Exploring Symmetric and Asymmetric Bimanual Eating Detection with Inertial Sensors on the Wrist

Edison Thomaz
The University of Texas at Austin
ethomaz@utexas.edu

Abdelkareem Bedri
Carnegie Mellon University
akareem.bedri@gmail.com

Temiloluwa Prioleau
Rice University
temiprioleau@gmail.com

Irfan Essa
Georgia Institute of Technology
irfan@gatech.edu

Gregory D. Abowd
Georgia Institute of Technology
abowd@gatech.edu

ABSTRACT

Motivated by health applications, eating detection with off-the-shelf devices has been an active area of research. A common approach has been to recognize and model individual intake gestures with wrist-mounted inertial sensors. Despite promising results, this approach is limiting as it requires the sensing device to be worn on the hand performing the intake gesture, which cannot be guaranteed in practice. Through a study with 14 participants comparing eating detection performance when gestural data is recorded with a wrist-mounted device on (1) both hands, (2) only the dominant hand, and (3) only the non-dominant hand, we provide evidence that a larger set of arm and hand movement patterns beyond food intake gestures are predictive of eating activities when L1 or L2 normalization is applied to the data. Our results are supported by the theory of asymmetric bimanual action and contribute to the field of automated dietary monitoring. In particular, it shines light on a new direction for eating activity recognition with consumer wearables in realistic settings.

Keywords

Dietary Monitoring; Food Tracking; Food Logging; Eating Detection; Activity Recognition; Inertial Sensing

1. INTRODUCTION

Automatically detecting eating activities is a cornerstone of a wide range of health applications, helping behavioral researchers understand the link between diet and disease [Hatori et al. 2012], and enabling new forms of dietary self-monitoring such as semi-automated food journaling [Choe et al. 2017]. Over the last few years, computing researchers have developed new approaches for automatic eating detection by making use of the inertial sensing capabilities in off-the-shelf devices such as mobile phones, smart watches, activity trackers, and wearable devices [Dong et al. 2013, Thomaz et al. 2015, Junker et al. 2008, Amft and Tröster 2009, Merck et al. 2016, Rahman et al. 2015, Rahman et al. 2016]. This methodology, referred to as commodity sensing, opportunistically leverages technologies that the general population has begun to incorporate into their everyday lives, hence greatly facilitating long-term data collection in naturalistic settings.

Despite the possibilities afforded by commodity sensing, a key question of relying exclusively on devices that individuals have adopted is whether these devices provide the sensing coverage that is required to fully recognize certain behaviors. In the context of eating detection, researchers have used sensors in popular smart watches to identify unimanual food intake gestures. While this seems like a straightforward task in principle, numerous challenges exist in practice. The most significant one is that in everyday living, many people choose to wear a wristwatch on their non-dominant hand while food intake gestures are usually performed by the dominant hand. Another difficulty emerges when utensils are involved. There are many styles individuals embrace when it comes to consuming foods with fork and knife. Most people in the U.S. follow either the European eating style, the American eating style or a hybrid of the two. According to the European style, the dominant hand is used exclusively for holding the knife and cutting while in the American style, the dominant hand is used for both cutting the food and bringing it to the mouth.

Assuming no prior knowledge of how individuals eat and which hand they regard as their dominant one, it is easy to see how eating detection based on the recognition and modeling of one unimanual gesture (i.e., food intake) is rather limiting. The contribution of this work is to provide evidence of a new type of behavior marker in eating detection with commodity sensors. Technically, we show that by applying normalization to the sensor data over longer time windows, it is possible to recognize symmetric and asymmetric bimanual actions beyond unimanual intake gestures [Guiard 1987].

This was accomplished with a user study where 14 participants performed eating and non-eating activities. In practice, this is significant because it provides grounds for eating behavior tracking with one wrist-mounted device placed on the non-dominant hand.

2. EATING AS A BIMANUAL TASK

Manual tasks have been of scientific interest across disciplines for many decades. For example, primatologists have studied prehensile movements and the relationship between hand posture and activity [Napier and Tuttle 1993], while HCI researchers have investigated two-handed mode switching techniques and the benefits of two-handed manipulation [Li et al. 2005, Hinckley et al. 2010].

In the context of bimanual action, the work of Guiard has been particularly relevant [Guiard 1987]. While proposing a framework for investigating asymmetry in bimanual action, he described three classes of human manual activities: unimanual (one-handed tasks such as dart throwing), asymmetric bimanual (two hands taking different roles, such as playing a stringed musical instrument), and symmetric bimanual (two hands taking the same role, as in rope skipping). Based on this description, eating activities could be placed into any of these three classes depending on the context, such as what is being consumed and how, as shown in Table 1. It is also worth noting that within just one eating episode, such as breakfast or lunch, it is typical for individuals to perform unimanual and bimanual gestures that fall into all of the proposed human manual activity classes. In other words, the theoretical construct proposed by Guiard validates the intuition that it would be advantageous to capture data from both wrists in eating detection, even though this level of tracking may not be possible in practice.

An additional contribution of Guiard's work is particularly relevant to the discussion of identifying food intake gestures, especially with regards to the value of capturing and analyzing eating gestures with only one wrist-mounted sensor. His observations suggest that there is, in fact, a "division of labor" between hands spanning all human bimanual activities; for any task for which only one hand seems to be involved (i.e. unimanual), it is impossible to demonstrate that the other hand plays no role. For instance, in an activity such as writing, the non-dominant hand seems to play a complementary role to the dominant hand, repositioning the paper and assisting with spatial reference. The implication of this finding for eating detection is that it might be possible to identify when eating is taking place by analyzing subtle, postural gestures and patterns performed by the non-instrumented wrist. This is the research question that underlies this work.

3. RELATED WORK

Perhaps the most comprehensive analysis of the impact of sensor placement and modality on human activity recognition with inertial sensors is provided by Bulling et al. [Bulling et al. 2014]. They used three inertial measurement units (IMU) to classify activities in a tutorial on human activity recognition. This work neither examines recognition performance with sensors that are symmetrically placed on both sides of the body nor focuses on eating detection.

Several research efforts have applied wrist-mounted inertial sensors towards automated dietary monitoring. How-
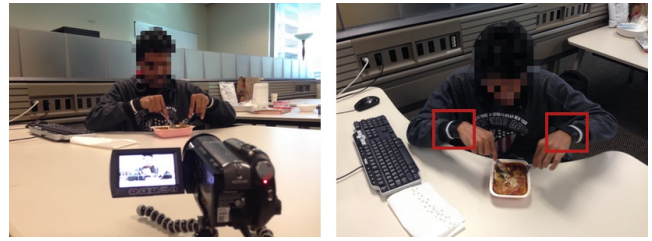

Figure 1: Participants were video-recorded as they performed eating and non-eating activities. The image on the right highlights the wrist-mounted sensor participants wore during the study.

ever, only a fraction of these efforts instrumented both participants' wrists at the same time during data collection. Amft et al. demonstrated how a jacket instrumented with a variety of sensors, including inertial sensors in the lower arms, could be used to detect eating gestures across various gesture categories [Junker et al. 2008, Amft and Tröster 2009]. However, the authors did not provide a disaggregated view of the impact of each one of the sensors on classification. Dong et al. showed an approach for detecting eating periods using an iPhone 4 on the wrist for data collection [Dong et al. 2013]. Only one wrist was instrumented and the phone was placed inside a case that could be tied around the forearm. Using a smart watch with inertial sensing capabilities, Thomaz et al. demonstrated eating detection by conducting studies both in a laboratory setting and in real-world conditions [Thomaz et al. 2015]. Similar to Dong et al., the wrist-mounted device was placed on the subjects' dominant hand only.

Other researchers have worked in the area of eating detection and recognition with inertial sensors. Kim et al. used the Chronos wrist-mounted sensing platform to identify eating activities and recognize food types [Kim et al. 2012]. Like Amft et al. and Thomaz et al., these efforts focused on classification performance given a chosen sensing modality and did not explore how sensor placement and hand movement patterns might impact classification results. More recently, Rahman et al. leveraged two wrist-mounted sensors to predict about-to-eat moments, but both sensors were placed on the dominant hand [Rahman et al. 2016]. Merck et al. recognizes the challenge of needing to track the dominant hand while sensing devices (e.g., smartwatches) are usually worn on the non-dominant hand [Merck et al. 2016]. While it does a good job comparing the impact of dominant versus non-dominant wrist sensing, it does not discuss eating behavior tracking beyond unimanual food intake actions.

4. USER STUDY

To explore symmetric and asymmetric bimanual hand actions in eating activity prediction, we conducted an IRB-approved study. The aim of our study was to compare the performance of inertial sensor-based food intake gesture classification with data collected from (1) both hands, (2) only the dominant hand, and (3) only the non-dominant hand over time windows of varying duration. The study protocol, designed in close collaboration with nutritional epidemiologists who are experienced at running dietary assessment experiments, centered on capturing behavioral sensor data

Activity Class	Description	Eating Example
Unimanual	One-handed tasks	Having soup with spoon
Asymmetric Bimanual	Two hands taking different roles	Eating with fork and knife
Symmetric Bimanual	Two hands taking the same role	Holding a sandwich with both hands

Table 1: The diversity of eating modes places the eating activity into any of the three activity classes defined by Guiard [Guiard 1987].

as participants ate a variety of foods and performed non-eating activities.

A convenient sample of 14 participants (10 males, 4 females) were recruited as participants across two educational institutions; they were graduate and undergraduate students between the ages of 18 and 55; all of them claimed to be right-handed. The study lasted an average of 55 minutes and took place around lunchtime. We instrumented participants with two Microsoft Bands, one on each wrist, for collecting accelerometer and gyroscope inertial sensor data. Participants performed eating and non-eating activities interchangeably under no time constraints. For the eating activities, the foods offered were: *popcorn, lasagna, and yogurt.* These food choices were served so that we could collect data that represented different eating styles (i.e., with fork and knife, with spoon, and with hands only). All participants were offered the exact same food types and amount for each food. Some eating activities required the use of utensils and some did not. Participants were told which foods would be served and allowed to eat as much as they wanted; drinking activities were coded as separate from eating activities. Although none of the participants had food restrictions, all foods served were vegetarian. We constrained the set of available food options to mitigate confounds and maintain consistency across study sessions.

Seven non-eating activities were included in the study: *watch a movie trailer, read a magazine, take a walk, use smartphone, place a phone call, use a computer, and brush teeth.* These tasks were selected to represent activities that required physical movement (i.e., walking), represented common everyday tasks (e.g., use a computer, use mobile phone, read a book), or could be confused with intake gestures due to the hand coming in proximity to the head (e.g., brush teeth, place a phone call). While it is often desirable to conduct studies in naturalistic settings, we chose to conduct a semi-controlled study in our labs in order to obtain reliable and detailed ground truth annotations for intake gestures. While methods such as context-aware experience sampling and wearable cameras have facilitated the collection of ground truth labels in naturalistic settings, the annotation resolution these techniques offer is at the level of activities, not gestures. To make the lab study as ecologically valid as possible, it was minimally controlled and multitasking was allowed. We gave participants the freedom to have social interactions during the study, and perform the requested activities anyway they chose. For example, we asked participants to take a short walk but did not explicitly instruct them where to go. Likewise, when asked to interact with their phones, some participants checked email and exchanged text messages with friends while others browsed the web.

Figure 2: Examples of bimanual eating action observed in the user studies. At the top, a male participant performs a symmetric bimanual intake action while having yogurt (i.e., the non-dominant hand lifts the cup, mirroring the dominant hand). At the bottom, the same type of action can be observed as the female participant lifts a water cup with both hands. On the right side, asymmetric bimanual actions (i.e., cutting food) are weaved with unimanual intake gestures with both hands.

4.1 Annotation

A video camera was setup in front of participants and each study session was recorded so that gestures and activities could be annotated (Figure 1). The annotation process involved coding food and drinking intake gestures. Based on empirical observations, we considered each intake event to last 4 seconds; this interval corresponded to 2 seconds before and 4 seconds after the exact timing the food or drinking cup reached the mouth. The video and corresponding sensor data were annotated with 8 possible labels that qualified food intakes gestures by grasp type (i.e., hand, spoon, fork, and drink) and by side (i.e., left and right). Non-eating activities were left unannotated, which resulted in the creation of a null class. To aid the annotation process, we used the ChronoViz tool [Fouse et al. 2011].

5. CAPTURE AND CLASSIFICATION

The Microsoft Band[1] was used as the sensor capture device. It contains both an accelerometer and a gyroscope, thus providing 6 DoF inertial sensor data. Since participants wore two Bands at a time during the study, one on each wrist, 12 streams of inertial data were captured; the sampling rate was set to 30Hz. The data was transmitted in real-time to a mobile phone (iPhone 6S, iOS version 10.2.1)

---

[1]http://www.microsoft.com/microsoft-band

using the Bluetooth Low Energy communication protocol, and was saved locally on the phone.

The data was analyzed at the end of the study sessions. We first preprocessed the sensor streams using an exponentially-weighted moving average (EMA) for noise reduction, and applied either a scaling (i.e., MinMax) or normalization (i.e., L1 or L2 norm) function to the data. For feature extraction, we used a sliding window approach with 50% overlap and computed 5 features for each frame: *mean*, *variance*, *skewness*, *kurtosis*, and *root mean square (RMS)*. We chose these features because we have found them to be effective at providing a discretized and compact representation of the raw sensor signals in gesture recognition tasks. The output of the frame extraction step was a 60-dimensional feature vector per frame (5 features for each of the 12 streams of sensor data), which we passed on to a Random Decision Forest (RDF) classifier. We chose to use a RDF classifier because in our experience it has shown superior performance at non-linear modeling tasks if compared to other algorithms. We used the Scikit-learn [Pedregosa et al. 2011] RDF implementation.

6. RESULTS & DISCUSSION

In total, the 14 participants performed 1184 food intake gestures across all eating activities. Of the total, 679 were performed with the right hand, and 505 with the left hand. Considering that all participants declared themselves as right-handed, the number of intake gestures performed with the left hand was surprisingly high. This discrepancy occurred because some participants reported hand dominance that was not reflected in practice, underscoring the difficulty of instrumenting individuals with sensors for activity tracking. Additionally, some individuals consider themselves right-handed for some tasks but not others (e.g., writing vs. eating).

During analysis, we experimented with different sliding window sizes and preprocessing operations. The charts in Figure 3 shows the effect of sliding window size on *aggregate* eating detection performance (F-Score) as a function of wrist instrumentation across all participants (leave-one-participant-out cross-validation). When L1 or L2 normalization is applied to the sensor data, it is possible to recognize eating activities with a wrist-mounted device on the non-dominant hand (blue line) almost as well as with a sensing device on the dominant hand (green line), or when both hands are instrumented (yellow line).

While the non-dominant hand is not the one performing unimanual intake gestures, these gestures are not the only indicators of eating activity; there are many eating-related hand gestures performed by the non-dominant hand that are also behavior markers of eating. These other gestures reflect Guiard's asymmetric division of labor in bimanual action; our classifier was able to learn these complementary, and sometimes mirroring motions during training.

Interestingly, we discovered that normalization was key to this result. As shown in the bottom chart in Figure 3, when the data is *not* normalized, performance drops significantly in the case where the sensing device is placed on the non-dominant hand only. This is because L2 normalization is not only rotational invariant [Ng 2004], but it also equalizes all the gestural data by establishing a common reference and scale, giving prominence to the subtler motions of the non-dominant hand that would have been largely discarded as
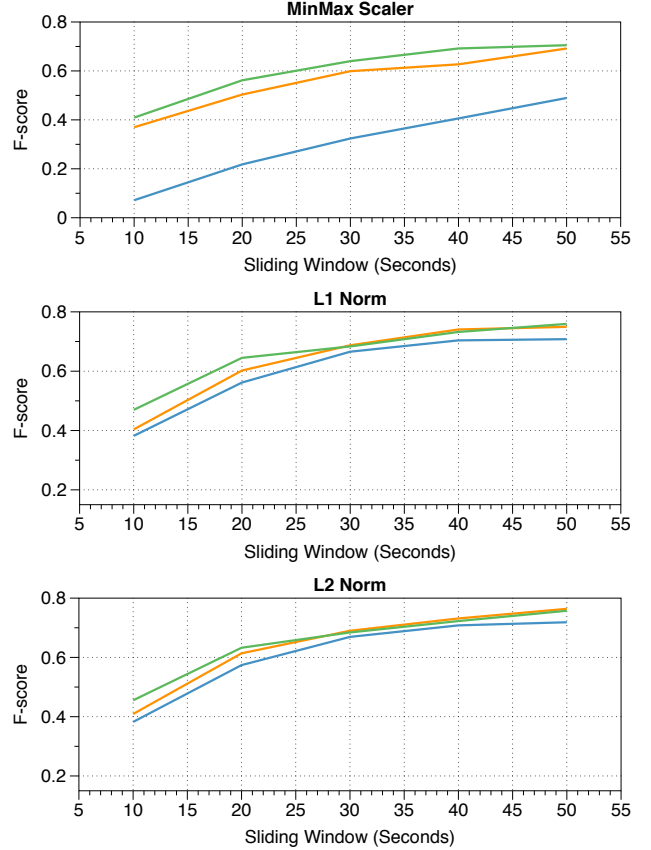

Figure 3: In these graphs, it is possible to see the effect of sliding window size on eating detection performance (F-Score) across all participants in the lab study (LOPO). When L1 or L2 normalization is applied to the sensor data, it is possible to recognize eating activities with a wrist-mounted device on the non-dominant hand (blue line) almost as well as with a sensing device on the dominant hand (green line) or on both hands (yellow line).

noise otherwise. As evidence of these subtler gestures of the non-dominant hand, we noticed that many participants regularly mixed their food together in a circular pattern when holding a fork and knife; this could have been an attempt to cool the food off, or was indicative of participants' eating styles. Regardless of the motivation, the mixing gestures were unique to the eating activities. Additionally, upon review of the study session videos, we saw evidence of symmetric bimanual food intake actions. While having a yogurt, P2 held the food container with his non-dominant hand and the spoon with his dominant hand (Figure 2). At the same time he lifted the dominant hand to bring yogurt to his mouth with the spoon (i.e., intake gesture), he also slightly raised the container with his non-dominant hand. In effect, the non-dominant hand was mirroring the intake gesture performed by the dominant hand. This effect was also observed as participants drank a cup of water.

Preprocessing	Left Hand (F-score)	Right Hand (F-score)	Both Hands (F-score)
MinMax Scaling	**48.9%**	70.5%	69.1%
L1 Norm	**70.7%**	75.9%	74.9%
L2 Norm	**71.8%**	75.7%	76.3%

Table 2: Best leave-one-participant-out performance results for eating detection (sliding window = 50 seconds) when taking into account sensor data from the instrumented left hand, right hand, and both. Notice the gain in performance with the left-hand when either L1 or L2 normalization is applied to the data during pre-processing (also shown in Figure 3).

7. CONCLUSION

The current paradigm in eating detection with practical, wrist-mounted inertial sensing hinges on identifying intake gestures with off-the-shelf devices placed on the dominant hand. Despite numerous attempts, tracking food intake gestures this way remains a difficult undertaking, due in large part to the high number of false positives. In this work, we provide evidence that subtle gestures and hand motions beyond unimanual food intake gestures are closely tied to dietary activity. This perspective is supported by Guiard's asymmetric bimanual action theory[Guiard 1987] and offers the possibility of new approaches for practical eating detection.

The key contribution of our paper is provide evidence of a new direction in eating detection with commodity sensors. This is particularly significant because it provides grounds for eating behavior tracking with one wrist-mounted device placed on the non-dominant hand. Technically, we show that a larger set of arm and hand movements beyond food intake gestures can be used to predict eating activities when L1 or L2 normalization is applied to the sensor data. It is worth noting that the primary aim of this work is not to advance the state of the art in terms of performance numbers; instead we propose an alternative way for reasoning about eating detection with wrist-mounted sensor, offer empirical evidence of its validity, and provide a theoretical underpinning for the approach. Finally, to encourage further work in the field of dietary monitoring and facilitate the validation of our analysis, we are making the data collected in our study available to the community as a public dataset.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[Amft and Tröster 2009] Oliver Amft and Gerhard Tröster. 2009. On-Body Sensing Solutions for Automatic Dietary Monitoring. *IEEE pervasive computing* 8, 2 (April 2009).

[Bulling et al. 2014] A Bulling, U Blanke, and B Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* (2014).

[Choe et al. 2017] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A. Epstein, Felicia Cordeiro, Matthew Kay, Gregory D. Abowd, Tanzeem Choudhury, James Fogarty, Bongshin Lee, Mark Matthews, and Julie A. Kientz. 2017. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing* 16, 1 (2017), 74–84. DOI: http://dx.doi.org/10.1109/MPRV.2017.18

[Dong et al. 2013] Yujie Dong, Jenna Scisco, Mike Wilson, E Muth, and A Hoover. 2013. Detecting periods of eating during free living by tracking wrist motion. *IEEE Journal of Biomedical Health Informatics* (Sept. 2013).

[Fouse et al. 2011] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D Hollan. 2011. ChronoViz: a system for supporting navigation of time-coded data. *CHI Extended Abstracts* (2011), 299–304.

[Guiard 1987] Yves Guiard. 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior* 19, 4 (1987), 486–517.

[Hatori et al. 2012] Megumi Hatori, Christopher Vollmers, Amir Zarrinpar, Luciano DiTacchio, Eric A Bushong, Shubhroz Gill, Mathias Leblanc, Amandine Chaix, Matthew Joens, James AJ Fitzpatrick, and others. 2012. Time-restricted feeding without reducing caloric intake prevents metabolic diseases in mice fed a high-fat diet. *Cell metabolism* 15, 6 (2012), 848–860.

[Hinckley et al. 2010] Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. 2010. Pen + Touch = New Tools. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 27–36. DOI: http://dx.doi.org/10.1145/1866029.1866036

[Junker et al. 2008] Holger Junker, Oliver Amft, Paul Lukowicz, and Gerhard Tröster. 2008. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition* 41, 6 (June 2008), 2010–2024.

[Kim et al. 2012] Hyun-Jun Kim, M Kim, Sun-Jae Lee, and Young Sang Choi. 2012. An analysis of eating activities for automatic food type recognition. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*. 1–5.

[Li et al. 2005] Yang Li, Ken Hinckley, Zhiwei Guan, and James A. Landay. 2005. Experimental Analysis of Mode Switching Techniques in Pen-based User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*.

ACM, New York, NY, USA, 461–470. DOI: http://dx.doi.org/10.1145/1054972.1055036

[Merck et al. 2016] Christopher Merck, Christina Maher, Mark Mirtchouk, Min Zheng, Yuxiao Huang, and Samantha Kleinberg. 2016. Multimodality Sensing for Eating Recognition. *Proceedings of PervasiveHealth* (March 2016), 1–8.

[Napier and Tuttle 1993] J Napier and R Tuttle. 1993. *Hands*. Princeton Science Library. Princeton University Press.

[Ng 2004] A Y Ng. 2004. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning.*

[Pedregosa et al. 2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[Rahman et al. 2015] Shah Atiqur Rahman, Christopher Merck, Yuxiao Huang, and Samantha Kleinberg. 2015. Unintrusive eating recognition using Google glass. In *PervasiveHealth '15: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering.

[Rahman et al. 2016] Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *DH '16: Proceedings of the 6th International Conference on Digital Health Conference*. Cornell University, ACM.

[Thomaz et al. 2015] Edison Thomaz, GD Abowd, and Irfan Essa. 2015. A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing. In *UbiComp '15: Proceedings of the 2015 ACM international joint conference on Pervasive and ubiquitous computing*. 1–12.