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 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 to enable natural conversation when logging meals, which increases the ease of reactive self-reporting. Lastly, we automatically enhance the food journal by assessing ecological moments of eating activity through our personal chronicle engine. 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, smartphone and smartband.

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.

The foods and drinks we put in our body 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

HealthMedia'18, Oct 2018, Seoul, Korea

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

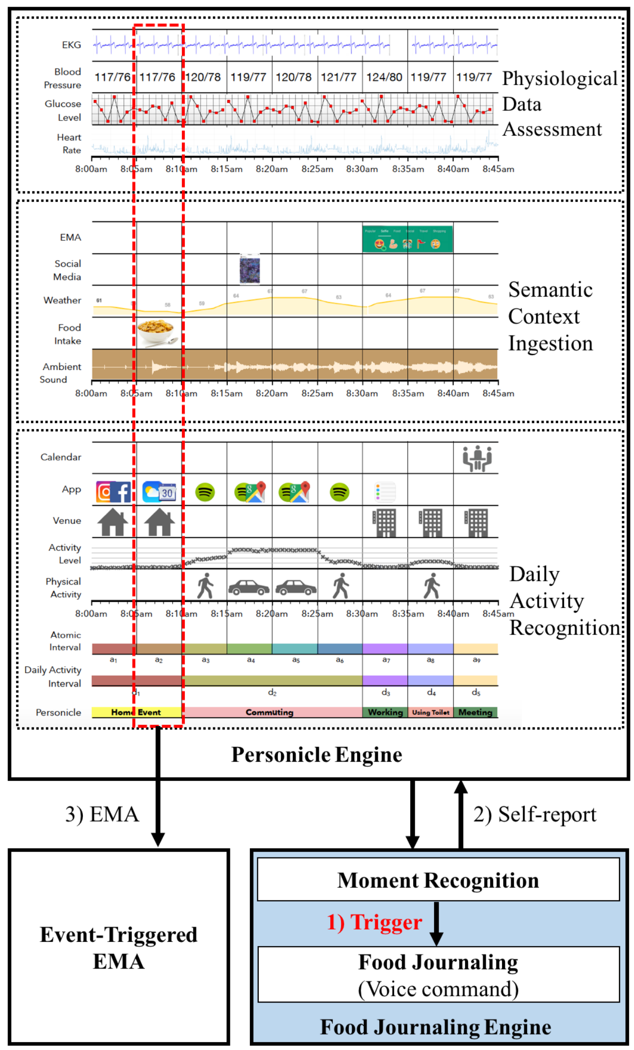

Figure 1: Assessing an enhanced ecological moment of eating activity through *Personicle* based food journaling.

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 plays a significant role in behavior change. Therefore, people who suffer from health related disorders as well as health care professionals 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 now, 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 afterwards, or even early abandonment. Over the years, many researchers in the field of multimedia have made attempts to automate food journaling by utilizing food image recognition in order to predict meals eaten and to estimate food calories. In particular, Aizawa et al, have prominently contributed to food journaling through their multimedia food-recording tool, FoodLog, making use of image processing for dietary assessment [1, 2, 20]. Such techniques have been effectively utilized in various commercial mobile applications, such as MyFitnessPal[2], Fitbit[3], or Bixby Vision [4]. Although remarkable technical progress has been made in the field, food journaling is still highly dependent on the user to take initiative, and 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 endemic problems of food journaling, 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 proactive and reactive perspective of the food journaling that can enhance the event-triggered EMA as shown 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 restaurant", and the other one is "eating at home". The former eating moments have been automatically recognized through smartphone based *Personicle*, which provides a person's time-ordered list of daily activities, in our previous research [15, 26, 27]. However, there have been difficult recognizing "eating at home" due to the lack of available smartphone sensors. In this paper, we try to solve the latter problems by pulling heart rate signal in the *Personicle*.

We then try to keep a food journal through what we call the event-triggered EMA, which is an automatically generated ecological momentary assessment. 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* engine 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, and thus 2) encouraging people to keep a well-balanced diet, as well as 3) helping them develop healthier dietary habits. We try to make it by providing a general eating moment model that can automatically recognize starting moment of eating, and a voice command food journaling method. We validate our approach with an experiment for 3 months with 3 users who are using *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 or bell so that it can let the user that it's the time to make a food journal.
(3) Then, the user just speaks whatever he is eating now, 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

[1]Cardiovascular disease
[2]https://www.myfitnesspal.com/
[3]www.fitbit.com/
[4]https://www.samsung.com/global/galaxy/apps/bixby/vision/

driven by advancements in technology. Commercial mobile applications such as MyFitnessPal[5] 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. 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, 18, 19, 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 [19]) 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, 17, 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, 16, 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 $\mathbf{R}$;
Set $\mathbf{S}_j = \mathbf{A}_i$ if $i = 0$ and $j = 0$, or $\mathbf{S}_j = \emptyset$, and then set $k = 0$, $m = 0$, $p = 0$;
Repeat
  Wait for next atomic-interval, $\mathbf{A}_i = \mathbf{A}_{i+1}$, $i = i + 1$;
  Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{A}_i$;
  Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{S}_j$;
  Calculate $\delta(i)$;
  If $\delta(i) = 1$, make a daily-activity-interval $\mathbf{R}_k$ by segmenting from $\mathbf{A}_j$ to $\mathbf{A}_i$;
  Set $p = i$, reset $i = j$, $\mathbf{a}_i = \mathbf{S}_j$, $\mathbf{s}_j = \mathbf{S}_j$;
  Repeat in the daily-activity-interval $\mathbf{R}_k$
    Assign next atomic-interval $\mathbf{a}_i = \mathbf{a}_{i+1}$;
    Extract activity level $\mathbf{l}_i$, and total amount of moving time $\mathbf{t}_i$ from $\mathbf{a}_i$;
    Extract activity level $\mathbf{l}_j$, and total amount of moving time $\mathbf{t}_j$ from $\mathbf{s}_j$;
    Calculate $\delta(i)$;
    If $\delta(i) = 1$, make a daily-activity-interval $\mathbf{r}_m$ by segmenting from $\mathbf{a}_j$ to $\mathbf{a}_i$;
    Set $m = m + 1$, $j = i$, set new seed atomic-interval $\mathbf{s}_j = \mathbf{a}_i$
  until i == p
  Set $k = K + 1$, $j = p$ set new seed atomic-interval $\mathbf{S}_j = \mathbf{A}_p$
until the system is terminated.
return R, r

---
[5]https://www.myfitnesspal.com/

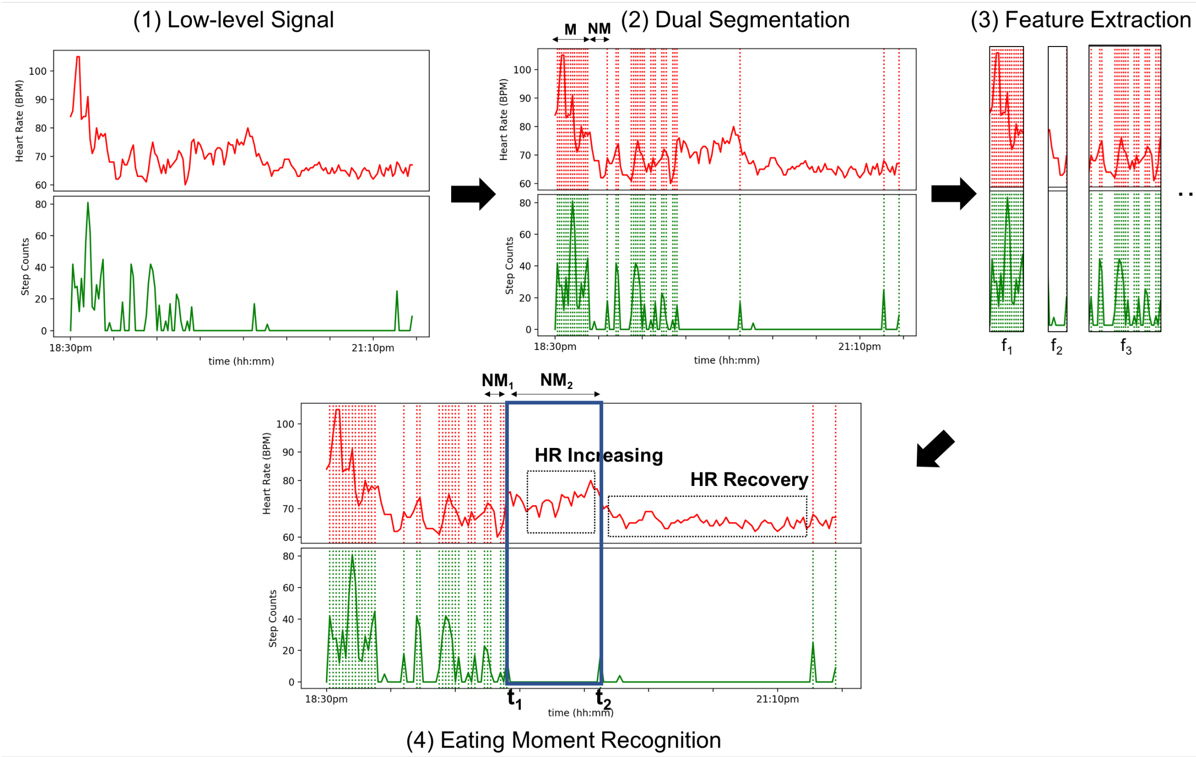

Figure 2: The sensor data processing pipeline for the eating moment recognition. M**: a moving type of daily-activity-interval,** NM**: a non-moving type of daily-activity-interval,** NM_1**: a sample of non-moving type of daily-activity-interval before eating,** NM_2**: a sample of non-moving type of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

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 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" or "commuting". 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. We 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

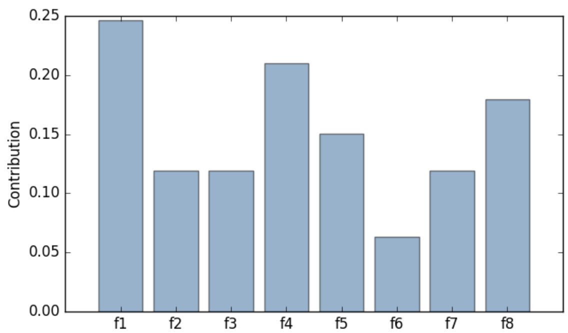

Figure 3: Contribution of each feature for eating moment classifier. $f1$: past average heart rate, $f2$: current average heart rate, $f3$: Δ average heart rate, $f4$: the amount of past NM time, $f5$: the amount of past M time, $f6$: the number of past steps, $f7$: the amount of moving time (past 30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

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 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 we hypothesize means that 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 of heart rate signal due to the fact that everyone has varying 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 based features, such as 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, there are day-night differences in body temperature and heart rate, and thus normal heart rate varies over a 24-hr period. To reflect these differences to the model, thus, we split the dataset depending on whether the current meal is breakfast, lunch, or dinner, and then tried to train each eating moment 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* engine finds the fine-grained daily-activity-intervals. If the interval is classified as a starting moment of "eating at home", we cut the whole range of this 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[6], 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[7]. 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,

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

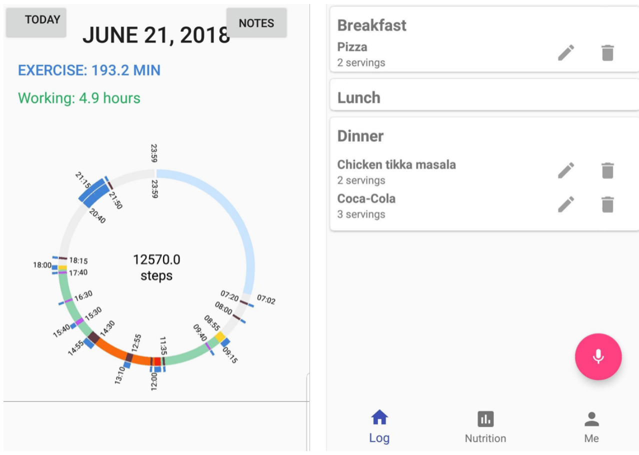

Figure 4: Screenshots of *Personicle* system including daily event recognition and food journaling.

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 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[8], 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* engine continuously

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

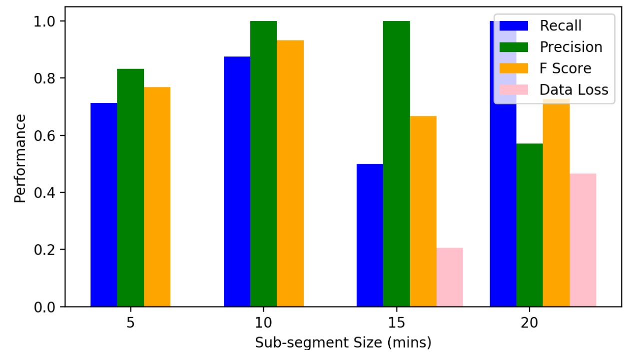

Figure 5: Performance (F-measure) of the eating moment recognition across different sub-segment sizes.

Table 1: Performance of eating moment classifiers trained with and without SAX algorithm in terms of Recall (R), Precision (P), and F-measure (F).

Model	Without SAX			With SAX		
	R	P	F	R	P	F
10-NN	0.4286	1	0.6	0.875	1	0.9333
NaiveBayes	0.8571	0.4615	0.6	0.7778	0.875	0.8235
Random Forest	0.5714	0.8	0.6667	0.75	1	0.8571
SVM	0.2857	0.6667	0.4	**0.875**	**1**	**0.9333**

monitors the chronicle of daily events as well as semantic contexts and physiological signals. Therefore, once the *Personicle* engine recognizes an eating moment, we can then 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 automate the food entry process, and thus keep a food journal without any user interventions. 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 describe experimental results regarding the accuracy of eating moment recognition. We implemented an android application as in Figure 4, and asked 3 participants for 3 months to label their eating moments. The total numbers of atomic-intervals, daily-activity-intervals, and eating labels were 90720, 1785, and 255 respectively. Next, we verified the voice command food journaling performance by evaluating the two kinds of protocols defined in Section 4.1. We tested 41 sample sentences with 3 distinct accents, native English, Indian, and Korean.

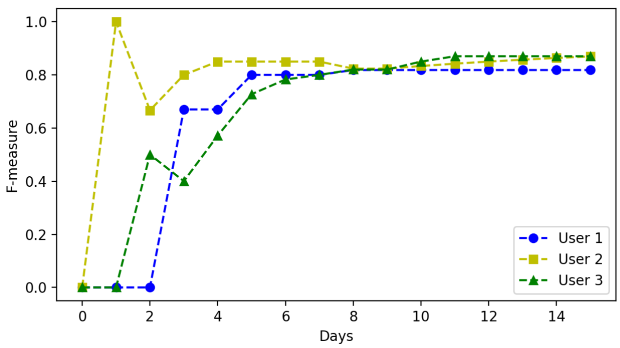

Figure 6: Performance of eating moment recognition using SVM classifier over time.

6.1 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.

We first tried to find the optimal sub-segment size. We found that every daily-activity-interval had a different set of segment sizes since they were split by the pattern of physical activity. Thus, it is essential to find an optimal sub-segment size to train and test the eating moment classifier. To do this, we experimented with four different sub-segment sizes. We chose intervals of 5, 10, 15, and 20 minutes, and trained the classifiers separately. Afterwards, we tested the classifiers to see which sub-segment size recognized the starting moment of eating activity the most accurately. From this experiment, we could clearly see that the performance of eating moment recognition is highly dependent on the sub-segment size. This indicates that if a sub-segment size is too long or too short, it is difficult to properly represent the features of eating moment. We also could see that there was data loss when the sub-segment sizes were too long, such as 15 or 20 minutes. This happened because there were some eating moments that do not last more than 10 minutes. Based on these findings, we found that 10 minutes was the most optimal sub-segment size that can reflect well the selected features.

With the 10 minute sub-segment size, we next verified the performance of our general eating moment classifier. Table 1 presents the precision, recall and F-measure of the eating moment classifiers with and without SAX algorithm. From this comparison, we could clearly see that the performance of all classifiers improved with the SAX algorithm. This can prove that the pattern of heart rate and activity around the eating moment are relatively similar from person to person. The results also could show that once the features are correctly extracted and selected, eating moments can be precisely classified by the basic supervised learning algorithms. The high precision and relatively low recall in the classifiers with SAX mean that more unique features are needed to classify more diverse eating moments. Among the results in Table 1, we finally chose the SVM classifier considering that it is fairly robust against overfitting, especially in high-dimensional space.

Table 2: 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

Lastly, we tested the performance of eating moment recognition using the SVM classifier over time. The SAX algorithm requires continuous data that follows a normal distribution so that it can correctly discretize the data into symbolic values. This means that it takes time to follow the normal distribution if there is not enough data. 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. Figure 6 indicates that there was a wide fluctuation in the recognition performance for first few days, but all the users showed an F-measure of about 80% within one week.

Given that there are constraints in our eating moment recognition, such as sub-segment size, we are unable to recognize some eating moments if those are less than 10 minutes or happen in moving type of daily-activity-interval. Nevertheless, our results showed the great potential to automate the eating moment recognition in a practical manner using smartphone and smartband.

6.2 Voice Command Food Journaling

In this section, we evaluated the performance of voice command food journaling by testing the 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, two kinds of protocols (Protocol 1: 16, Protocol 2: 25) were made. We evaluated these 41 test cases with three users who have different English accents so that we can also see the effect of accent on the voice command results. Table 2 shows all the incorrect results that we obtained from our experiment. All the meal types and quantities of the foods were correctly extracted from the test cases, and thus were excluded in the Table.

We found that there are three issues when trying to recognize food items. First, if the user is not familiar with the food item, such as Wonton, Idly, Szechuan, Lasagna, Quesadilla, or Sambhar, it is relatively difficult to be correctly extracted. Second, the voice recognition API has been mainly trained by a native English accent. For example, even though Bingsu and Idly were pronounced by Korean and Indian accent respectively, Google Voice API converted them into Kingsville and Elite. Third, if there is a short silence between words, recognition accuracy is compromised, such as 6 cake. However, the results were great if we describe common food items, such as "I had a big bucket of chicken nuggets for breakfast", "Hot dog" or "Dim sum". Even a very long sentence like "I had a

French toast which was dripping with maple syrup and chocolate sauce and whipped cream and strawberries" is correctly analyzed as "one French toast", "one maple syrup", "one chocolate sauce", "one whipped cream", and "strawberries" for "lunch". These results verify that the voice command via natural conversation can be successfully applied into food journaling.

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 may play a very important role in applications for health and well-being study. Results obtained from the three participants show the potential of such an approach for the unobtrusive food journaling. Further research would allow for automatically recognizing broad categories of food items by analyzing the variation of heart rate, 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, 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.

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

[17] 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.

[18] 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.

[19] 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.

[20] Keigo Kitamura, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2008. Food log by analyzing food images. In *Proceedings of the 16th ACM international conference on Multimedia.* ACM, 999–1000.

[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.