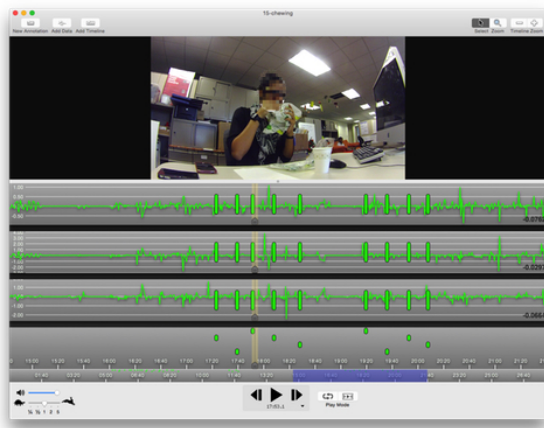
We conducted a user study in our laboratory and examined how our method performed when discriminating between eating and non-eating moments. We recruited 21 participants (13 males and 9 females) between the ages of 20 and 43. All participants were right-handed. Due to a data collection error, we had to discard the data for one of the participants.

The study lasted an average of 31 minutes and 21 seconds and participants were invited to arrive around lunch time, between 11AM and 1PM. Participants were asked to wear the smartwatch on the arm they deemed dominant for eating activities. We did not compensate subjects monetarily, but provided them lunch, which they ate as part of the study itself. Before the activities began, we told them the foods we would be serving and gave them the freedom to eat as much as they wanted. We never had more than one subject participating in the study at a time.

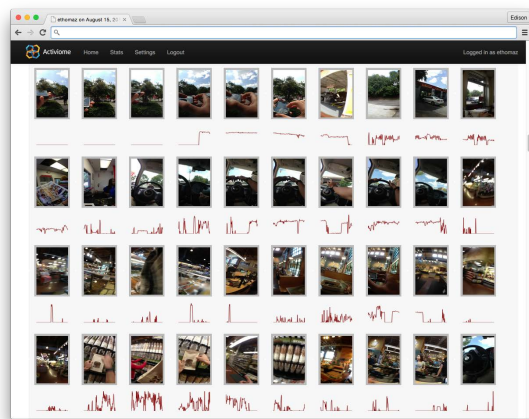
The study was designed so that participants performed a sequence of activities. Participants were assigned to one of two activity groups (Table 2), which contained a mix of eating moments and non-eating activities. The order in which subjects performed these activities varied depending on the activity group. There were no time constraints, and activities were performed in succession without a significant pause in-between. At the end of each activity, except for the last one, the experimenter instructed participating on what to do next. Although this study was scripted and took place in a lab, participants were free to eat completely naturally. Some participants chose to check news and messages on their phone while eating; others were more social, and ate the food provided while having a conversation with the experimenter and others non-participants who happened to be in lab.

The eating moments involved eating different kinds of food, such as rice and beans, and popcorn. For consistency, all foods offered were vegetarian, even though many participants did not have any food restrictions. Subjects were provided with utensils for the activities that required them, and a water-filled cup and napkins were made available to them throughout the study. Although drinking is often linked with food consumption, it was not annotated as an eating moment in this study.

The non-eating activities either required physical movement, or made participants perform hand gestures and motions close to or in direct contact to the head. These activities typically lasted no more than a few minutes, and as little as a few seconds, and were chosen because they are typically performed in daily life and could be confused with food intake in terms of the gestures associated with them. For the "Walking" activity, we asked participants to walk down a hallway, take the stairs down to the floor below, turn around and come back to the study area. The "Phone Call" task involved placing a phone call and leaving a voice message. For the "Comb Hair" and "Brush Teeth" activities, we provided each participant with a hair brush, a tooth brush, toothpaste and they performed these tasks on the spot, with the exception of teeth brushing, which took place in the bathroom.



**Figure 1.** We estimated ground truth by recording each study session with a video camera and then coding the data with the ChronoViz tool.



**Figure 2.** Participants of the in-the-wild study wore a wearable camera that captured photos automatically every minute. After the study, participants were asked to review the photographs and label all eating moments using a web tool specifically designed for this purpose.

### Ground Truth

Participants were continuously audio and video recorded during the study as they performed their assigned activities (Figure 1). The only exceptions were the “Walking” and “Brushing Teeth” activities, when subjects left the user study room momentarily. The acquired video footage served as the foundation for the ground truth we estimated; all coding was performed using the ChronoViz tool [14].

For eating activities, we coded every food intake gesture and differentiated between gestures made with the instrumented arm versus the non-instrumented arm. For food intake, we marked the absolute time the food reached the mouth, and then added a fixed pre and post offset of 3 seconds to each intake event. This offset made it possible to model the entirety of food intake gestures, which often begin and end moments before and after the food is placed in the mouth. A three second offset was chosen empirically based on our observations of participants’ eating gestures. Non-eating activities were coded from the moment they began until their conclusion. In

other words, coding for non-eating activities was not focused on modeling any specific gesture.

The reliability of our ground truth estimation scheme was verified by having an external coder review 15% of the recorded audio and video. This was equivalent to 3 study sessions. To account for minor temporal differences in the assigned codes, we established that as long as they were within 3 seconds of each other, the codes referred to the same activity. By following this protocol, there was agreement in 96.7% of the coded gestures.

### In-the-Wild Studies

To evaluate the ecological validity of our method, we conducted two in-the-wild studies. For the first one, we recruited 7 participants (2 males and 5 females, between the ages of 21 and 29), who did not participate in the laboratory study. They were asked to wear the smartwatch on their dominant arm for an average of 5 hours and 42 minutes for one day while performing their normal everyday activities, which included taking public transportation, reading, walking, doing computer work, and eating. Four participants started the study in the morning and 3 in the afternoon and at least one eating moment was documented for each participant. Of a total data collection time of 31 hours and 28 minutes, 2 hours and 8 minutes corresponded to eating activities (6.7% of the total).

In the second study, one of the authors (male, 38 years of age) collected and annotated free-living inertial sensor data for 31 days. The author wore the smartwatch throughout the entire day, accumulating a total of 422 recorded hours during this period. For this dataset, 3.7% of all sensor data collected reflected eating activities; non-eating activities spanned personal hygiene (e.g., brushing teeth), transportation (e.g., driving), leisure (e.g., watching tv), and work (e.g. computer typing).

### Ground Truth

In the field of activity recognition, one of the critical challenges of in-the-wild studies is collecting reliable ground truth data for model training and evaluation. Self-reports are typically used for this purpose, but they are known to be susceptible to biases and memory recollection errors. To improve the reliability and objectivity of ground truth in our in-the-wild studies, we built an annotation platform around first-person images. In addition to the smartwatch, participants wore a wearable camera on a lanyard that captured photographs automatically every 60 seconds, depicting participant’s activities throughout the day. These images were uploaded in real-time to a server, and participants could access and review them at any time by logging into a password-protected web application. With this system, participants were able to indicate when they were engaged in eating moments from photographic evidence without having to share their photos with our research team, mitigating privacy concerns.

This method offered greater confidence for the ground truth labels, because the annotation was based on picture evidence. The camera was outfitted with a wide-angle lens to maximize

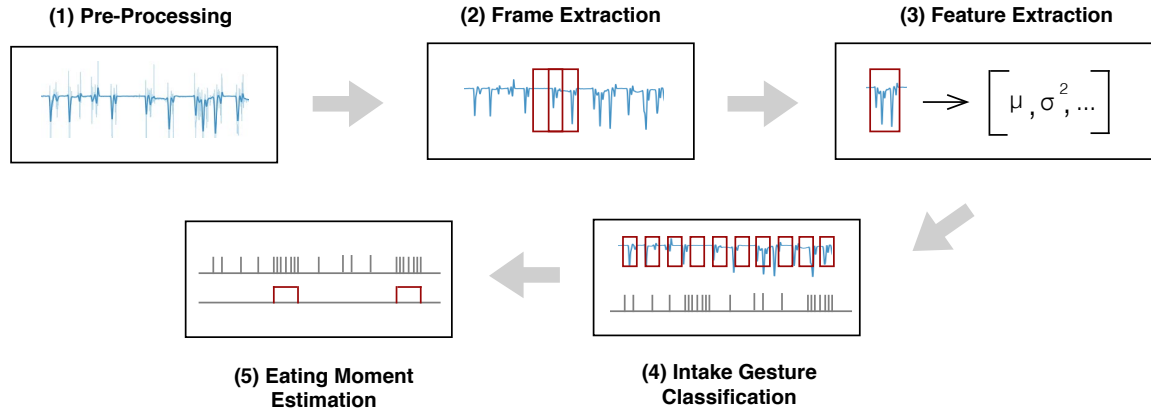


Figure 3. The data processing pipeline of our eating moment detection system. In our approach, food intake gestures are firstly identified from sensor data, and eating moments are subsequently estimated by clustering intake gestures over time.

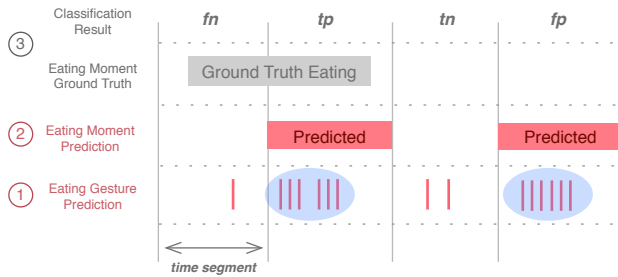


Figure 4. Going from bottom to top, the first step to eating moment recognition involves recognizing eating gestures (1). These are clustered temporally to identify eating moments (2). Finally, estimated eating moments are compared against ground truth in terms of precision and recall measurements at the level of time segments ranging from 3 to 60 minutes (3).

the field-of-view and capture food and eating-related activities and objects even if they were not directly in front of the individual. However, since photos were taken only every 60 seconds, there is a small possibility that a short eating moment (e.g., a snack) occurred in-between two photos and was not recorded. We set the interval to 60 seconds as a compromise between maximizing battery life and photo capturing for as long as possible on a given day.

### Public Datasets

To encourage research in the domain of intake gesture spotting and eating moment recognition, we are making the Lab-20, Wild-7, and Wild-Long datasets publicly available to the research community<sup>3</sup>.

### IMPLEMENTATION

Our system was designed to learn to identify moments when individuals are eating food. The sensor data processing pipeline consists of data capture and pre-processing, frame and feature extraction, food intake gesture classification, and eating moment estimation (Figure 3).

<sup>3</sup><http://www.ethomaz.com>

### Sensor Data Capture

Practicality was one of the key driving forces guiding this work. Thus, for data capture we relied on a non-specialized, off-the-self device with inertial sensing capabilities: the Pebble Watch<sup>4</sup>. We wrote custom logging software for capturing continuous 3-axis accelerometer sensor data from the device. The version of the smartwatch we employed did not contain a gyroscope. We also developed an iOS smartphone companion application for data storage and retrieval. Subjects wore the smartwatch on the wrist of their dominant hand. Sensor data was captured at 25Hz.

### Frame & Feature Extraction

The first steps in the data processing pipeline involved filtering the sensor streams using an exponentially-weighted moving average (EMA) filter and scaling the resulting data to unit norm (l2 normalization).

We extracted frames from the pre-processed data streams using a traditional sliding window approach with 50% overlap. The frame size plays an important role in classification since it needs to contain an entire food intake gesture. The gesture duration is determined by many factors, such as individuals' eating styles and whether they are multitasking (e.g., reading a book, socializing with friends) while eating. Based on data observed in our laboratory user study, we noticed that an intake gesture might last between 2 and 10 seconds. An analysis examining the sensitivity of window size suggested best classification results when the frame size was close to the mid-point of this range, around 6 seconds.

We computed five statistical functions for each frame, shown in Table 4: the signal's *mean*, *variance*, *skewness*, *kurtosis*, and *root mean square (RMS)*. These frame-level features comprise a concise and commonly used representation for the underlying inertial sensor data. The end result of the feature extraction step were 5-dimensional feature vectors for each axis of the accelerometer.

<sup>4</sup><http://www.getpebble.com>



Feature	Description	Definition
1 mean	average value of the samples of signal $\mathbf{x}$	$\mu_{\mathbf{x}} = \frac{1}{N} \sum_{n=0}^{N-1} x_n$
2 variance	power of values of signal $\mathbf{x}$ with its mean removed	$\sigma_x^2 = \frac{1}{N} \sum_{n=0}^{N-1}  x_n - \mu_x ^2$
3 skewness	measure of (lack of) symmetry in data distribution	$\frac{\sum_{n=1}^N (x_n - \mu)^3}{(N-1)s^3}$
4 kurtosis	measure of the shape of the data distribution	$\frac{\sum_{n=1}^N (x_n - \mu)^4}{(N-1)s^4}$
5 RMS	square root of the average power of signal $\mathbf{x}$	$\sqrt{P_x}$ , where $P_x = \frac{E_x}{N} = \frac{1}{N} \sum_{n=0}^{N-1}  x_n ^2$

Table 4. Feature definitions used for food intake gesture classification

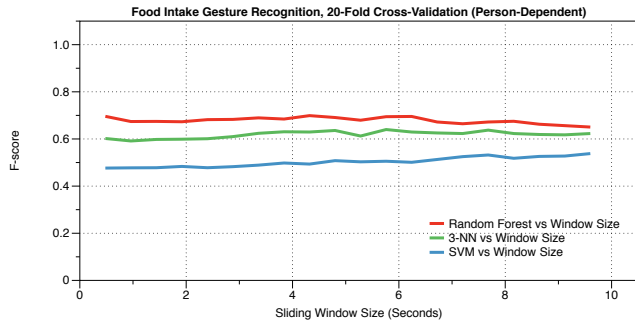


Figure 5. We evaluated the person-dependent performance of three food intake gesture classifiers with respect to window size. Each classifier was trained with a different learning algorithm: Random Forest, SVM (RBF kernel), and 3-NN. We achieved best results with the Random Forest classifier.

### Food Intake Gesture Classification

The first classification task in our system is the identification of food intake gestures, which we define as the arm and hand gestures involved in bringing food to the mouth from a resting position on a table, for instance, and then lowering the arm and hand back to the original resting position. In practice, this task is made much harder by intra-class diversity. For example, individuals eat differently if compared to each other and different types of food consumption require different gestures. Additionally, an individual might perform other tasks while eating, such as gesticulate when talking to others, hold a mobile phone or magazine, etc.

For food intake gesture classification, we evaluated classifiers using the Scikit-learn Python package [26]. Best results were obtained with the Random Forest learning algorithm.

### Eating Moment Estimation

We estimated eating moments by examining the temporal density of observed food intake gestures. When a minimum number of inferred intake gestures were within a certain temporal distance of each other, we called this event an eating moment. We employed the DBSCAN clustering algorithm for this calculation [13].

DBSCAN has three characteristics that make it especially compelling for our scenario; there is no need to specify the number of clusters ahead of time, it is good for data that con-

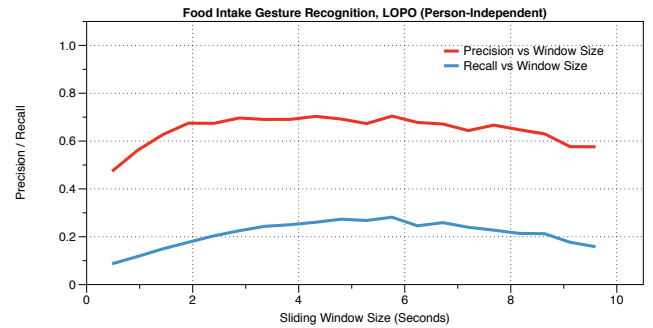


Figure 6. We performed a leave-one-participant-out (LOPO) evaluation of the food intake gesture classifier trained with the Random Forest learning method. The figure shows its sensitivity to window size.

tains clusters of similar density, and it is capable of identifying outliers (i.e., food intake gestures) in low-density regions. A well-defined method for pinpointing outliers is important because there are many gestures that could be confused with intake ones throughout one's day. Once areas of high intake-gesture densities have been identified as clusters in the time domain, we calculate their centroids and report them as eating moment occurrences.

## RESULTS

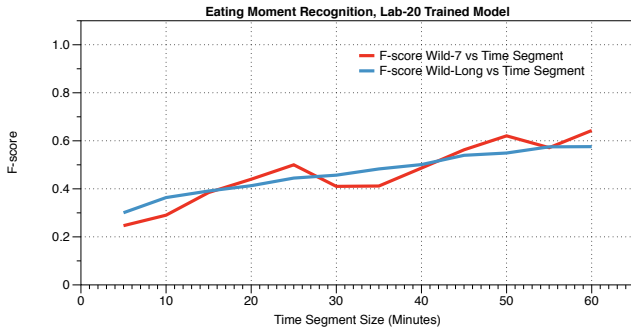
To reiterate, our goal is to develop and evaluate a *practical* approach to detect eating moments, using sensor data from an off-the-shelf smartwatch. To that end, the primary performance metric we wished to assess was whether the system could distinguish eating moments from non-eating moments. In this section we first review our eating gesture classification findings and then discuss our eating moment recognition results.

### Eating Gesture Recognition

In our system, predicting eating moments hinges on the detection of food intake gestures. Using the Lab-20, we evaluated the performance of three food intake gesture classifiers (Random Forest, SVM, and 3-NN) as a function of sliding window size for the person-dependent (Figure 5) and person-independent cases. The Random Forest classifier outperformed the SVM and 3-NN classifiers using the F-score measure for comparison. We attribute this result to the Random

	Other	Eat FK	Eat FS	Eat Hand	Movie	Walk	Chat	Phone	Comb	Brush	Wait
Other	26%	6.6%	4%	13.2%	13.7%	1.5%	28.5%	3%	0%	3%	0%
Eat FK	2.4%	35.6%	34.2%	14.3%	1.6%	0.2%	10.2%	0.2%	0.2%	0.7%	0%
Eat FS	0.2%	6.2%	74.7%	7.1%	1.1%	0.6%	7.5%	0.5%	0%	1.7%	0%
Eat Hand	1%	4.2%	9.6%	72.9%	1.7%	0.9%	8.8%	0.2%	0%	0.1%	0.1%
Movie	2.2%	0.8%	2.9%	4.7%	77.3%	0.82%	10.1%	0.6%	0%	0%	0.2%
Walk	0.3%	0.3%	0.3%	0.7%	0%	91.3%	5.5%	0%	0%	1.3%	0%
Chat	2.6%	4.5%	15.9%	10.7%	6.9%	1.5%	53%	0.8%	0.3%	3.1%	0.3%
Phone	2.4%	2.4%	24.7%	14%	1.6%	0%	5.7%	47.1%	0%	1.6%	0%
Comb	7.1%	14.2%	17.8%	3.5%	0%	0%	7.1%	0%	39.2%	10.7%	0%
Brush	1.4%	3.3%	16.8%	16.8%	0%	11%	11%	0.9%	0.9%	37.5%	0%
Wait	3%	5.1%	17.3%	5.1%	5.1%	4%	9.1%	0%	0%	6.1%	44.9%

**Table 5.** Confusion matrix showing the percentage of actual vs. predicted activities by the Random Forest model. The FK and FS acronyms refer to eating activities employing fork and knife, and fork or spoon, respectively.



**Figure 7.** F-score results for a model trained with lab data (Lab-20 dataset) and tested with in-the-wild data, Wild-7 (red), and Wild-Long (blue). The x-axis correspond to time segment size, in minutes.

Forest’s powerful nonlinear modeling capability. This learning algorithm was also appealing to us because it does not require much parameter tuning.

A person-independent evaluation of the Random Forest classifier using the leave-one-participant-out strategy (LOPO) is shown in Figure 6. Note that the reported precision, recall and F-score measurements in Figures 5 and 6 reflect the classifiers’ ability to spot intake gestures at the frame level, and best performance was achieved with a frame size of just under 6 seconds.

Table 5 provides a detailed picture of how the Random Forest model performed at classifying eating gestures in relation to non-eating activities. The data for all laboratory study participants was combined and randomly split into one training and one test set; approximately one third of the data was held out for testing. This procedure was performed with Scikit-learn’s train-test-split cross-validation function [26]. For purposes of reporting results, we further distinguish 3 different eating gestures to gain a richer understanding of model classification and error rates: eating with fork and knife (i.e., Eat FK), eating with fork or spoon only (i.e., Eat FS), and eating with hands (i.e., Eat Hand).

### Eating Moment Recognition

As previously described, our approach for inferring eating moments depends on the temporal density of observed food intake gestures; we cluster these intake gestures over time

using the DBSCAN algorithm, which takes two parameters, a minimum number of intake gestures (minPts), and a distance measure given as a temporal neighborhood (eps). To assess how well eating moments were recognized, we compared ground truth and predictions over a time window that is longer than a frame size. This is necessary because an eating moment is in the range of minutes, not seconds. In this paper, we refer to this longer time window for eating moment recognition as a *time segment*, shown in Figure 4. When one or more eating moments are recognized within a time segment, the entire time segment is assigned the eating label.

One of the questions our work explores is whether it is feasible to build a model for eating moment recognition based on semi-naturalistic behavior data captured in a laboratory. To answer this question, we trained a model with the Lab-20 dataset and tested it on both in-the-wild datasets (Wild-7 and Wild-Long). Figure 7 plots F-scores as a function of time segment size ranging from 5 to 60 minutes (DBSCAN parameters set to minPts=1, eps=10, meaning at least 1 intake gesture that is within 10 seconds from another recognized intake gesture). The charts show an upward trend in recognition performance as time segment duration increases. This is because more data points become available in terms of recognized and non-recognized food intake gestures, leading to improved density estimation, and thus better eating moment recognition results. When the time segment size is set to 60 minutes, the F-scores are 64.8% and 56.8%.

Our intuiting guiding eating moment recognition is that making a prediction about a 60-minute time segment would suffice for most practical applications of our work. Given that intuition, it is valuable to understand how much we can optimize our classifier when the time segment is fixed at 60 minutes. Varying the minPts and eps parameters of the DBSCAN algorithm, but still using the Lab-20-trained intake gesture recognition model, (shown in Figures 8 and 9), F-scores of 76.1% (66.7% Precision, 88.8% Recall) and 71.3% (65.2% Precision, 78.6% Recall) could be achieved when evaluating the classifier with the Wild-7 and Wild-Long datasets, respectively.



## DISCUSSION

In this section, we discuss our classification results, the instrumentation strategy we chose, characteristics of the data collected, and the practical implications of our findings.

### Classification Challenges

To more realistically assess our system's classification performance, we purposely included gestures that required arm movements similar to food intake gestures. Activities such as placing a phone call, combing hair and brushing teeth are all similar to eating in that they all require hand-arm motions around the head and mouth areas. Other observed movements that occurred in our laboratory study closely matching eating gestures included wiping the face with a napkin, scratching the head, and assuming a resting position by supporting the head and chin with the instrumented hand and wrist. Because of the semi-controlled nature of our laboratory study, these movements occurred naturally during sessions, and did not have to be scripted.

Based on our results, shown in the the confusion matrix in Table 5, we found that one of the most challenging activities to discriminate from eating was 'Chat'. This is because when people are having a conversation, they typically gesticulate. This effect varies in intensity amongst individuals but it was significant enough across all participants in the laboratory study that between 7.5% and 10% of each eating intake class (Eat FK, Eat FS, Eat Hand) was misclassified as 'Chat'.

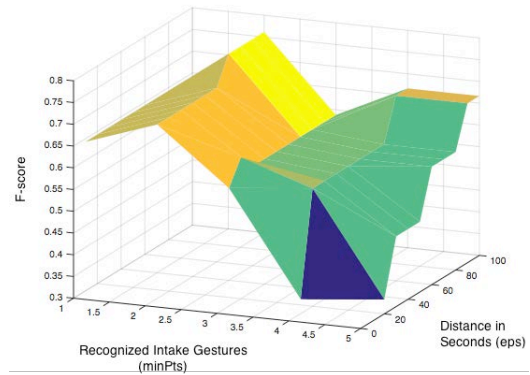
In Table 5, it is also possible to see false positives originating from the 'Phone', 'Comb', and 'Brush' activities. In the context of the lab study this is not surprising since these activities were specifically included to induce misclassifications. Common to these non-eating activities gestures was a movement bringing the hand close to the head; the *temporality* of subsequent movements was one of the key characteristic differentiating them. In the 'Phone' activity, the hand stayed up holding the phone close to the ear; in effect there is no subsequent gesture in this case. For the 'Comb' activity, the hand was lifted up and remained in motion, moving slowly in a pattern that depended on the hairstyle of the participant. The 'Brush' activity pattern was distinguished by quick-moving hand gestures while holding a toothbrush. We believe we can lower the rate of false positives by incorporating time-dependent features that can better characterize these types of non-eating activities.

### Intra-Class Diversity

We observed a large amount of variability in participants' eating styles. Some held a sandwich with two hands, others with one hand, sometimes alternating between them. A minority of participants took bites of their food at regular intervals (P4 in Figure 10). Others were not so regular; they gesticulated more while talking and eating (P5 in Figure 10).

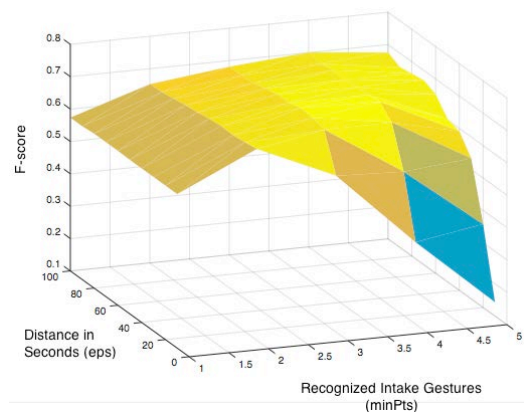
When using utensils, and in the short intervals between bites, some participants kept mixing their food in a regular pattern. This could be attributed to an individual's own eating style or an attempt to cool off the food, for example. There was significant variation in the way participants ate smaller foods as well. Several participants held several kernels of popcorn in

Eating Moment Recognition, Clustering Parameters (Wild-7)



**Figure 8.** F-score results for estimating eating moments given a time segment of 60 minutes as a function of DBSCAN parameters (minPts, and eps). Tested on the Wild-7 dataset, eating moments can be estimated with an F-score of up to 76.1% when minPts=2 and eps=80 (at least 2 intake gestures that are within 80 seconds from another intake gesture).

Eating Moment Recognition, Clustering Parameters (Wild-Long)

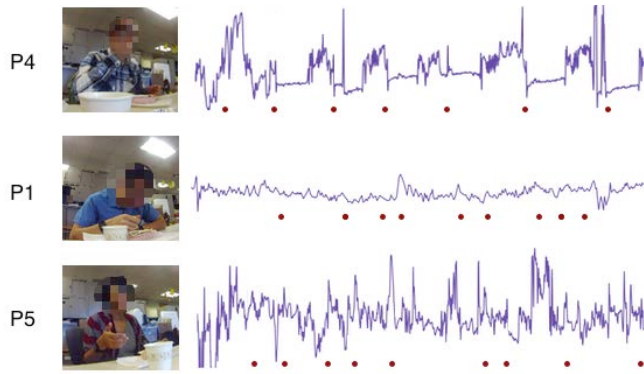


**Figure 9.** F-score results for estimating eating moments given a time segment of 60 minutes as a function of DBSCAN parameters (minPts, and eps). Tested on the Wild-Long dataset, eating moments can be estimated with an F-score of up to 71.3% when minPts=3 and eps=40 (at least 3 intake gestures that are within 40 seconds from another intake gesture).

hand and ate them continuously until they were gone. Others liked to eat more than one popcorn at a time.

While many participants performed the “traditional” food intake gesture of bringing food to the mouth using utensils, hands, or by lifting a bowl, we noticed that many participants did the opposite; they bent over their plate, brought their head close to the food and then moved their arm in a modified, shorter and subtler version of the traditional intake gesture. This was particularly common when participants were trying to avoid food spillage (P1 in Figure 10).

In this study we did not create a separate model for each observed eating style; all intake gestures were given one label: “eating”. Without any question, this posed an additional challenge to the classification task. Fitting a model to user-specific data might be the most effective way to address intra-class diversity, and we hope to explore this in future work. Also, face-mounted wearable computing systems like Google



**Figure 10.** The accelerometer data (x-axis) of three participants as they ate a serving of lasagna depicts personal variation in eating styles and makes intra-class diversity evident. The red dots are intake gesture markers.

Glass are becoming more popular; these devices offer the opportunity to capture inertial sensing data reflecting head movements, which might contribute significantly to the identification of eating and chewing activities despite individual differences.

### Instrumentation

We provided participants with one wrist-worn device, a smartwatch, and placed it on their dominant hand. There are two key reasons why we decided on a strategy of minimal instrumentation. Firstly, in real-world settings, people wear only one smartwatch at a time. In this context, with an eye towards the practical applicability of this research, we were interested in the extent to which eating moments can be estimated with just one sensor data capture device. Secondly, we felt that asking participants to wear one additional device would be unnatural, and thus result in a level of discomfort that could compromise the validity of the data.

We chose participants' dominant hand because it is the one that is typically used in food intake gestures. However, the dominant hand might play different roles while eating, such as cutting with a knife, and this has an effect in modeling intake gestures; it is possible to observe in Table 4 that the "eating with a fork and knife" class was misclassified as "eating with fork or spoon only", and with "eating with hand". This is inconsequential if the goal is to identify "whether" eating is taking place, but it presents modeling opportunities for characterizing "what" is being eaten.

### Ecological Validity

Our evaluation results demonstrate the promise of a minimally-instrumented approach to eating moment detection. However, it is important to situate our findings in light of our study design and aspects of our system implementation. An issue that might arise in practice while collecting data with only one device is that certain eating gestures might not get captured. For instance, a person might be wearing a smartwatch on the non-dominant hand while eating with a fork held by the dominant hand. Although this scenario represents a challenge, we believe it can be addressed in two ways: by modeling non-eating gestures performed by the

non-dominant hand during eating, and by leveraging additional modalities such as ambient sounds. In future work, we plan to explore the combination of these two different paths.

With regards to the validity of our results, the types of foods that we served participants and the enforcement of which utensils they were allowed to use, if any, were in line with current western eating traditions. We aimed for a representative sample of eating activities and styles by picking foods such as rice, popcorn, and sandwiches apples but our scientific claims do not and cannot generalize to all populations and cultures. For instance, none of participants in the study ate with chopsticks.

### Practical Applications

Despite the importance of high precision and recall measures for both benchmarking and practical applications, our experiments showed that since there are usually many intake gestures within one eating moment, a slightly lower recall in food intake gesture classification does not have a large affect in the results. In contrast, consecutive false positives have a direct effect in the misclassification of eating moments. With respect to the applications we envision leveraging this work, there are two paths to consider. In a system designed to facilitate food journaling, lower precision means that individuals might be frequently prompted to provide details about meals that did not occur, which is undesirable. However, as a tool for health researchers to determine when individuals eat meals, what is critically important is to not miss any eating activities. In this case, false positives are preferable to false negatives.

### Battery Performance

Our data capture setup employed a Pebble smartwatch and an iPhone 4S. Smartwatch accelerometer data was captured at 25Hz and transmitted to the smartphone every second using Bluetooth. For the laboratory study, the sensor data was saved locally on the phone and retrieved at the end of each session. Sessions in the lab lasted 31 minutes and 21 seconds on average, and battery performance was never a concern.

On the other hand, the in-the-wild studies posed a significant challenge in terms of power consumption. In this context, the smartphone played three roles. Worn on a lanyard, it was programmed to take snapshots automatically every 60 seconds. This was necessary to obtain a measure of ground truth of participants' activities over the course of their day. Secondly, the smartphone continued to serve as an end-point buffer for all the incoming smartwatch sensor data over Bluetooth. Finally, the phone uploaded the sensor data to a server using a cellular connection every minute, and thus in near real-time.

Starting on a full charge, the smartphone was able to perform all these tasks for an average of 5 hours and 42 minutes, which determined the duration of our one-day in-the-wild studies. For the 31-day in-the-wild study, the same instrumentation was used but with the addition of one 15,000mAh battery pack connected to the phone. Carrying the battery pack proved to be an additional inconvenience, but it allowed data collection to take place for the entire day.

Throughout the studies, the smartwatch, the smartphone and the battery pack were restored to full charge overnight and used again the following day. The Pebble watch never represented a limiting factor in data collection. We attribute its low power consumption to its e-ink display and lack of a more sophisticated inertial measurement unit (IMU).

### FUTURE WORK

Technically, there are numerous opportunities to extend this work. In the near term, our goal is to continue to improve our eating gesture detection by experimenting with methods such as Dynamic Time Warping (DTW) and new feature representations.

One area we believe is particularly promising in the context of eating moment recognition is personalization. Eating styles vary from person to person to a large degree (Figure 10), and we intend to investigate the effect of a truly personalized model on performance results.

Finally, we are interested in fusing on-body inertial sensing with additional sensing modalities for eating moment recognition, such as location, and continuing to explore approaches for identifying not only when individuals are eating but also what they are consuming.

### CONCLUSIONS

We describe the implementation and evaluation of an approach that identifies eating moments using 3-axis accelerometer sensor data from an off-the-shelf smartwatch. An eating moment classifier trained with participants in a semi-controlled lab setting was able to recognize eating moments in two in-the-wild studies with F-scores of 76.1% (66.7% Precision, 88.8% Recall), and 71.3% (65.2% Precision, 78.6% Recall).

These results are promising for three main reasons. Firstly, they represent a baseline for practical eating detection; we anticipate performance gains when employing additional inertial sensing modalities. As a means of comparison, Amft et al. obtained 84% recall and 94% precision with accelerometer and gyroscope in drinking gesture spotting [1]. Secondly, our studies explored one type of sensing modality, and many other contextual cues could be utilized to improve eating moment detection, such as location and perhaps even ambient sounds [32]. Thirdly, and more broadly, this work suggests that it might be possible to build ecologically valid models of complex human behaviors while minimizing the costly acquisition of annotated data in real-world conditions; the dataset we compiled and used in our analysis is being made public so that others can validate our results and build upon our work.

Building a truly generalizable system for eating moment detection, and automatic food intake monitoring in general, represents a significant challenge. We believe such a system could provide the foundation for a new class of practical applications, benefiting individuals and health researchers. Despite limitations and opportunities for improvement, we believe this work provides compelling evidence that a practical solution around commodity sensing can play an important role towards this vision.

### ACKNOWLEDGMENTS

This work was supported by the Intel Science and Technology Center for Pervasive Computing (ISTC-PC), and by the National Institutes of Health under award 1U54EB020404-01.

### REFERENCES

1. Amft, O., Bannach, D., Pirkl, G., Kreil, M., and Lukowicz, P. Towards wearable sensing-based assessment of fluid intake. In *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2010 8th IEEE International Conference on (2010), 298–303.
2. Amft, O., and Tröster, G. On-Body Sensing Solutions for Automatic Dietary Monitoring. *IEEE pervasive computing* 8, 2 (Apr. 2009).
3. Arab, L., Estrin, D., Kim, D. H., Burke, J., and Goldman, J. Feasibility testing of an automated image-capture method to aid dietary recall. *European Journal of Clinical Nutrition* 65, 10 (May 2011), 1156–1162.
4. Archer, E., Hand, G. A., and Blair, S. N. Validity of U.S. Nutritional Surveillance: National Health and Nutrition Examination Survey Caloric Energy Intake Data, 1971–2010. *PLoS ONE* 8, 10 (Oct. 2013), 76632.
5. Burke, L. E., Swigart, V., Warziski Turk, M., Derro, N., and Ewing, L. J. Experiences of self-monitoring: successes and struggles during treatment for weight loss. *Qualitative health research* 19, 6 (June 2009), 815–828.
6. Burke, L. E., Wang, J., and Sevick, M. A. Self-Monitoring in Weight Loss: A Systematic Review of the Literature. *YJADA* 111, 1 (Jan. 2011), 92–102.
7. Cheng, J., Zhou, B., Kunze, K., Rheinländer, C. C., Wille, S., Wehn, N., Weppner, J., and Lukowicz, P. Activity recognition and nutrition monitoring in every day situations with a textile capacitive neckband. In *the 2013 ACM conference*, ACM Press (New York, New York, USA, 2013), 155.
8. Cordeiro, F., Epstein, D., Thomaz, E., Bales, E., Jagannathan, A. K., Abowd, G. D., and Fogarty, J. Barriers and negative nudges: Exploring challenges in food journaling. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2015).
9. Dhurandhar, N. V., Schoeller, D., Brown, A. W., Heymsfield, S. B., Thomas, D., Sorensen, T. I. A., Speakman, J. R., Jeansonne, M., and Allison, D. B. Energy balance measurement: when something is not better than nothing. *International Journal of Obesity* (Nov. 2014).
10. Doherty, A. R. A., Hodges, S. E. S., King, A. C. A., Smeaton, A. F. A., Berry, E. E., Moulin, C. J. A. C., Lindley, S. S., Kelly, P. P., and Foster, C. C. Wearable cameras in health: the state of the art and future possibilities. *American journal of preventive medicine* 44, 3 (Mar. 2013), 320–323.

11. Dong, Y. Tracking Wrist Motion to Detect and Measure the Eating Intake of Free-Living Humans. *Thesis (Ph. D.) Clemson University* (May 2012), 1–106.
12. Dong, Y., Scisco, J., Wilson, M., Muth, E., and Hoover, A. Detecting periods of eating during free living by tracking wrist motion. *IEEE Journal of Biomedical Health Informatics* (Sept. 2013).
13. Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD* (1996), 226–231.
14. Fouse, A., Weibel, N., Hutchins, E., and Hollan, J. D. ChronoViz: a system for supporting navigation of time-coded data. *CHI Extended Abstracts* (2011), 299–304.
15. Jacobs, D. R. Challenges in research in nutritional epidemiology. *Nutritional Health* (2012), 29–42.
16. Junker, H., Amft, O., Lukowicz, P., and Tröster, G. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition* 41, 6 (June 2008), 2010–2024.
17. Kadamura, A., Li, C.-Y., Tsukada, K., Chu, H.-H., and Siio, I. Persuasive technology to improve eating behavior using a sensor-embedded fork. In *the 2014 ACM International Joint Conference*, ACM Press (New York, New York, USA, 2014), 319–329.
18. Kim, H.-J., Kim, M., Lee, S.-J., and Choi, Y. S. An analysis of eating activities for automatic food type recognition. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific* (2012), 1–5.
19. Kleitman, N. *Sleep and wakefulness*. The University of Chicago Press, Chicago, July 1963.
20. Liu, J., Johns, E., Atallah, L., Pettitt, C., Lo, B., Frost, G., and Yang, G.-Z. An Intelligent Food-Intake Monitoring System Using Wearable Sensors. In *Wearable and Implantable Body Sensor Networks (BSN), 2012 Ninth International Conference on*, IEEE Computer Society (2012), 154–160.
21. Martin, C. K., Han, H., Coulon, S. M., Allen, H. R., Champagne, C. M., and Anton, S. D. 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, 03 (July 2008), 446.
22. Michels, K. B. A renaissance for measurement error. *International journal of epidemiology* 30, 3 (June 2001), 421–422.
23. Michels, K. B. Nutritional epidemiology—past, present, future. *International journal of epidemiology* 32, 4 (Aug. 2003), 486–488.
24. Noronha, J., Hysen, E., Zhang, H., and Gajos, K. Z. Platemate: crowdsourcing nutritional analysis from food photographs. *Proceedings of the 24th annual ACM symposium on User interface software and technology* (2011), 1–12.
25. O’Loughlin, G., Cullen, S. J., McGoldrick, A., O’Connor, S., Blain, R., O’Malley, S., and Warrington, G. D. Using a wearable camera to increase the accuracy of dietary analysis. *American journal of preventive medicine* 44, 3 (Mar. 2013), 297–301.
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
27. Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., and Hansen, M. Image browsing, processing, and clustering for participatory sensing: lessons from a DietSense prototype. In *EmNets ’07: Proceedings of the 4th workshop on Embedded networked sensors*, ACM Request Permissions (June 2007).
28. Sazonov, E., Schuckers, S., Lopez-Meyer, P., Makeyev, O., Sazonova, N., Melanson, E. L., and Neuman, M. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. *Physiological Measurement* 29, 5 (Apr. 2008), 525–541.
29. Stellar, E., and Shrager, E. E. Chews and swallows and the microstructure of eating. *The American journal of clinical nutrition* 42, 5 (1985), 973–982.
30. Sun, M., Fernstrom, J. D., Jia, W., Hackworth, S. A., Yao, N., Li, Y., Li, C., Fernstrom, M. H., and Sciabassi, R. J. A wearable electronic system for objective dietary assessment. *Journal of the American Dietetic Association* 110, 1 (2010), 45.
31. Thomaz, E., Parnami, A., Essa, I. A., and Abowd, G. D. Feasibility of identifying eating moments from first-person images leveraging human computation. *SenseCam* (2013), 26–33.
32. Thomaz, E., Zhang, C., Essa, I., and Abowd, G. D. Inferring meal eating activities in real world settings from ambient sounds: A feasibility study. *Proceedings of the ACM Conference on Intelligent User Interfaces* (2015).
33. Willett, W. *Nutritional Epidemiology*. Oxford University Press, Oct. 2012.
34. Yatani, K., and Truong, K. N. BodyScope: a wearable acoustic sensor for activity recognition. *UbiComp ’12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (2012), 341–350.
35. Zhang, S., Ang, M. H., Xiao, W., and Tham, C. K. Detection of activities by wireless sensors for daily life surveillance: eating and drinking. *Sensors* 9, 3 (2009), 1499–1517.