Multimodal Food Journaling

Hyungik Oh
University of California, Irvine
Irvine, California
hyungiko@uci.edu

Jonathan Nguyen
University of California, Irvine
Irvine, California
jonatn8@uci.edu

Soundarya Soundararajan
University of California, Irvine
Irvine, California
soundas1@uci.edu

Ramesh Jain
University of California, Irvine
Irvine, California
jain@ics.uci.edu

ABSTRACT

A food journal is essential for improving health and well-being. However, journaling every meal is extremely difficult because it depends on user initiative and intervention. Current approaches to food journaling are both potentially inaccurate and tedious, causing people to abandon their journals very soon after they start. In this paper, we propose a proactive and reactive mechanism that can significantly reduce user initiative while still remaining highly accurate. We first suggest a novel eating moment recognition technique using heart rate and activity patterns to trigger a food journaling process in a proactive manner. We then begin the food journaling process via voice command which utilized natural language processing when logging meals, which increases the ease of reactive self-reporting. Lastly, we enhance the food journal by automatically assessing ecological moments of eating activity through our personal chronicle system. We verified the method from a feasibility study conducted with three people for three months in their day-to-day lives. Our approach is designed to be unobtrusive and practical by leveraging multi-modal sensor data through the most common device combination of a smartphone and wearable device.

CCS CONCEPTS

• Information systems → Mobile information processing systems; **• Human-centered computing → Ubiquitous computing**; **Mobile computing**; **Ambient intelligence**; **Smartphones**;

KEYWORDS

Eating Moment Recognition; Food Journaling; Ecological Momentary Assessment; Lifelogging; Pervasive Health; Personicle;

1 INTRODUCTION

You are what you eat.

HealthMedia'18, October 22, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5982-5/18/10.
https://doi.org/10.1145/3264996.3265000

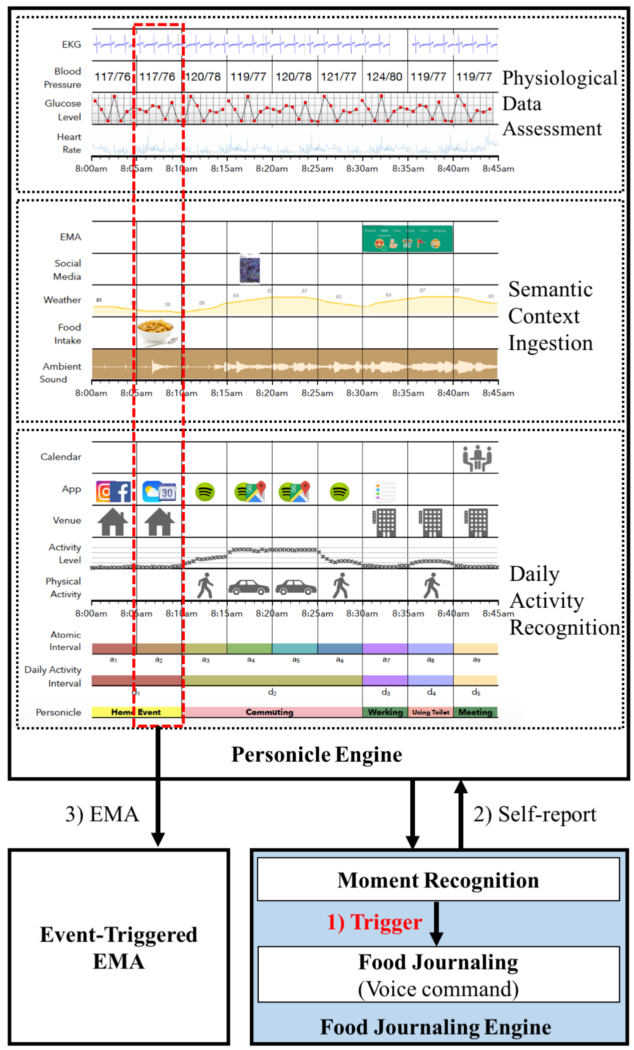

Figure 1: Assessing an enhanced ecological moment of eating activity through *Personicle* based food journaling.

The foods and drinks we put in our bodies have a direct impact on our health and well-being. There have been numerous medical studies showing that unhealthy dietary habits can be a major cause of diseases such as obesity, kidney disorder, CVD[1], cancer, and diabetes [21, 34]. Clearly a well-balanced diet is very important to stay healthy.

Food journaling has been demonstrated to encourage people to develop healthier dietary habits since it provokes self-reflection that can play a significant role in behavior change. Therefore, health care professionals and people who suffer from health-related disorders have tried to maintain a food journal so that they can analyze the health effects of their dietary intake [37]. However, even though food journaling has been the main method of monitoring dietary intake for a long time, unobtrusive ways of keeping a food journal remain relatively undeveloped.

The traditional method of keeping a food journal is manually recording meals in as much detail as possible by including the portion size, number of servings and calories, time, location, or even the people around us. This detailed description is effective, but it is very easy to forget or procrastinate logging food entries, which then results in more difficulty recalling meals eaten or even early abandonment of the journaling process. Although remarkable technical progress in automating the food journaling process has been made, it is still highly dependent on the user to take initiative and then requires them to do things such as taking pictures of their food, scanning barcodes, or searching for foods in a food database. These methods tend to be unreliable and require many actions on the user's part which can then lead to the problems endemic, such as inaccurate or missed food entries and early abandonment.

There are currently two main challenges in improving food journaling: (1) triggering a food journaling process in a timely, proactive manner, and (2) improving the reactive self-reporting procedure while preserving high measuring accuracy. In this paper, we offer an alternative method to current food journaling through a event-triggered Ecological Momentary Assessment (EMA). We try to consider both the proactive and reactive perspectives that can unobtrusively enhance the event-triggered EMA as in Figure 1 and thus move forward as fully-automated food journaling.

To solve the first challenge in food journaling, a timely reminder is essential. The best time for food journaling is when people start eating a meal since they know what they are eating at that moment. Thus, our approach begins with finding eating moments so that we can trigger a food journaling process at the correct time. More specifically, we find two kinds of eating moments; one is "eating at a restaurant", and the other one is "eating at home". Our previous research [15, 16, 26, 27] has developed technology to automatically recognize the former eating moment through smartphone based *Personicle*, which provides a person's time-ordered list of daily activities. However, there have been difficulties recognizing "eating at home" due to the lack of available smartphone sensors. In this paper, we try to solve the latter problem by pulling heart rate signal in *Personicle*.

We then try to keep a food journal through what we call the event-triggered EMA. To do this, we provide an environment that the user can log their meals by describing what they just ate via voice commands. Essentially, taking pictures of foods and barcodes to create food entries have shown to be inaccurate or inconvenient. For this reason, we offer an alternative, which is to use the voice commands to create food entries by using speech-to-text technologies and natural language processing. Meanwhile, the *Personicle* system automatically assesses the user's ecological moment by including the food entries as well as various contexts of the eating moment and thus unobtrusively complete the event-triggered EMA. Our main contribution in the area of food journaling is 1) providing an event-triggered EMA to automate the food journaling process, thereby 2) encouraging people to keep a well-balanced diet, as well as 3) helping them develop healthier dietary habits. We make these contributions by providing a general eating moment model that can automatically recognize the starting moment of eating, and then prompting the user to begin a voice command food journaling method. We validate our approach with an experiment for 3 months with 3 users who are using the *Personicle* system with Fitbit Charge 2 or Blaze. Our food journaling scenario is as follow:

(1) Users install *Personicle* on their Android phone and start using it with a Fitbit device, such as Charge 2, Blaze, Ionic, or Versa, which are the most common devices in the market.
(2) After a cold start period lasting a week, the *Personicle* system starts requesting the voice command food journaling whenever it recognizes a starting moment of breakfast, lunch, or dinner. It generates a unique pattern of vibration so that it can let the user know that it's time to make a food journal.
(3) Then, the user simply speaks whatever he is eating at that moment, such as "I'm eating a slice of pizza with buffalo wild wings and a cup of Coke for lunch".
(4) After that, the *Personicle* system extracts food items (e.g., pizza, buffalo wild wing, coke), quantity of the food (e.g., one slice, a cup), and meal type (e.g., lunch).
(5) Finally, the *Personicle* system makes an event-triggered EMA by capturing other contexts around the eating moment, such as glucose level, stress level, emotion, weather, location, other people with the user, or even past events before the eating activity.

2 RELATED WORK

Food journals are currently the most commonly used method for analyzing dietary intake. An early method of keeping a food journal was through anecdotal summaries, such as lengthy interviews and questionnaires [10, 24]. This method has shown to be a cumbersome and inefficient way of monitoring dietary intake. Recall-based paper diaries have been another popular alternative to understanding dietary habits of people [6]. However, both of these methods as of recently have been phased out in favor of mobile food journals driven by advancements in technology. Commercial mobile applications such as MyFitnessPal[2], Fitbit[3], or Bixby Vision [4], support food databases so that users can easily and accurately log calories and nutritional information. Utilizing databases takes the guess work out of logging calories and result in more accurate food journals.

[1] Cardiovascular disease

[2] https://www.myfitnesspal.com/

[3] http://www.fitbit.com/

[4] https://www.samsung.com/global/galaxy/apps/bixby/vision/

Additionally, most of these mobile applications also support features such as barcode scanners and shortcuts for commonly eaten foods in order to quickly journal food information [35]. However, even though these technologies increase convenience and usability for maintaining a food journal, it is still highly dependent on the user to take initiative and remain consistent in their food logging [30].

Researchers in the field of computer vision have started to incorporate food image recognition in order to make food journaling more convenient and consistent for users. They have addressed the challenges in image recognition by developing machine/deep learning algorithms to recognize food items [5, 9, 19, 20, 25, 39]. For example, FoodLog has contributed to a record of users' food intake simply by taking photos of their meals [1, 2]. It also allows users to input textual descriptions based on image retrieval techniques. This kind of approach mainly uses mobile applications or wearable cameras (e.g., DietCam [22], Menu-Match [4], FoodCam [20]) for food recognition, assessment, and journaling. In addition to food image recognition, food quantity estimation has been another important aspect of the research to automate the assessment of food intake. [28, 29, 36]. However, the classification of food image is still a very difficult task and is still fairly inaccurate, since there are various confounding factors, such as visually similar foods, home made foods, quality of photos, and lighting conditions [30].

Another important challenge of automated food intake monitoring involves eating moment recognition. Since Stellar et al. recognized eating event by measuring tongue pressure through oral strain gauge in 1980s [33], researchers have used various sensing modalities for eating moment recognition. One of these modalities used an acoustic sensor to monitor swallowing and chewing sound through the ear, laryngopharynx [32], or neck [8, 38]. Some others have utilized on-body inertial sensor to detect eating or utensil (e.g., fork, spoon) gesture [3, 11, 18, 34]. More recently, researchers have looked to analyzing the heart rate response for eating moment recognition. Shinji et al. analyzed short-term and long-term features of heart rate changes, and revealed that there is another heart rate peak after eating for few hours [14]. Despite all the progress made in this field, most of these proposed methods are impractical for real-life usage, requiring multiple on-body sensors, or suffer from several limitations, such as weak gesture model, or experimental constraints (e.g., time, situation). To the best of our knowledge, there is no approach seeking for unobtrusive food journaling that automates the process of keeping a food journal by utilizing all the contexts around eating activity. The biggest difference our work offers is that we generate a event-triggered EMA by automatically assessing ecological moments of eating activity at the correct time.

3 EATING MOMENT RECOGNITION

We try to trigger a food journaling process for two different kinds of eating moments, "eating at restaurant", or "eating at home". Currently, we have successfully recognized eating activity if people are eating outside of their homes, such as restaurants, or their favorite breakfast/lunch/dinner spot [26]. However, it has been difficult to recognize when the user is eating at home since we were unable to find useful features that can classify "eating" from "home event". In this section, we propose a novel method of recognizing eating moment by pulling heart rate signal in the chronicle of daily activity. This approach mainly focuses on finding the starting moment of the eating.

We first hypothesize that heart rate is increased when people start eating a meal, and then maintain that increased rate while they are eating. This is supported by the studies published in psychophysiology, nutritional science, and electro-cardiology, which have proved that heart rate is generally higher after meals [13, 17, 31]. We also propose another hypothesis that there is a unique activity pattern before "eating at home", such as preparing food, or moving to dinning room. We try to build a general eating moment model, which is designed to learn the aforementioned features. Figure 2 shows the sensor data processing pipeline starting with multi-modal sensor data tracking, such as heart rate and step count. We segment the low-level signals whenever the pattern of physical activity is changed, and then extract features from the segmented results, and finally recognize eating moments through machine learning algorithms.

Algorithm 1 Double Daily Activity Segmentation using RBIG

Input: current atomic interval $\mathbf{A}_i$, $\mathbf{a}_i$, seed atomic interval $\mathbf{S}_j$, $\mathbf{s}_j$ where $\mathbf{a}_i$ and $\mathbf{s}_j$ are 1 minute interval
Output: daily activity interval set $\mathbf{R}$;
1: Set $\mathbf{S}_j = \mathbf{A}_i$ if $i = 0$ and $j = 0$, or $\mathbf{S}_j = \emptyset$, and then set $k = 0$, $m = 0$, $p = 0$;
2: **Repeat**
3: Wait for next atomic interval, $\mathbf{A}_i = \mathbf{A}_{i+1}$, $i = i + 1$;
4: Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{A}_i$;
5: Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{S}_j$;
6: Calculate $\delta(i)$;
7: If $\delta(i) = 1$, make a daily activity interval $\mathbf{R}_k$ by segmenting from $\mathbf{A}_j$ to $\mathbf{A}_i$;
8: Set $p = i$, reset $i = j$, $\mathbf{a}_i = \mathbf{S}_j$, $\mathbf{s}_j = \mathbf{S}_j$;
9: **Repeat in the daily activity-interval** $\mathbf{R}_k$
10: Assign next atomic interval $\mathbf{a}_i = \mathbf{a}_{i+1}$;
11: Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{a}_i$;
12: Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{s}_j$;
13: Calculate $\delta(i)$;
14: If $\delta(i) = 1$, make a daily activity-interval $\mathbf{r}_m$ by segmenting from $\mathbf{a}_j$ to $\mathbf{a}_i$;
15: Set $m = m + 1$, $j = i$, set new seed atomic interval $\mathbf{s}_j = \mathbf{a}_i$
16: **until** i == p
17: Set $k = K + 1$, $j = p$ set new seed atomic interval $\mathbf{S}_j = \mathbf{A}_p$
18: **until** the system is terminated.
19: **return R, r**

3.1 Double Segmentation

We define segmentation as a process of partitioning a day into multiple sets of daily activity intervals. To segment a day, we assume that transitions of physical activity pattern (e.g., moving to non-moving, or non-moving to moving) can be involved in the changes of other attributes, such as location, which can then be considered

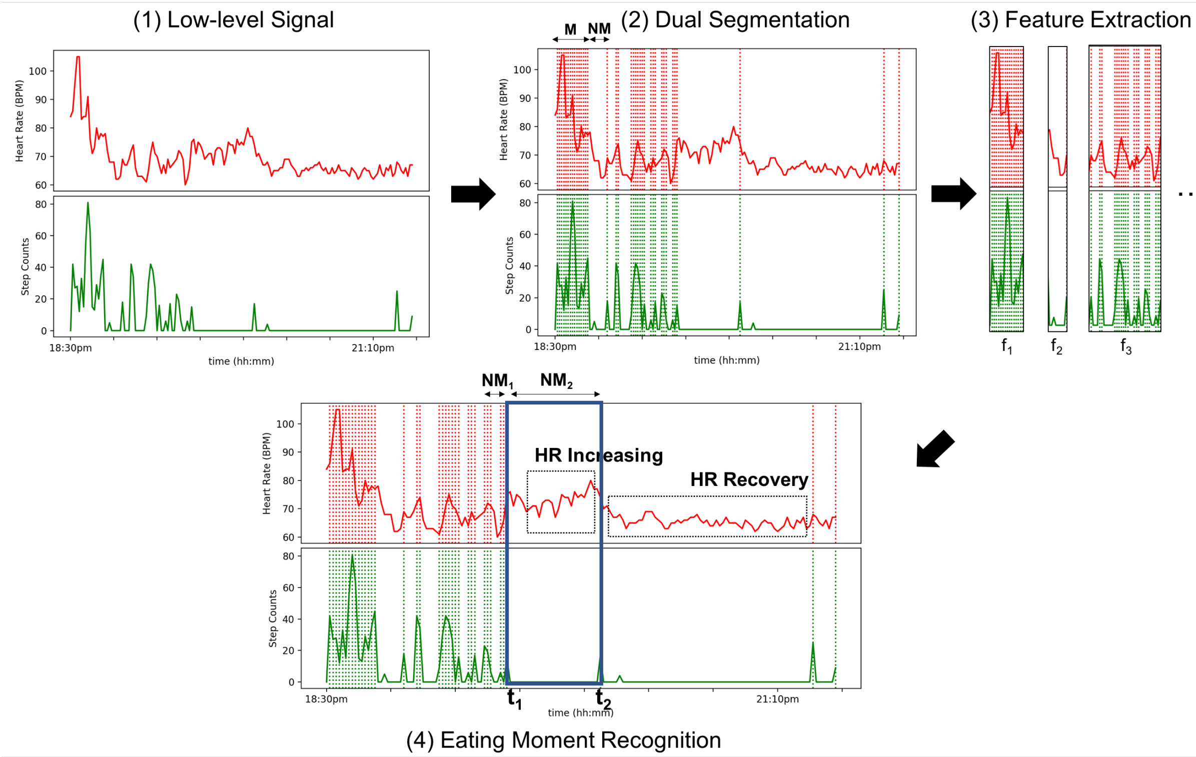

Figure 2: The sensor data processing pipeline for the eating moment recognition. M**: a moving type of daily activity interval,** NM**: a non-moving type of daily activity interval,** NM_1**: a sample of non-moving type of daily activity interval before eating,** NM_2**: a sample of non-moving type of daily activity interval indicating an eating moment.**

as ending one daily activity (e.g., "home event") and starting another (e.g., commuting) [26, 27]. Thus, as shown in Figure 2, we segment sequential atomic intervals of the day into moving (M) or non-moving (NM) type of daily activity intervals.

As shown in Figure 1, we have used five-minutes as the length of an atomic interval in order to segment the sequential atomic intervals into coarse-grained daily activity intervals. In our previous research [26], we finally classified this coarse-grained daily activity intervals into daily activity, such as "home event". However, in this research, we decrease the length of atomic interval from five-minutes to one-minute so that we can re-segment coarse-grained daily activity intervals into fine-grained daily activity intervals, and thus classify "eating at home" activity from "home event". To do this, we suggest a recursive binary interval growing technique (RBIG).

Recursive Binary Interval Growing (RBIG): Algorithm 1 shows the process of recursively re-segmenting a coarse-grained daily activity interval into fine-grained daily activity intervals. We first use five-minute for segmenting a coarse-grained daily activity interval. Once we segment the coarse-grained daily activity interval, we come back to the seed atomic interval, and then re-segment the coarse-grained daily activity interval into fine-grained daily activity intervals by using one-minute atomic intervals. As shown in Algorithm 1, we assign the seed atomic interval S_j at the beginning of the process, and then start calculating the similarity $\delta(i)$ between the seed atomic interval S_j and the incoming atomic intervals A_i, every interval minutes (e.g., 5 min, 1 min). $\delta(i)$ can be formulated as follow:

$$\delta(i) = \|f(S'_j) - f(I'_i)\|_2^2 \quad (1)$$

where S'_j is $\{l_j, t_j\}$, I'_i is $\{l_i, t_i\}$, $f(x)$ is a classifier to identify if an atomic interval is the moving (1) or non-moving (0) type of interval. When $\delta(i)$ is equal to 1, we segment the atomic intervals from I_j to I_i, and then make it as a coarse-grained daily activity interval R_k. After that we try to segment R_k into fine-grained daily activity intervals r_m by repeating the same process with one-minute atomic interval.

3.2 Feature Extraction and Selection

We extract and select eating moment features from the fine-grained daily activity interval. We heuristically explored the 3 months worth of daily activity intervals so that we can identify latent features underlying the visible sensor data streams. In our exploration, as shown in Figure 2, we first could see that heart rate is increased when people start eating, as seen in NM_2 between time t_1 and t_2, and then the increased heart rate remained high during the meal time. We next tried to see how average heart rate is different between the eating moment and the past since it can be another unique feature that determines the starting moment of eating. To

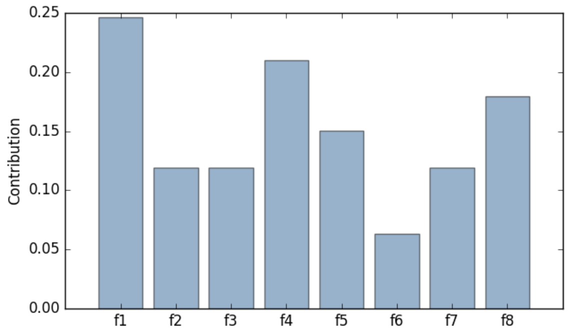

Figure 3: Contribution of each feature for eating moment classifier. $f1$: past average heart rate, $f2$: current average heart rate, $f3$: Δ average heart rate, $f4$: the amount of past NM time, $f5$: the amount of past M time, $f6$: the number of past steps, $f7$: the amount of moving time (past 30 min), $f8$: heart rate variation.

do this, we excluded all the moving type of daily activity intervals, which can highly affect the increase of heart rate, and then tried to correctly compare the difference in average heart rate. Such heart rate effects of dietary intake can be seen in between NM_2 and NM_1. Based on this finding, we extracted eating moment features, such as average heart rate, heart rate variation, and heart rate difference between current and past daily activity intervals. In addition, we also extracted more features from activity patterns before eating. As shown in Figure 2, there will always be a certain amount of moving time just before beginning an eating activity due to events like preparing the food, or moving to the dining room. Thus, we also extracted features such as step count, and the amount of moving and non-moving time.

Lastly, we used the Correlation-based Feature Selection (CFS) criteria so that we can select the best subset of extracted features [12]. This algorithm evaluates how accurately all the features in the feature subset are indicative of the target class. It also can evaluate which features are not correlated with each other by providing complementary information for each of them [30]. To evaluate our heuristically extracted features, we first trained the eating moment classifier with all the features, and then compared the recognition performance (F-measure) to those of other classifiers, which are trained without a particular feature subset. As shown in Figure 3, all the extracted features show some performance degradation, which means these features are all highly indicative of the starting moment of eating activity. Based on this result, we selected all the extracted features.

3.3 Eating Moment Recognition

With the eating moment features, we tried to build a general eating moment model that can classify the fine-grained daily activity intervals into eating or non-eating moments.

There are several things to keep in mind regarding general eating moment modeling. First, it requires relative values due to the fact that everyone has varying heart rate ranges. Thus, we converted the heart rate C of each individual data into heart rate levels, which are discretized w-dimensional space, by a vector $\bar{C} = \bar{c_1}, \bar{c_2}, ..., \bar{c_i}$. To do this, we used a discretization technique, Symbolic Aggregate Approximation (SAX) that reduces the time series of arbitrary length n into the w-dimensional space as follows [23]:

$$\bar{c_i} = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j \tag{2}$$

Second, we also converted the activity and time features (e.g., moving pattern, step count, and the amount of moving and non-moving time into discretized levels) given that these also differ from person to person. Furthermore, we reflected day-night differences in body temperature and heart rate by training breakfast, lunch, or dinner model separately.

For training the classifier, a Support Vector Machine (SVM) with a Radial Basis Function kernel (RBF) was applied to the training dataset [7]. In the test phase, we ran the classifier whenever the *Personicle* system finds the fine-grained daily activity intervals. If the interval is classified as a starting moment of "eating at home", we assume the whole range of this interval as the eating moment, as shown in t_1 and t_2 in Figure 2.

4 VOICE COMMAND FOOD JOURNALING

As a first step towards building a voice command food journaling engine, we define a basic sentence protocol that has to be spoken by a user to apply to text analysis. Then, the prototyped solution accepts a voice based input describing food intake, transcribes the input to text by using Google voice API[5], breaks down the input sentences according to the predefined protocol, and then extracts information for keeping a food journal. In this section, we take the full advantage of using APIs in order to ensure maximum results.

4.1 Protocol

The major components that are important on a food journal is food item, meal type and quantity. This information enables the ability to obtain nutrition information and calorie intake by querying a food database, such as USDA Food Composition Databases[6]. Therefore, we suggest users to include the aforementioned information with an actuating verb, such as "eat" or "have", when they describe what they are eating. The actuating verb would help to increase the accuracy of voice command analysis since it points out the most important sentences of all conversation. The quantity, food item, and meal type information should be listed sequentially after this actuating verb. Here are some examples.

- **Protocol 1:** *I'm eating* (actuating verb) / *a* (quantity) / *cheeseburger* (food item) / *for lunch* (meal type)
- **Protocol 2:** *Two* (quantity) / *garlic naans* (first food item) / *and a cup of* (quantity) / *Coke* (second food item)

The first example shows a complete protocol that we want to see from the voice command. It has an actuating verb "eat", and then food item "cheeseburger", quantity "one" and meal type "lunch". In addition to the complete form, we also can accept a simplified protocol as can be seen in second example. We then try to obtain

[5]https://cloud.google.com/speech-to-text/
[6]https://ndb.nal.usda.gov/ndb/doc/index

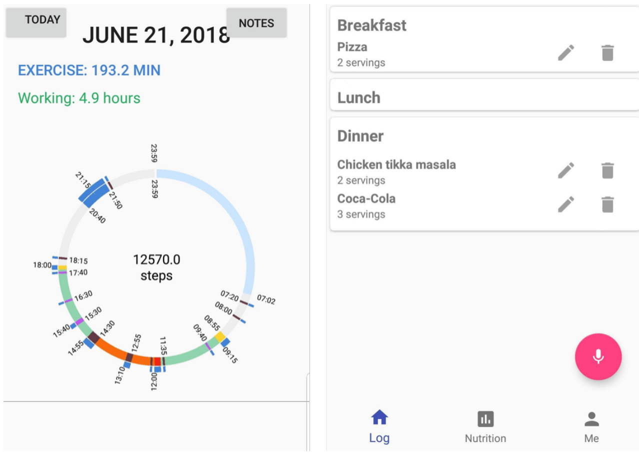

Figure 4: Screenshots of *Personicle* system including daily event recognition and food journaling.

food items "garlic naan", "Coke", and quantities of the foods "two", "a cup". Multiple food information also can be acceptable once it is listed sequentially.

4.2 Information Extraction

After the speech has been converted to text via Google Voice API, the next task is to extract the key information out of the text. We utilize a natural language processing API, TextRazor[7], so that we can accuarately extract the keywords in a sentence and the classification results of those keywords. Based on the result, we first find the sentences, which include actuating verbs, such as "eat", or "have". We then filter out all the keywords from the sentences by checking for foods, meal types and numbers, keeping only the necessary information to create a food entry. After that, we extract the nearest numbers from the food items to map the quantity to the food. Lastly, if there is no meal type in the sentence, we extract this information from the tense of the verb or current time that the food entry was created.

5 EVENT-TRIGGERED EMA

The current phase of our event triggered EMA is shown in Figure 1. We can automatically assess personal ecological moments without any questionnaires since our *Personicle* system continuously monitors the chronicle of daily events as well as semantic contexts and physiological signals. Therefore, once the *Personicle* system recognizes an eating moment, we can create an EMA of the eating activity consisting of stress level, glucose level, emotion, weather, location, other people with the user, and even past events before eating and their frequency. Additionally, if the user reacts to the voice command request, we also include the food eaten, the quantity of the foods, the nutrition value, and the calorie intake in the event-triggered EMA. The ultimate goal of our event triggered EMA is to fully automate the food entry process, and thus keep a food journal without any user interventions, such as taking pictures.

[7] https://www.textrazor.com/

We see the potential of automating the food entry process in that there are distinct differences in heart rate patterns depending on the food type. Such research could open many opportunities to support innovative studies in pervasive health.

6 EXPERIMENTAL VALIDATION

In this section, we first elaborate how participants collected multimodal sensor data and how they labeled starting moments of their eating activity. We then validate our eating moment recognition method by evaluating the performance of the general eating moment classifier. After that, we verify the voice command food journaling using the two kinds of protocols defined in Section 4.1.

6.1 Data Collection

We implemented a *Personicle* system on Android as in Figure 4. This *Personicle* system always runs in the background, collects both smartphone and wearable sensor data as described in Figure 1, and stores the data in Google Firebase Database[8] in real-time. To collect experimental data, we hired three participants who are using Galaxy S9 plus (OS 8.0.0) with Fitbit Blaze, Galaxy S8 (OS 8.0.0) with Fitbit Blaze, and Google pixel (OS 8.1.0) with Fitbit Charge2 for three months, and asked them to install the *Personicle* application on their smartphone.

The three participants manually labeled eating moments by using Samsung Health[9], KhanaPal[10], and the *Personicle* application, respectively. We then used the dual segmentation technique (RBIG) so that we can extract fine-grained daily activity intervals, which include the labeled moments as described in Section 3.1. Therefore, the participants didn't have to provide all the start and end times of each eating activity, but simply labeled a time stamp in the meal time. To guarantee the quality of the ground truth data, we requested them not to label the eating moment by guessing if they miss the food journaling time. The total numbers of atomic intervals, daily activity intervals, and eating labels were 90720, 1785, and 255 respectively.

6.2 Eating Moment Recognition

We evaluated our model by calculating precision, recall, and F-measure. We performed 10-fold cross validation on each participant's daily activity interval data and then averaged the results to obtain an overall result.

Since eating-labeled daily activity intervals had different segment sizes, from five minutes to more than an hour, we needed to make their size uniform so that we can train and test the eating moment classifier. Thus, we tried to find the optimal sub-segment size of the eating-labeled daily activity intervals that most represents a starting moment of eating activity. To do this, we first explored all the sub-segment sizes of eating-labeled daily activity intervals and found that 80.9% of them belong to a range between five minutes and twenty minutes. Then, we chose the intervals of 5, 10, 15, and 20 minutes among those ranges considering that *Personicle* system has a five-minute data processing interval. Afterwards, we trained

[8] https://console.firebase.google.com/
[9] https://www.samsung.com/us/support/owners/app/samsung-health
[10] https://play.google.com/store/apps/details?id=com.foodie.android

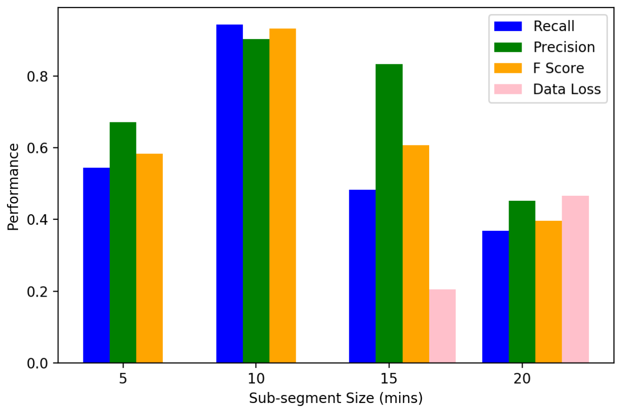

Figure 5: F-measure of the eating moment recognition across different sub-segment sizes.

the classifiers separately with the chosen sizes and tested the classifiers to see which sub-segment size most accurately recognized the starting moment of eating. Figure 5 shows the performance of eating moment recognition across the four different sub-segment sizes. From this experiment, we can see that using a 10-minute sub-segment size provides better results compared to other sub-segment sizes. This indicates that if a sub-segment size is too long or too short, it is difficult to properly represent the features of eating moments. We also can see that there is data loss when the sub-segment sizes are too long. It means that there are many number of eating activities that are less than 15 minutes. Furthermore, the results shown in Table 1 indicate that the performance of eating moment recognition across different sub-segment sizes is not person-specific but can be generally applied for all the users. Based on these findings, we found that 10-minute is the most optimal sub-segment size for accurately reflecting the selected features as described in Section 3.2.

With the 10-minute sub-segment size, we next verified the performance of our general eating moment classifier. Table 2 presents the precision, recall, and F-measure of the eating moment classifiers with and without SAX algorithm. From this comparison, we can clearly see that the performance of all classifiers improved with the SAX algorithm. This indicates that the pattern of heart rate and activity around the eating moment are relatively similar from person to person. More interestingly, Table 2 shows that all the tested classifiers achieved significantly good performance. Even the 10-NN classifier shows comparable results to SVM on the precision, recall, and F-measure. It can signify that the eating moment sub-segment set can be easily separable even though it is in high dimensional space. However, the relatively low recall in the classifiers with SAX means that more unique features are still needed to classify more diverse eating moments. Considering that we are planning to add more features to the eating moment model, we chose the SVM classifier, which is low cost, and high speed, and fairly robust against over-fitting, especially in high-dimensional space.

Table 1: F-measure of each user's eating moment recognition across different sub-segment sizes.

User	Sub-segment Size			
	5 minutes	10 minutes	15 minutes	20 minutes
User 1	0.7500	**0.8686**	0.8236	0.8571
User 2	0.6000	**0.8889**	0.6667	0.0000
User 3	0.4000	**1.0000**	0.3333	0.3333

Table 2: Performance of eating moment classifiers trained with and without SAX algorithm in terms of Recall (R), Precision (P), and F-measure (F). *10-NN*: 10 Nearest Neighbors, *NB*: Naive Bayes, *RF*: Random Forest, *SVM*: Support Vector Machine.

Model	Without SAX			With SAX		
	R	P	F	R	P	F
10-NN	0.4286	1.0000	0.6000	0.8750	1.0000	0.9333
NB	0.8571	0.4615	0.6000	0.7778	0.8750	0.8235
RF	0.5714	0.8000	0.6667	0.7500	1.0000	0.8571
SVM	0.2857	0.6667	0.4000	**0.8750**	**1.0000**	**0.9333**

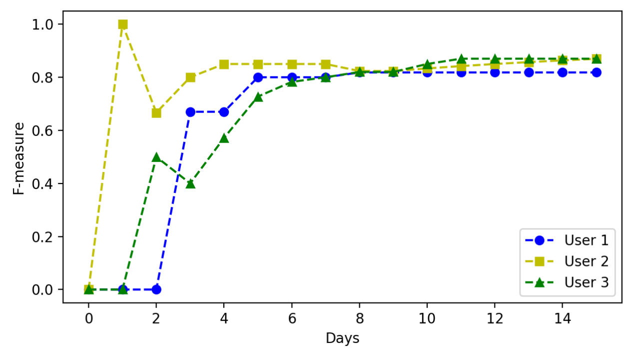

Figure 6: Performance of eating moment recognition using SVM classifier over time.

Lastly, we tested the performance of eating moment recognition over time. According to Lin et al., the small subset of data can deteriorate the efficiency of SAX since the discretization technique is based on normal distribution [23]. Considering that the *Personicle* system began with zero user data, our initial performance also could be highly affected by incorrectly discretized features. Because of this, we tried to find the cold start of our general eating moment classifier in order to understand how long it takes to obtain reasonable results. As we can see in Figure 6, there was a wide fluctuation in the recognition performance for the first few days. To determine the reason for this, we tried to draw a normal probability plot of heart rate data as shown in Figure 7. In the figure, the highly linear nature of the plots indicates that heart rate data comes from a Gaussian distribution. However, the first few days of each plot show that there was lack of heart rate data in high range, and thus resulted in incorrect heart rate discretization, especially between 70 bpm and 120 bpm. As an example, 70 bpm was discretized as level

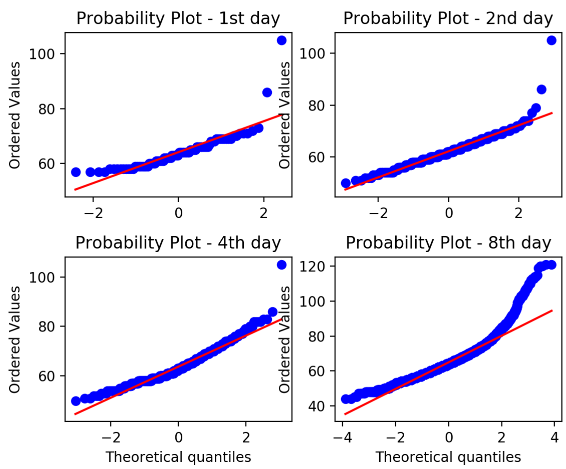

Figure 7: A normal probability plot of the distribution of evening heart rate over time.

8, 9, 8, and 7 out of 10 in the 1st, 2nd, 4th, and 8th day's normal probability plot, respectively. This small difference could significantly affect the eating moment recognition performance given that recognizing eating moment is highly dependent on heart rate variations.

Our results show great potential for automating the eating moment recognition in a practical manner using a smartphone and wearable device combination. However, the current research has a few limitations, namely sub-segment size (e.g., 10 minutes) and location (e.g., home) in the recognition process. While these constraints can be effective for the recognition of "eating at home" events, these can preclude the possibility of recognizing other eating events, particularly those that are less than 10 minutes, or those that happen during physical activity, or those occurring in outdoor places. Therefore, in our future research, we aim to analyze more distinct heart rate patterns, which are not only the analysis of starting moments, but also overall fluctuations in the entire eating moments. To do this, we will collect more diverse eating cases with more participants and try to analyze the effect of different factors on heart rate so that we can better understand correlations between heart rate and foods.

6.3 Voice Command Food Journaling

In this section, we evaluated the performance of voice command food journaling by testing predefined protocols. We selected the most popular food items from the following seven different food types: Indian, Korean, American, Chinese, Italian, Mexican, and Japanese. From this list of different food types, 41 test cases of the two protocols were made (sixteen of Protocol 1, and twenty five of Protocol 2). We evaluated these 41 test cases with three users who have different English accents, which are native English, Indian, and Korean, so that we can also see the effect of accent on the voice command results. Table 3 shows all the incorrect results that we obtained from our experiment. The meal types and quantities of

Table 3: List of incorrect results among all the test cases. A_1: Korean accent, A_2: Indian accent, A_3: Native English accent.

Food	Voice Command	Food Extracted		
		A_1	A_2	A_3
Indian	Idly	Italy	Elite	O
	Hot spicy sambhar rice	Some hard	O	sambar
Korean	Bingsu	Kingsville	O	O
American	I had large stack of 6 pancakes	O	6 cake	O
Chinese	Szechuan chilli chicken	Sachin	Sichuan	2chan
	Wonton	1 ton	O	O
Italian	Lasagna	O	Sonya	O
Mexican	Quesadilla	O	Jesse Diaz	O

the foods were correctly extracted from the test cases, and thus were excluded in the Table.

In the experiment, we found that there are three issues when trying to recognize food items. First, if the user is not familiar with food items, such as Wonton, Idly, Szechuan, Lasagna, Quesadilla, or Sambhar, these are not correctly recognized and extracted as shown in Table 3. Second, the voice recognition API has been mainly trained by a native English accent. For example, even though Bingsu and Idly were pronounced by their native accents, which are Korean and Indian respectively, the voice recognition API converted them into wrong words, such as Kingsville and Elite. Third, if there is a short silence between or within words, such as a pause in the word "pancakes" in Table 3, recognition accuracy is compromised. Articulating the voice recording method in the programming level will be able to improve this issue. To solve all of the aforementioned problems, we will use more contexts around the voice command moment or apply approximate string matching algorithms so that we can find the most close food item to the converted text.

7 CONCLUSIONS

Food journaling is a great way to improve health because it lets us monitor our dietary intake, but it can potentially be inaccurate and difficult to maintain. This paper builds towards the research to develop a unobtrusive food journaling method that automates the process of keeping a food journal via common wearable devices. Specifically, this paper focuses on recognizing a starting moment of eating activity to trigger a food journaling process in a timely, proactive manner. Thus, it describes the methodology behind automatically recognizing eating moment with the goal to build an event-triggered EMA. We also propose a voice command food journaling method which makes it simple to keep a food journal while still remaining highly accurate, and thus include the food entries in the event-triggered EMA. Such methods could play a very important role in applications for health and well-being study using personal food journaling data. Results obtained from the three participants show the potential of such an approach for the unobtrusive food journaling. We expect that our further research would allow for automatically recognizing food items at broad level considering that there are distinct differences in heart rate patterns depending on the food type, and thus building a fully-automated food journaling engine.

REFERENCES

[1] Kiyoharu Aizawa, Kazuki Maeda, Makoto Ogawa, Yohei Sato, Mayumi Kasamatsu, Kayo Waki, and Hidemi Takimoto. 2014. Comparative study of the routine daily usability of foodlog: A smartphone-based food recording tool assisted by image retrieval. *Journal of diabetes science and technology* 8, 2 (2014), 203–208.

[2] Kiyoharu Aizawa and Makoto Ogawa. 2015. Foodlog: Multimedia tool for healthcare applications. *IEEE MultiMedia* 22, 2 (2015), 4–8.

[3] Oliver Amft and Gerhard Troster. 2009. On-body sensing solutions for automatic dietary monitoring. *IEEE pervasive computing* 8, 2 (2009).

[4] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. 2015. Menu-match: Restaurant-specific food logging from images. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*. IEEE, 844–851.

[5] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101–mining discriminative components with random forests. In *European Conference on Computer Vision*. Springer, 446–461.

[6] Bertha S Burke et al. 1947. The dietary history as a tool in research. *Journal of the American Dietetic Association* 23 (1947), 1041–1046.

[7] Yin-Wen Chang, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin. 2010. Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research* 11, Apr (2010), 1471–1490.

[8] Jingyuan Cheng, Bo Zhou, Kai Kunze, Carl Christian Rheinländer, Sebastian Wille, Norbert Wehn, Jens Weppner, and Paul Lukowicz. 2013. Activity recognition and nutrition monitoring in every day situations with a textile capacitive neckband. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 155–158.

[9] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2017. Food recognition: a new dataset, experiments, and results. *IEEE journal of biomedical and health informatics* 21, 3 (2017), 588–598.

[10] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3207–3216.

[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] Mark Andrew Hall. 1999. Correlation-based feature selection for machine learning. (1999).

[13] Katerina Hnatkova, Donna Kowalski, James J Keirns, E Marcel van Gelderen, and Marek Malik. 2014. QTc changes after meal intake: Sex differences and correlates. *Journal of electrocardiology* 47, 6 (2014), 856–862.

[14] Shinji Hotta, Tatsuya Mori, Daisuke Uchida, Kazuho Maeda, Yoshinori Yaginuma, and Akihiro Inomata. 2017. Eating moment recognition using heart rate responses. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 69–72.

[15] Laleh Jalali, Da Huo, Hyungik Oh, Mengfan Tang, Siripen Pongpaichet, and Ramesh Jain. 2014. Personicle: personal chronicle of life events. In *Workshop on Personal Data Analytics in the Internet of Things (PDA@ IOT) at the 40th International Conference on Very Large Databases (VLDB), Hangzhou, China*.

[16] Laleh Jalali, Hyungik Oh, Ramin Moazeni, and Ramesh Jain. 2016. Human Behavior Analysis from Smartphone Data Streams. In *International Workshop on Human Behavior Understanding*. Springer, 68–85.

[17] Jonathan D Johnston. 2014. Physiological responses to food intake throughout the day. *Nutrition research reviews* 27, 1 (2014), 107–118.

[18] 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 (2008), 2010–2024.

[19] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. 2014. Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 1085–1088.

[20] Yoshiyuki Kawano and Keiji Yanai. 2015. Foodcam: A real-time food recognition system on a smartphone. *Multimedia Tools and Applications* 74, 14 (2015), 5263–5287.

[21] Nathaniel Kleitman. 1963. *Sleep and wakefulness*. University of Chicago Press.

[22] Fanyu Kong and Jindong Tan. 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing* 8, 1 (2012), 147–163.

[23] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and knowledge discovery* 15, 2 (2007), 107–144.

[24] Jean W Marr. 1971. Individual dietary surveys: purposes and methods. In *World review of nutrition and dietetics*. Vol. 13. Karger Publishers, 105–164.

[25] Yuji Matsuda, Hajime Hoashi, and Keiji Yanai. 2012. Recognition of multiple-food images by detecting candidate regions. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. IEEE, 25–30.

[26] Hyungik Oh and Ramesh Jain. 2017. From Multimedia Logs to Personal Chronicles. In *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 881–889.

[27] Hyungik Oh, Laleh Jalali, and Ramesh Jain. 2015. An intelligent notification system using context from real-time personal activity monitoring. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on*. IEEE, 1–6.

[28] Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi. 2014. Measuring calorie and nutrition from food image. *IEEE Transactions on Instrumentation and Measurement* 63, 8 (2014), 1947–1956.

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

[30] Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting About-to-Eat Moments for Just-in-Time Eating Intervention. In *Proceedings of the 6th International Conference on Digital Health Conference*. ACM, 141–150.

[31] Katherine A Sauder, Elyse R Johnston, Ann C Skulas-Ray, Tavis S Campbell, and Sheila G West. 2012. Effect of meal content on heart rate variability and cardiovascular reactivity to mental stress. *Psychophysiology* 49, 4 (2012), 470–477.

[32] 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.

[33] Eliot Stellar and E Eileen Shrager. 1985. Chews and swallows and the microstructure of eating. *The American journal of clinical nutrition* 42, 5 (1985), 973–982.

[34] 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, 1029–1040.

[35] Christopher C Tsai, Gunny Lee, Fred Raab, Gregory J Norman, Timothy Sohn, William G Griswold, and Kevin Patrick. 2007. Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. *Mobile networks and applications* 12, 2-3 (2007), 173–184.

[36] Gregorio Villalobos, Rana Almaghrabi, Parisa Pouladzadeh, and Shervin Shirmohammadi. 2012. An image procesing approach for calorie intake measurement. In *Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on*. IEEE, 1–5.

[37] Walter Willett. 2012. *Nutritional epidemiology*. Oxford University Press.

[38] Koji Yatani and Khai N Truong. 2012. BodyScope: a wearable acoustic sensor for activity recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 341–350.

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