Bite detection and differentiation using templates of wrist motion

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Electrical Engineering

by
Soheila Eskandari
December 2013

Accepted by:
Dr. Adam Hoover, Committee Chair
Dr. John N. Gowdy
Dr. Eric R. Muth

Abstract

We introduce a new algorithm of bite detection during an eating activity based on template matching. The algorithm uses a template to model the motion of the wrist over a 6-second window centered on the time when a person takes a bite. We also determine if different types of bites (for example food vs. drink, or using different types of utensils) have different wrist motion templates. This method is implemented on 22,383 bites and 5 different types of templates are built. We then describe a method to recognize different types of bites using the set of templates. The obtained accuracy was 46%. Finally, we describe a method to detect bites using the set of templates and compare its accuracy to the original threshold-based algorithm. We get positive predictive value of 75 % and true positive rate of 47% found across all bites.

Acknowledgments

I would like to thank my advisor Dr. Hoover, for his guidance and advice throughout my Master's program. It has been my great pleasure to work closely with you. Sincere thanks to my committee members, Dr. Gowdy and Dr. Muth for their guidance and help. It is my honor to have you as members of my committee. I would like to thank my sister, Sonia, for her unconditional support, both financially and emotionally throughout my degree. Thanks to all my friends in my research group for guiding me in different situations. We gratefully acknowledge the support of the NIH via grant 1R41DK091141-A1 to collect the data used in this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis considers the problem of developing a template-based algorithm for tracking wrist motion to determine when a person has taken a bite of food. Previously, our group developed an algorithm for detecting bites that used thresholds on wrist roll motion and time [7]. That algorithm was shown to have a 76% true positive rate and 87% positive predictive value on a large data set of 276 subjects eating a total of 518 courses in a cafeteria setting [14]. In this thesis, we develop a template-based algorithm to determine if it can achieve a higher accuracy. The algorithm uses a template to model the motion of the wrist over a 6-second window centered on the time when a person takes a bite (places food into the mouth). We also determine if different types of bites (for example food vs. drink, or using different types of utensils) have different wrist motion templates. We first describe methods to construct templates of different types of bites from the cafeteria data set. We then describe a method to recognize different types of bites using the set of templates. Finally, we describe a method to detect bites using the set of templates and compare its accuracy to the original threshold-based algorithm.

1.1 Obesity Problem

Obesity is a serious and growing problem throughout the world and has recently been recognized as a disease by the U.S. American Medical Association. BMI (body mass index) is a measure commonly used to determine the condition. A BMI between 25 and 29.9 indicates overweight and a BMI greater than 30 indicates obesity [3]. According to the World Health Organization (WHO), 400 million adults were overweight and 1.6 billion were obese based on BMI values in 2006 [22]. According to the 2003-2004 National Health and Nutrition Examination Survey (NHANES), roughly one third of U.S. adults were overweight and another one third were obese [20]. The percentages increased to 35.5 percent for adult men and 35.8 percent for adult women in 2009-2010 [9]. Costs associated with obesity were estimated to be more than $117 billion in the U.S. in 2000 [16]. Obesity is strongly linked to a number of health problems such as heart disease, arthritis, hernia, sleep apnea and hypoxia [32].

Several obesity treatments are used in practice including very low-calorie diets, pharmacotherapy and surgery [10, 31]. The most important goal in all of these treatments is to balance energy consumption and expenditure. Currently, tools to monitor this balance are error-prone and cumbersome to use. The most common method for monitoring energy intake is to track all foods eaten and calculate calories manually. This method is well known to suffer from under-reporting and under-estimation [5]. Developing a simple and accurate method to calculate energy intake automatically during eating is the main goal of this research. With such a device, eating data can also be stored to examine habits over time to offer data for the study of long term healthy eating patterns.

1.2 Wrist motion tracking

Mobile health technology and body-worn sensors are being explored as new methods to monitor daily behaviors [17]. From the health perspective, they have been used for mobile monitoring of the human electrocardiogram (ECG) [12], measuring the heart rate, breathing frequency, blood pressure variations and breathing amplitude [11] and detection of different sleep phases [6]. Since a wrist-worn device allows for freedom of activities and has a greater likeness to a watch, it makes the wrist a natural place for implementing body-worn sensor devices in comparison to other areas of the body.

Eating involves the throat and mouth. Previous researchers have studied the monitoring of muscle activity at the throat and ear and multiple sensors tracking limb movements to measure eating-related behaviors [2, 24]. However, monitoring at the wrist has the potential to be less burdensome if eating activites can be reliably detected through the tracking of wrist motion.

Wrist motion has been tracked for a number of applications. One example is the "E-watch" developed to recognize location and detect temperature, act as a calendar, and communicate with a computer or cell phone [20,26]. The study of hand motions and gesture recognition are notable applications for using a device on the wrist instead of on the other parts of body. Sugimoto et al. [30] developed a wrist-mounted device which transmits wireless data with Bluetooth to evaluate the energy expenditure by assessing used oxygen. This is useful for heart rate measurement. A wrist-worn device which was presented by Harland [12] had two watch style sensors to obtain the human electrocardiogram (ECG) and display that on a computer wirelessly. It could be used for ECG mobile monitoring. Gagnadre et al. [11] described a device which could use for sleep phases detection. The device used several parameters like blood pressure alternation, breathing frequency and amplitude and heart rate by optic fiber sensor.

Hand motions are also studied by wrist motion tracking for different problems. Howard et al. [13] presented a wrist worn device which is the virtual typing and pointing system. Ogris et al. [21] tracked the gesture of a pre-defined bicycle repair task with determining distance and motion by 3 sensors; accelerometers and gyroscopes, ultrasonic. Schmidt et al. [25] developed a device named eWatch for the virtual orchestra systems. It could be used for gesture recognition based on acceleration data and light. Lementec et al. [18] could recognize different motion gestures with compound of sensor and position states. The sensors were mounted on right and left wrists, upper part of the right and left arms.

Amft and colleagues [1,15] used a wrist-mounted sensor to detect eating action. In addition they implemented sensors on upper arms, head and ears. They looked for pre-defined sensors signals patterns in order to classify them as drinking or utensiling patterns to eat. However we are looking for a device which is simple and has less sensors for bite detection regardless what type of food is eaten by the wearer.

1.3 Previous related work done by our group

Our group previously developed a method for detecting bites by tracking wrist roll motion [7]. The algorithm was based on the tracking of roll motion only, which was found to have a pattern indicative of the taking of a bite of food that was independent of the actual wrist orientation. The motion pattern for eating was defined based on four events. In two events, velocity should surpass a positive and negative threshold. The other two events are related to the minimum amounts of time between the two rolls of one bite, and between consecutive bites. The algorithm for detecting the bite using the events is shown as follows:

```
Let EVENT = 0
Loop
  Let V_t = measured roll vel. at time t
  if V_t > T1 and EVENT = 0
    EVENT = 1
    Let s=t
  if V_t<T2 and T-s>T3  and EVENT = 1
    Bite detected
    Let s=t
    EVENT = 2
  if EVENT = 2 and  T-s>T4
    EVENT = 0
```

The variable named EVENT should iterate through all the events of roll motion. T1 and T2 are the roll velocities and T3 is the time interval between the first and second events of roll motion. T4 is the time interval between the end of one bite and beginning of the next bite.

The method has been tested in a series of experiments that successively relaxed control conditions and therefore provided increased challenges. In the first test, 139 meals were eaten by 51 subjects (each subject participated 2-3 times). The food in every meal was the same (waffles), the container was the same (plate), and the utensil was the same (fork). The method was found to have a 94% true detection rate and 80% positive predictive value. The second test was implemented in less controlled conditions. Subjects consumed their own foods and drinks however they liked. This test was implemented across 49 meals eaten by 47 subjects (two subjects participated twice). This test yielded an 86% true detection rate and an 81% positive predictive value. The first two tests were conducted in a laboratory setting with each person eating by themselves. The third test was executed in a cafeteria setting with participants seated up to 4 at a time at a common table. A total of 276 subjects ate a meal. In total, the data set consists of 22,383 bites of 380 different foods and drinks. The results showed a true detection rate of 76% with a positive predictive value of 87%. It was found that bites were taken somewhat quicker in the third experiment as compared to the first two experiments, probably due to the environment or social setting of dining with up to 3 strangers. The timing thresholds in the original algorithm were tuned to provide maximum results based on the data in the first 2 tests. Adjusting the second timing threshold (T4, the time between the end of a bite and the beginning of the next bite) down a value of 6 seconds (from its original 8 seconds) yielded a true detection rate of 82% and a positive predictive value of 82% [14].

All of these results are encouraging. They have shown that it is possible to track wrist motion across a wide variety of people and foods and reliably detect bites. The purpose of this thesis was to explore if an alternative algorithm based upon templates could provide a better detection rate.

1.4 Template matching

There are several methods to determine similarity between templates and an unknown signal, including cross correlation and template matching [4, 19]. Template matching is a technique which locates the position of a desired signal inside a large signal. The desired signal is named a template. Template matching is the process of moving the template over the main signal and calculating the resemblance between the template and the window in the large signal where the template is positioned. The position in the main signal where the most similarity is obtained. Areas of major likeness can be determined via the sum of the cross correlation coefficient and the value of absolute difference. The cross-correlation is defined as:

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f^*[m]g[n+m] \tag{1.1}$$

The absolute difference is defined as:

$$SAD(x,y) = \sum_{m=-\infty}^{\infty} |f[m] - g[m]| \tag{1.2}$$

In this method the lowest SAD value gives the estimate for the best position of template within the main signal. The method is simple to understand and implement.

1.5 Novelty

In this thesis we describe a new method for bite detection using templates of wrist motion. We created different templates based on different types of bites. We extracted the data from a total of 276 people were recorded consuming 518 courses which includes 22,383 bites. Templates were built on the utensils which were used for taking a bite. Using the templates, we performed the template matching algorithm against raw motion data to recognize the unidentified bite. For purposes of detection, template matching will be explored to determine if bites could be detected among all activities during eating and compared its accuracy to the original threshold-based algorithm.

Chapter 2

Research Design and Methods

2.1 Overview

The data used for this thesis was collected in a cafeteria setting. It has been used previously for a study evaluating the accuracy of the original threshold-based bite detection algorithm [14] and for another study evaluating the accuracy of bite-derived estimates of calories consumed by people [23]. This chapter first provides a brief description of that data; further details can be found in the above cited references. We then describe our methods for building bite templates, differentiating bite types through template matching, and detecting bites via template matching.

2.2 Data Collection

The data used for this work was collected in the Harcombe Dinning Hall at Clemson University. The cafeteria provides a large variety of foods and beverages. Typcial foods include soups, vegetables, a salad bar, pasta, rice, chicken, fish, hamburgers, French fries, cakes, fruits and other deserts. Beverages include milk, sodas, tea, coffee, juices and water. Multiple kinds of containers are used including plates, bowls, mugs and glasses. Utensil types include knives, forks, spoons, chopsticks, fingers and straws. The dining hall can seat up to 800 people simultaneously. Collectively these characteristics suggest that a wide variety of eating behaviors can be observed in this setting. For our experiments, 276 subjects participated in data collection including 145 females and 131 males between 18-75 years old.

To record the data, a table with four seats was instrumented and prepared for participants. Figure 2.1 shows an image of the instrumented table. A custom wrist motion tracker was developed by our group and connected to a nearby laptop to record the wrist motion data during meal consumption. Participants were asked to wear the device on the hand which they normally use to eat with. The device contained three sensors, one accelerometer and two gyroscopes, in a small circuit board. The sensors were a STMicroelectronics LIS344ALH [27] three-axis linear accelerometer, a LPR410AL [28] two-axis gyroscope for roll and pith detection, and a LPY410AL [29] two-axis gyroscope for pitch and yaw detection. The sampling frequency was set to 15 Hz. A Gaussian-weighted window was used for smoothing accelerometer and gyroscope data [7, 14].

Two other instruments were used for data collection. A digital scale was set under the tray for each participant to continuously measure tray weight. This data was not used for the experiments described in this thesis. Four digital cameras were set in the ceiling to record the visual actions of each participant. The video was zoomed to see the head and torso of the participant along with the food tray in front of the participant. The data of all three instruments were synchronized using time-stamps so that they can be played back for review and for processing.

Figure 2.1: Instrumented table.

2.3 Ground truth

The data recorded by all the instruments was manually reviewed to determine the times of bites taken by participants. Each recording was observed and the following information was recorded for each bite: time (data index), food, utensil, container, and hand used (left, right or both). The times of actual bites were determined by looking at the hands and mouth of the subject in the video. The actual bites are used as ground truth to determine the accuracy of our method for bite differentiation and for bite detection. A total of 22,383 bites comprise the ground truth.

2.4 Bite Templates

The original threshold-based algorithm was developed using casual observations of approximately 1,800 bites over 49 meals [7]. Given the wider variety of bites

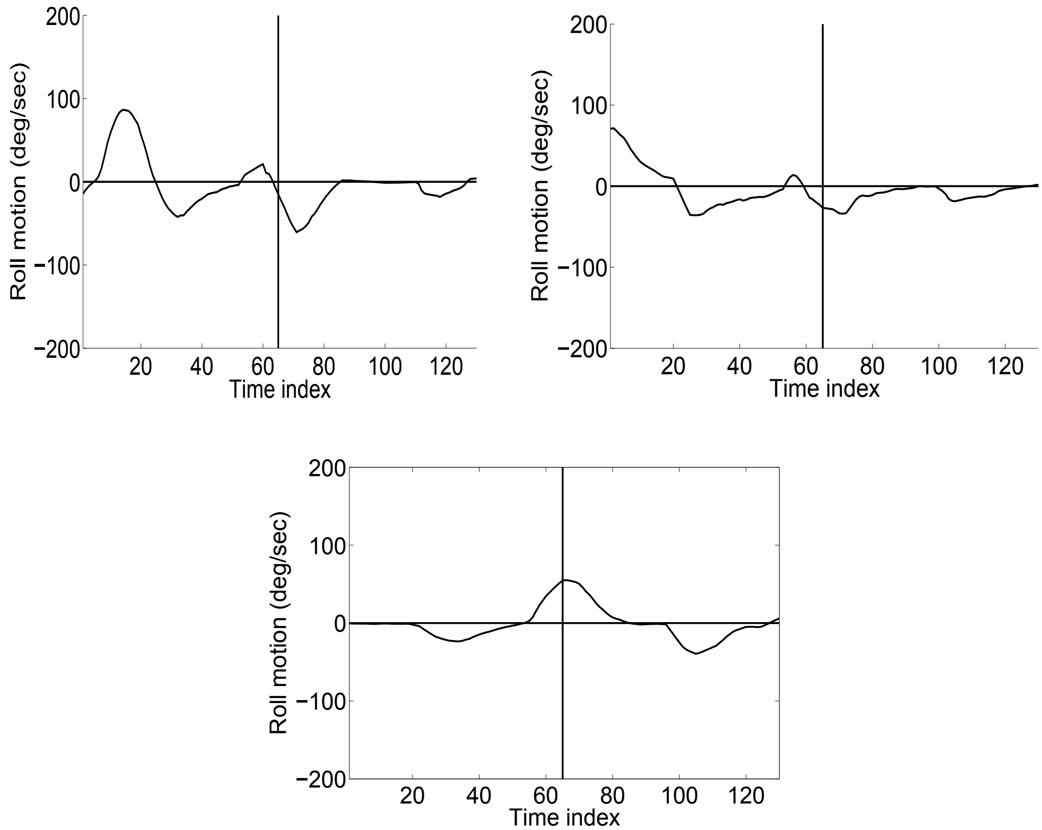


Figure 2.2: Bite examples.

available in the cafeteria data set, along with known food and utensil characteristics, it is possible to study the average motion sequences of different types of bites. For example, three different bites taken are shown in Figure 2.2, one each by a fork, spoon and hand.

The first experiment seeks to determine the overall pattern and variability pattern of wrist motion of a bite. A template is created by averaging the motion data across all the bites in the 22,383 total ground truth bites that match the desired characteristics. A template can be created for any set of characteristics, for example all drink bites, or all bites eaten with a fork, or all bites of a specific food, or all bites. Each template is created over a six second window centered on the bite time in order to capture pre- and post-bite motion. The template includes data collected from both the accelerometers and gyroscopes, including information about x-, y-, and z-axis acceleration as well as yaw, pitch, and roll.

We created a set of templates (also known as a filter bank) of motion patterns of different types of bites. Specifically, we studied a total of five different types of bites: drink bites, bites taken with a fork, bites taken with a spoon, food bites eaten using one hand (the instrumented hand), and food bites eaten using both hands (e.g. a sandwich picked up with both hands simultaneously).

2.5 Bite differentiation

The second experiment sought to determine if different types of bites could be reliably recognized using template matching against the typical (average) motion pattern. Figure 2.3 illustrates the process, where a bite of unknown type is being compared against templates for a hand bite, drink bite and fork bite.

In order to identify different bite types, each actual bite in the ground truth was matched against the five created average templates using the sum of absolute difference, as follows:

Let: $C_i = 1, ..., C$ courses

$\forall$ C_i let $b_j = 1, ..., b_c$ bites for course C

Let: $T_k = 1, ..., T_N$ different templates

$\forall$ bites:

$$\text{Max similarity } = \min[\sum_{b_j-3sec}^{b_j+3sec} \{|b_j - T_{ij}|\}]_{k=1}^{k=N}$$

The minimum scoring template identifies the most closely matching bite. The algorithm was run across all 22,383 manually labeled bites and a confusion matrix was generated to evaluate the performance of the algorithm.

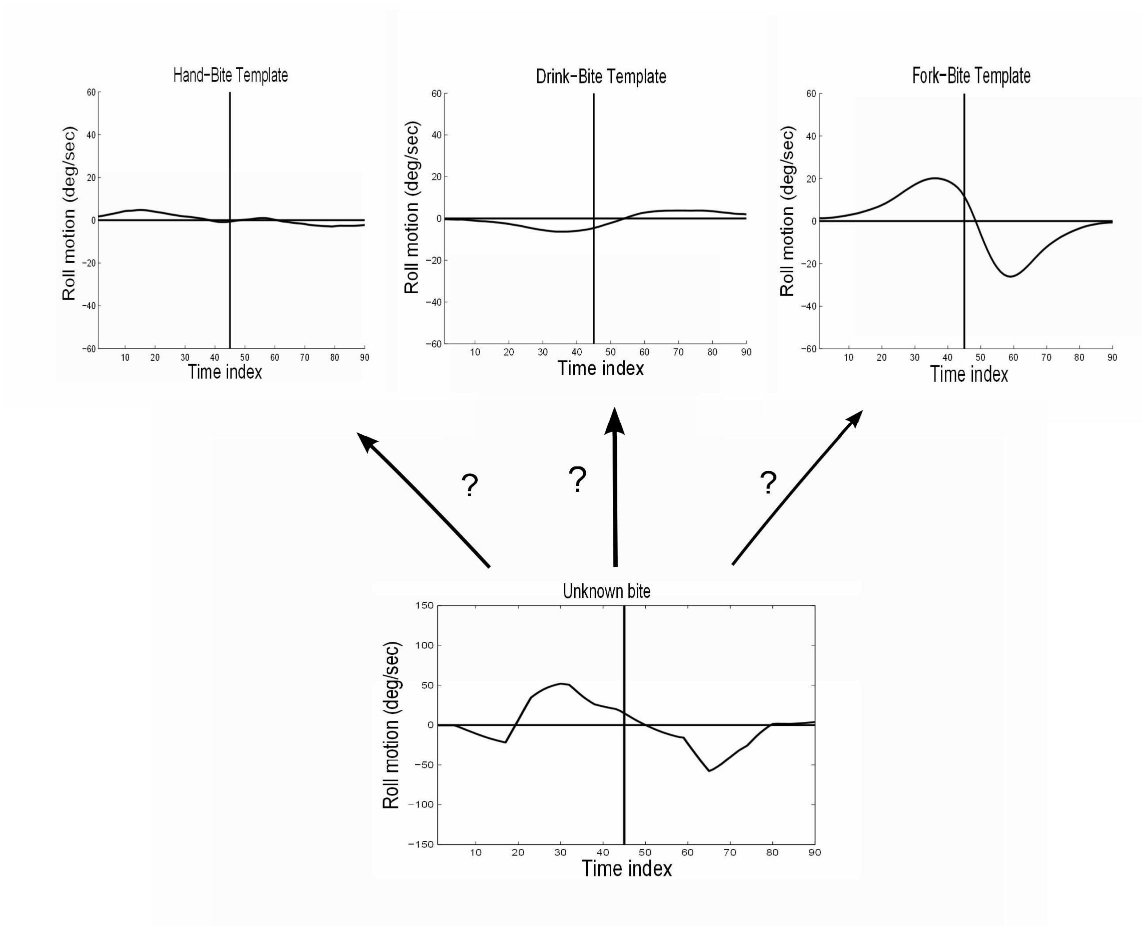


Figure 2.3: Bite recognition using templates.

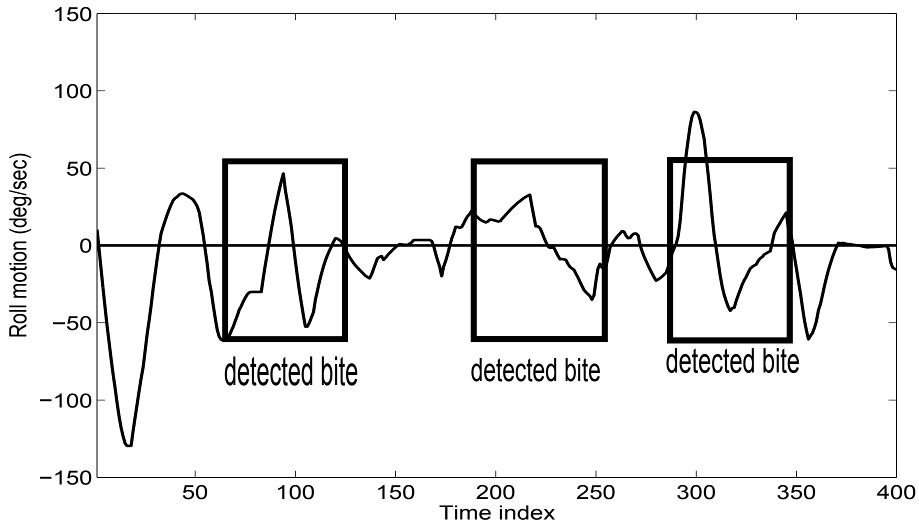


Figure 2.4: Bite detection during a meal.

2.6 Bite detection

The third experiment sought to determine if bites could be reliably detected using template matching. Figure 2.4 illustrates the idea. Eating activities during a meal include several different actions such as using a utensil, taking a bite and drinking. The purpose of this experiment is to determine if the characteristic motion of a bite is regular enough that a rigid template can reliably detect it and differentiate it from other activities.

Bites were detected by calculating the sum of absolute difference between a bite template and the wrist motion data at every time step, and then by detecting local minima in the sum of absolute difference. The minima indicate where the template best matched the motion data. Local minima were found using the peak detection algorithm described in [8]. The algorithm was inverted to find local minima instead of local maxima. Each local minima was taken to be the time of a computer detected bite. While the bite differentiation experiment used all 6 axes of motion, this experiment used only roll motion in order to simplify the calculations.

Figure 2.5: The developed software for bite detection

For evaluation, computer detected bites were compared with ground truth bites using the method described in [7]. The evaluation of each computer detected bite was done by finding correspondences within a window of time from the previous computer detected bite to the next computer detected bite. Figure 2.5 shows a screen shot of the software we developed for this. The red lines show the ground truth bites and the black dots the computer detected bites using our method. The first ground truth bite within this window, that has not previously been associated with another computer detected bite, is considered as a true detection. If there are no ground truth bites within that window, then the computer detected bite is considered a false positive. After all computer detected bites have been iterated through this process, any remaining ground truth bites that were not corresponded with computer detected bites are considered undetected bites. The overall detection rate is then calculated as:

$$detection\,rate = \frac{true\,detected\,bites}{true\,detected\,bites + undetected\,bites} \tag{2.1}$$

The positive predictive value (PPV) is a measure of the number of false positives which can be calculated as:

$$PPV\ = \frac{true\,deteceted\,bites}{true\,deteceted\,bites + false\,detected\,bites} \tag{2.2}$$

Chapter 3

Results

3.1 Bite Templates

We first created the total bites templates of all motion data. Figure 3.1 shows the templates for all motion axes of the total bites which were 22,383 bites. Then we separated the bites according to food and drink bites. Using the ground truth there were 17,166 bites that could be identified as food and 3,185 bites identified as drink. Figure 3.2 shows the average accelerometer and gyroscope motions for food and drink bites. The solid lines show the average motion and the bars show the standard deviations at each point in time. Food bites show a larger average motion in the Z and roll axes, while drink bites show a larger average motion in the X and yaw axes. Drink bites also show a longer (slower) motion than food bites in the yaw axis. It can also be seen that the roll motion for drink bites is opposite to that of food bites, with negative roll preceding positive roll.

We then calculated additional average motions for food bites based on the different types of utensils used to take the bite: bites taken with a fork, bites taken with a spoon, food bites eaten using one hand (the instrumented hand), and food bites

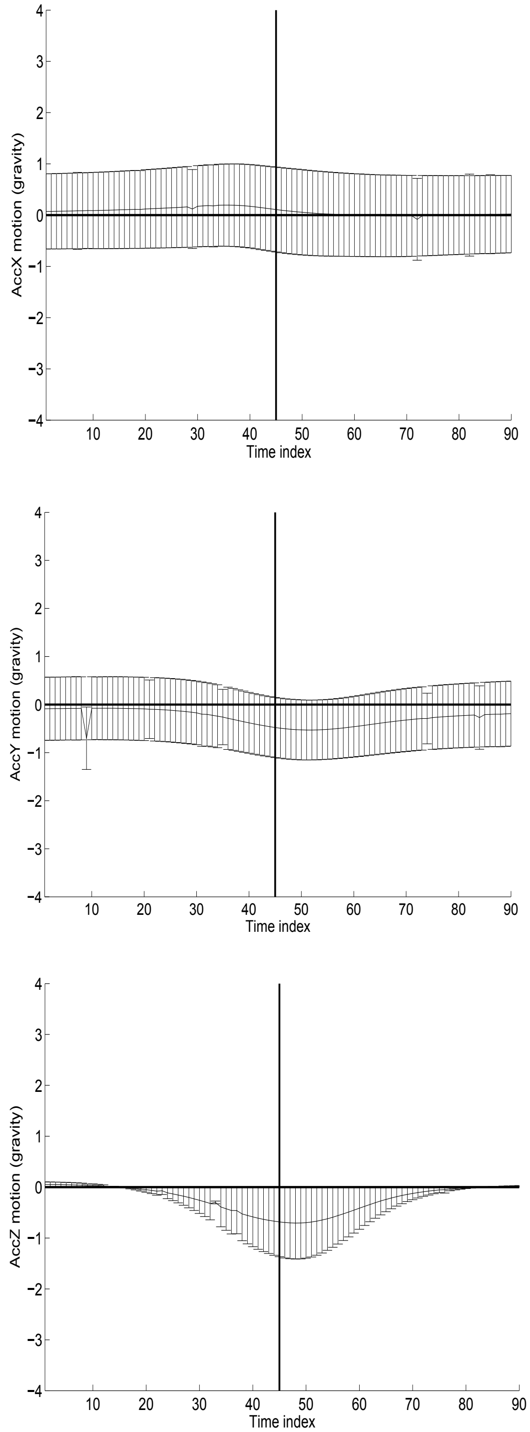
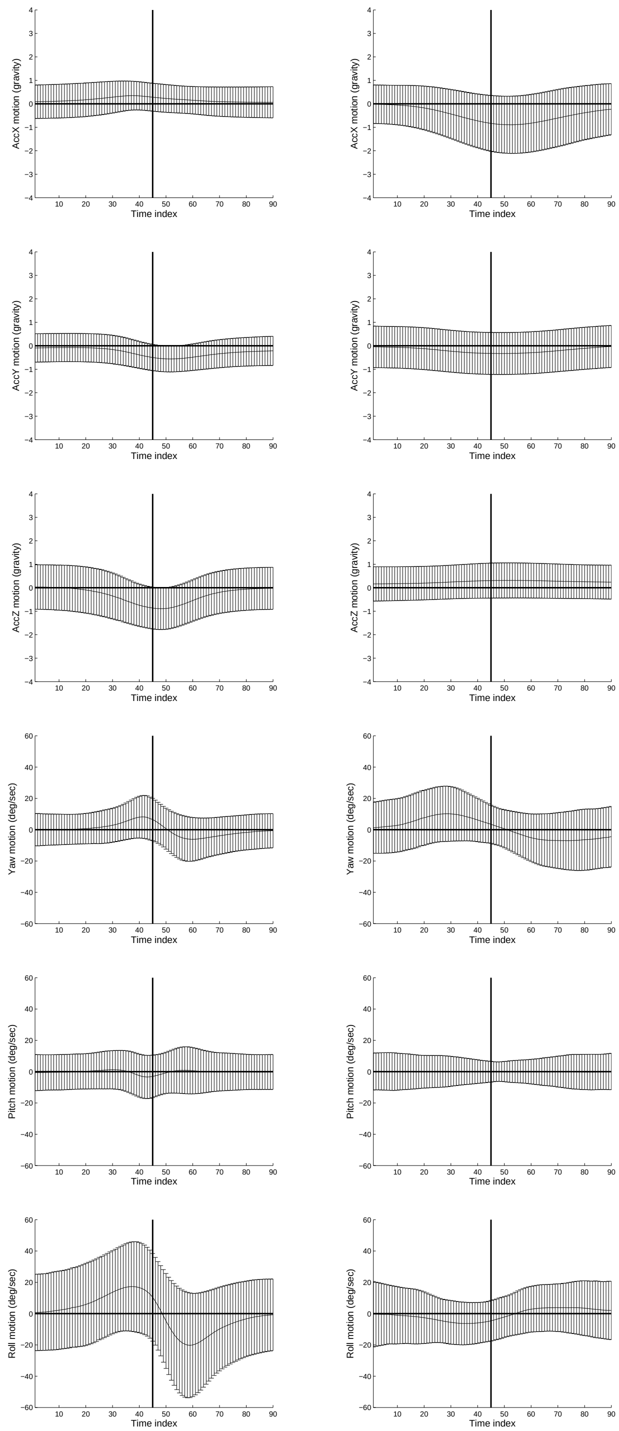
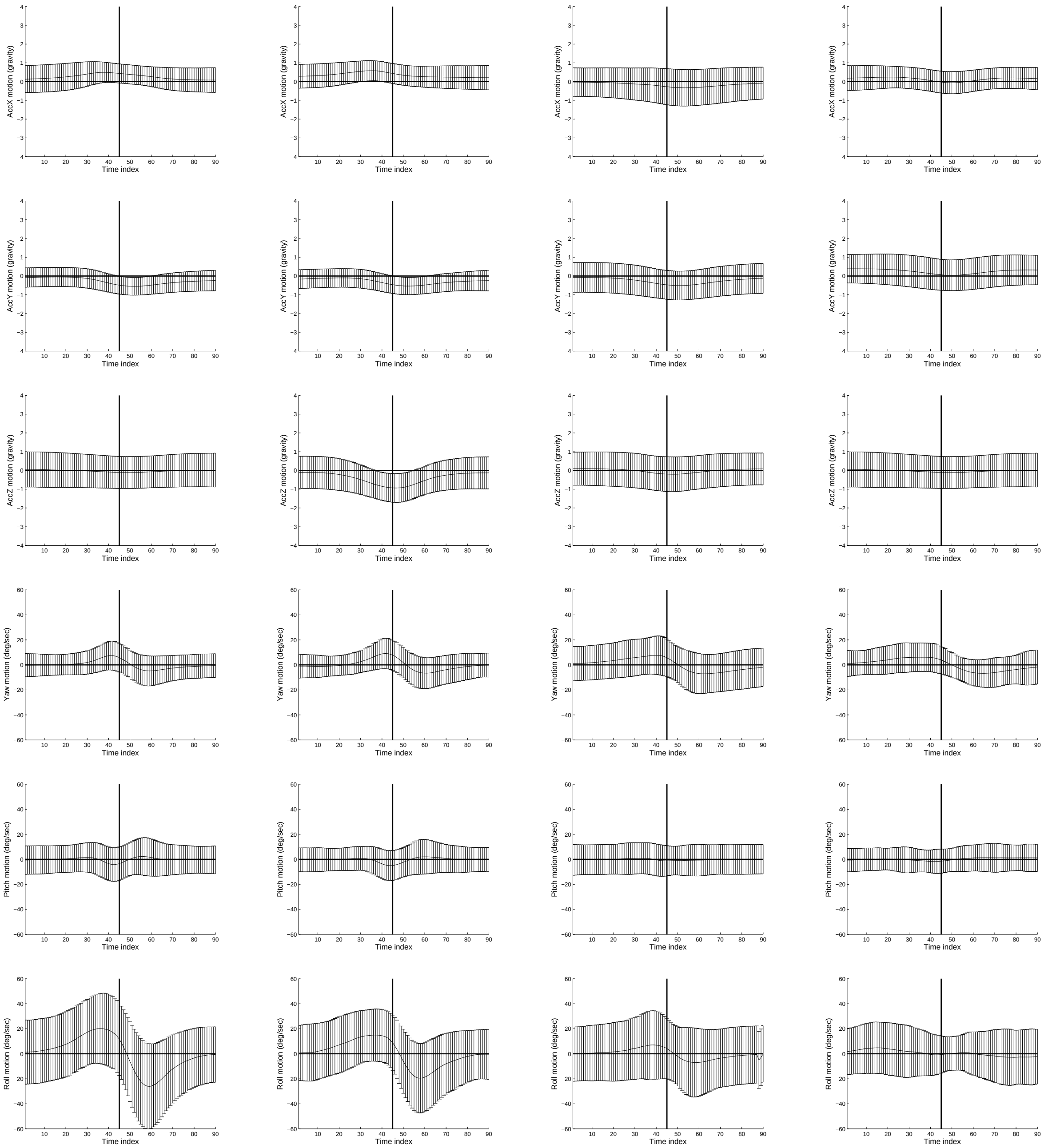


Figure 3.1: Template for all bites.



(a) food bites (b) drink bites

Figure 3.2: Templates for food vs drink bites. Top to bottom is acceleration x, y, z and gyroscope yaw, pitch, roll.



(a) fork bites (b) spoon bites (c) single hand bites (d) both hand bites

Figure 3.3: Templates for different utensils for food bites. Top to bottom is acceleration x, y, z and gyroscope yaw, pitch, roll.

eaten using both hands (e.g. a sandwich picked up with both hands simultaneously). Drink bites were retained as a fifth category. A total of 8,764 bites were taken by fork and 1,986 bites were taken by spoon. Subjects had 9,241 bites with a single hand and 2,441 bites with both hands. Figures 3.3 shows the average and variations in motion for each acceleration and each gyroscope axis for each of these four different types of food bites. In the roll axis, fork and spoon bites have noticeably more motion than other types of bites. For foods eaten using both hands, the most prominent motion is in the yaw axis, similar to drink bites. The largest variability in motion appears in the roll axis for all bite types.

3.2 Bite differentiation

The average motions calculated in the previous result were used as templates to study bite differentiation. Each average motion is considered a template. The set of templates is considered a filter bank. After creating the filter banks, we implemented template matching based on the sum of absolute difference to see whether it could be used to recognize unknown bites. We used a confusion matrix to visualize the performance of each template.

Table 3.1 shows the results for differentiating food and drink bites across each of the 6 axes of motion. Different axes contributed different amounts to separating the two types of bites. Table 3.2 shows the result combining all 6 axes. Drink bites were recognized with 95% accuracy but 30% of food bites were confused as drink bites.

Tables 3.3-3.8 show the confusion matrices for the five types of bites according to utensil, for each axis. Table 3.9 shows the confusion matrix after combining the yaw and roll axes. Table 3.10 shows the confusion matrix after combining the sum of

		Computer detected	
Ground truth		Food(Ax,Ay,Az,Yaw,Pitch,Roll)	Drink(Ax,Ay,Az,Yaw,Pitch,Roll)
	Food	75%,72%,68%,72%,43%,64%	25%,28%,32%,27%,57%,36%
	Drink	13%,10%,12%,40%,19%,5.6%	87%,90%,88%,60%,81%,94%
Accuracy	81%,81%,78%,66%,62%,79%		

Table 3.1: Accelerometer and gyroscope motions confusion table for food & drink bites recognition.

		Computer detected	
Ground truth		Food	Drink
	Food	12045(70%)	5248(30%)
	Drink	114(5%)	2056(95%)
Accuracy	83%		

Table 3.2: Bite differentiation of food and drink bites using all 6 motion axes.

absolute difference result from all 6 axes. Overall, the accuracy for recognizing these 5 different types of bites was low ranging from 19-48% for the 4 different types of utensils used for food to 80% for drink bites.

3.3 Bite detection

For purposes of detecting bites, template matching was executed using a single template on the roll axis that was the average of all 22,383 bites. This yielded an overall detection rate of 48% and a positive predictive value of 75%. Different axes and different combinations of axes were examined but none yielded a higher performance.

		Computer detected				
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	1991(23%)	4174(49%)	294(4%)	1112(13%)	1009(12%)
	Spoon	372(19%)	1113(20%)	86(4.3%)	288(14%)	156(8%)
	Drink	20(1%)	113(5%)	1744(81%)	129(6%)	149(7%)
	Both hand	112(8%)	411(30%)	242(18%)	384(28%)	230(17%)
	Single hand	685(15%)	949(21%)	931(20%)	864(19%)	1114(25%)
Accuracy	35%					

Table 3.3: AccX motion confusion table.

		Computer detected				
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	1751(20%)	4791(56%)	388(5%)	762(9%)	888(10%)
	Spoon	288(14%)	1288(64%)	92(5%)	171(8.5%)	176(9%)
	Drink	23(1%)	95(4.5%)	1811(84%)	124(6%)	102(5%)
	Both hand	84(6%)	492(36%)	381(28%)	254(18%)	166(12%)
	Single hand	491(11%)	1236(27%)	1275(28%)	673(15%)	868(19%)
Accuracy	41%					

Table 3.4: AccY motion confusion table.

		Computer detected				
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	1785(21%)	4313(51%)	514(6%)	797(10%)	1171(1.4%)
	Spoon	357(18%)	1206(60%)	101(5%)	176(9%)	175(9%)
	Drink	20(1%)	127(6%)	1776(82%)	139(6.5%)	93(5%)
	Both hand	89(7%)	428(31%)	501(37%)	227(16%)	134(10%)
	Single hand	452(10%)	1267(28%)	1588(35%)	485(11%)	751(17%)
Accuracy	39%					

Table 3.5: AccZ motion confusion table.

		Computer detected				
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	3632(42%)	2673(31%)	1188(14%)	490(6%)	597(7%)
	Spoon	791(40%)	754(38%)	228(11%)	94(5%)	148(7%)
	Drink	603(28%)	205(10%)	1105(51%)	208(10%)	34(1.5%)
	Both hand	562(41%)	233(17%)	382(28%)	120(9%)	82(6%)
	Single hand	1615(36%)	1308(29%)	896(20%)	304(7%)	420(10%)
Accuracy	30%					

Table 3.6: Yaw motion confusion table.

	Computer detected					
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	1244(14%)	284(33%)	2173(25%)	1122(13%)	1194(14%)
	Spoon	177(9%)	797(40%)	509(25%)	220(11%)	312(15%)
	Drink	68(3%)	218(10%)	1043(48%)	573(27%)	253(12%)
	Both hand	82(6%)	265(19%)	432(31%)	427(31%)	173(13%)
	Single hand	364(8%)	1210(27%)	1390(31%)	706(16%)	873(19%)
Accuracy	30%					

Table 3.7: Pitch motion confusion table.

	Computer detected					
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	4218(49 %)	1470(17 %)	687(8%)	1000(12%)	1205(14 %)
	Spoon	748(37 %)	399(20%)	263(13 %)	282(14%)	323(16%)
	Drink	30(1.5%)	35(1.6%)	1528(71%)	385(18%)	177(8%)
	Both hand	50(4%)	78(6%)	555(40%)	442(32%)	254(19%)
	Single hand	1584(35%)	531(12%)	805(18%)	829(18%)	794(18%)
Accuracy	38%					

Table 3.8: Roll motion confusion table.

	Computer detected					
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	4286(50%)	1734(20%)	570(6.7%)	1025(12%)	965(11%)
	Spoon	690(34%)	553(27%)	217(11%)	301(15%)	254(13%)
	Drink	34(1.6%)	51(2.4%)	1489(69%)	437(20%)	144(6.7%)
	Both hand	63(4.6%)	93(6.8%)	480(35%)	507(37%)	236(17%)
	Single hand	1556(34%)	667(15%)	640(14%)	965(21%)	715(16%)
Accuracy	40%					

Table 3.9: Confusion combining roll and yaw motions.

	Computer detected					
		Fork	Spoon	Drink	Both Hand	Single Hand
Ground truth	Fork	4123(48%)	2337(27%)	280(3%)	1329(15%)	511(6%)
	Spoon	596(30%)	774(38%)	113(6%)	423(21%)	109(5%)
	Drink	32(1.5%)	32(1.5%)	1708(80%)	218(10%)	165(8%)
	Both hand	67(5%)	74(5%)	384(28%)	635(46%)	219(16%)
	Single hand	1490(33%)	579(13%)	642(14%)	986(22%)	846(19%)
Accuracy	46%					

Table 3.10: Confusion combining all 6 motion axes.

Chapter 4

Conclusions and future Work

In this thesis, we have built a filter bank of different bite motions including templates of bites taken by fork, spoon, hand, both hands and drink. The templates were created by averaging motion data collected for each of the five types of bites across 22,383 recorded and labeled bites. The templates were created over six second windows centered on the point in time at which the bite was recorded.

Food and drink bites appear to have different wrist motion patterns and different types of utensils for food bites also appear to have different wrist motion patterns. The bite templates were matched against the raw motion data using the mean absolute difference to determine the similarity between each of the filter bank templates and the unidentified bite in question. The algorithm was run across 22,383 manually labeled bites, allowing the creation of a confusion matrix to examine the performance of the algorithm. A 46% accuracy for single bite recognition was found using all 6 axes of motion. Drink bites could be fairly reliably recognized compared to food bites, but the recognition accuracy for type of utensil was low and it shows they are not consistent enough to enable differentiation via template matching.

For purposes of detection, template matching was explored to determine if bites could be detected among all activities during eating. The overall true detection rate found was 48% with a positive predictive value of 75% while the true positive rate for threshold based algorithm was 77% and the PPV was 86%. It shows template matching is too rigid for detecting bites because there is too much variability in appearance but interestingly, it yielded the close PPV in the threshold-based algorithm suggesting it might be useful for suppressing false positives.

Bibliography

[1] O. Amft, H. Junker, and G. Troster, “Detection of eating and drinking arm gestures using inertial body-worn sensors,” in *Proceedings of the ninth IEEE International Symposium on Wearable Computers*, 2005, pp. 160–163.

[2] O. Amft and G. Troster, “On-body sensing solutions for automatic dietary monitoring,” *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 62–70, 2009.

[3] G. Billington, Epstein, “Overweight, obesity, and health risk.” *Arch Intern Med*, vol. 160, pp. 898–904, 2000.

[4] M. Boninsegna and M. Rossi, “Similarity measures in computer vision,” *Pattern Recognition Letters*, vol. 15, no. 12, pp. 1255–1260, 1994.

[5] C. Champagne, G. Bray, A. Kurtz, J. Monteiro, E. Tucker, J. Volaufova, and J. Delany, “Energy intake and energy expenditure: a controlled study comparing dietitians and non-dietitians,” *Journal of the American Dietetic Association*, vol. 102, no. 10, pp. 1428–1432, 2002.

[6] C. Ching, M. Jenu, and M. Husain, “Fitness monitor system,” in *Proceeding of Conference on Convergent Technologies for Asia-Pacific Region*, vol. 4, 2003, pp. 1399–1403.

[7] Y. Dong, “Tracking wrist motion to detect and measure the eating intake of free-living humans,” Ph.D. dissertation, Electrical and Computer Engineering Department, Clemson University, 2012.

[8] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, “Detecting periods of eating during free living by tracking wrist motion,” *IEEE of Biomedical Health Informatics*, 2013.

[9] K. Flegal, M. Carroll, B. Kit, and C. Ogden, “Prevalence of obesity and trends in the distribution of body mass index among us adults, 1999-2010,” *The journal of the American Medical Association*, vol. 307, no. 5, pp. 491–497, 2012.

[10] J. Foreyt and W. Poston II, *Overview and the future of obesity treatment*. Springer, 1999.

[11] C. Gagnadre, M. Billon, and S. Thuillier, "Fibre optic sensor for physiological parameters," *Electronics Letters*, vol. 34, no. 21, pp. 1991–1993, 1998.

[12] C. Harland, T. Clark, and R. Prance, "High resolution ambulatory electrocardiographic monitoring using wrist-mounted electric potential sensors," *Measurement Science and Technology*, vol. 14, no. 7, pp. 923–928, 2003.

[13] B. Howard and S. Howard, "Lightglove: Wrist-worn virtual typing and pointing," in *Proceedings of IEEE Fifth International Symposium on Wearable Computers*, 2001, pp. 172–173.

[14] Z. Huang, "An assessment of the accuracy of an automated bite counting method in a cafeteria setting," Master's thesis, Electrical and Computer Engineering Department, Clemson University, 2013.

[15] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognition*, vol. 41, no. 6, pp. 2010–2024, 2008.

[16] E. Kelly, *Obesity: Health and Medical Issues Today.* Greenwood Publishing Group, 2006.

[17] S. Kumar, W. Nilsen, M. Pavel, and M. Srivastava, "Mobile health: Revolutionizing healthcare through transdisciplinary research." *IEEE Computer*, vol. 46, no. 1, pp. 28–35.

[18] J. Lementec and P. Bajcsy, "Recognition of arm gestures using multiple orientation sensors: gesture classification," in *Proceedings of the 7th IEEE International Conference on Intelligent Transportation Systems*, 2004, pp. 965–970.

[19] A. Mikhail, C. Frederic, and D. Philippe, "An algorithm for estimating all matches between two strings," INRIA, Tech. Rep., 2001.

[20] C. Ogden, M. Carroll, L. Curtin, M. McDowell, C. Tabak, and K. Flegal, "Prevalence of overweight and obesity in the united states, 1999-2004," *The journal of the American Medical Association*, vol. 295, no. 13, pp. 1549–1555, 2006.

[21] G. Ogris, T. Stiefmeier, H. Junker, P. Lukowicz, and G. Troster, "Using ultrasonic hand tracking to augment motion analysis based recognition of manipulative gestures," in *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, 2005, pp. 152–159.

[22] W. H. Organization, "Overweight and obesity," http://www.who.int/mediacentre/factsheets/fs311/en/index.html, 2008.

[23] J. Salley, "Accuracy of a bite-count based calorie estimate compared to human estimates with and without calorie information available," Master's thesis, Psychology Department, Clemson University, 2013.

[24] E. Sazonov and S. Schuckers, "The energetics of obesity: A review: Monitoring energy intake and energy expenditure in humans," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 1, pp. 31–35, 2010.

[25] D. Schmidt, R. Dannenberg, A. Smailagic, D. Siewiorek, and B. Biigge, "Learning an orchestra conductor's technique using a wearable sensor platform," in *Proceeding of the 11th IEEE International Symposium on Wearable Computers*, 2007, pp. 113–114.

[26] A. Smailagic, D. Siewiorek, U. Maurer, A. Rowe, and K. Tang, "Ewatch: Context sensitive system design case study," in *Proceedings of the IEEE Computer Society Annual Symposium on VLSI*, 2005, pp. 98–103.

[27] STMelectronics. (2013) Mems inertial sensor 3-axis linear accelerometer. http://www.st.com/web/catalog/sense_power/FM89/SC444/PF207281.

[28] STMelectronics. (2013) Mems motion sensor 2-axis pitch and roll gyroscope. http://www.st.com/web/catalog/sense_power/FM89/SC1288/PF248621.

[29] STMelectronics. (2013) Mems motion sensor 2-axis pitch and yaw gyroscope. http://www.st.com/web/catalog/sense_power/FM89/SC1288/PF248616.

[30] C. Sugimoto, H. Ariesanto, H. Hosaka, K. Sasaki, N. Yamauchi, and K. Itao, "Development of a wrist-worn calorie monitoring system using bluetooth," *Microsystem technologies*, vol. 11, no. 8-10, pp. 1028–1033, 2005.

[31] L. Terre, W. Poston II, and J. Foreyt, *12 Overview and the Future of Obesity Treatment*. Springer, 2005.

[32] N. Wellman and B. Friedberg, "Causes and consequences of adult obesity: health, social and economic impacts in the united states," *Asia Pacific journal of clinical nutrition*, vol. 11(s8), pp. S705–S709, 2002.