Using Hidden Markov Models to Segment and Classify Wrist Motions Related to Eating Activities

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Electrical Engineering

---

by
Raúl Ignacio Ramos García
May 2014

Accepted by:
Dr. Adam W. Hoover, Co-chair
Dr. John N. Gowdy, Co-chair
Dr. Richar R. Brooks
Dr. Eric R. Muth

Abstract

Advances in body sensing and mobile health technology have created new opportunities for empowering people to take a more active role in managing their health. Measurements of dietary intake are commonly used for the study and treatment of obesity. However, the most widely used tools rely upon self-report and require considerable manual effort, leading to underreporting of consumption, non-compliance, and discontinued use over the long term. We are investigating the use of wrist-worn accelerometers and gyroscopes to automatically recognize eating gestures. In order to improve recognition accuracy, we studied the sequential dependency of actions during eating. In chapter 2 we first undertook the task of finding a set of wrist motion gestures which were small and descriptive enough to model the actions performed by an eater during consumption of a meal. We found a set of four actions rest, utensiling, bite, and drink; any alternative gestures is referred as the other gesture. The stability of the definitions for gestures was evaluated using an inter-rater reliability test. Later, in chapter 3, 25 meals were hand labeled an used to study the existence of sequential dependence of the gestures. To study this, three types of classifiers were built: 1) a K-nearest neighbor classifier which uses no sequential context, 2) a hidden Markov model (HMM) which captures the sequential context of sub-gesture motions, and 3) HMMs that model inter-gesture sequential dependencies. We built first-order to sixth-order HMMs to evaluate the usefulness of increasing amounts of sequential dependence to aid recognition. The first two were our baseline algorithms. We found that the adding knowledge of the sequential dependence of gestures achieved an accuracy of 96.5%, which is an improvement of 20.7% and 12.2% over the KNN and sub-gesture HMM. Lastly, in chapter 4, we automatically segmented a continuous wrist motion signal and assessed its classification performance for each of the three classifiers. Again, the knowledge of sequential dependence enhances the recognition of gestures in unsegmented data, achieving 90% accuracy and improving 30.1% and 18.9% over the KNN and the sub-gesture HMM.

Dedication

This work is dedicated to God, the Virgin Mary, and to my beloved family and friends.

Acknowledgments

My time spent at Clemson University has been an opportunity to grow in many ways academically, personally, and spiritually. I would like to express my deepest gratitude to those that contributed to the fulfillment of this dream. First of all, I want to thank God and the Virgin Mary for guiding my path, hearing my prayers, and providing the strength to purse my goals. Likewise, to my parents (Victor Ramos and Yolanda Garcia) and my brothers (Victor, Alejandro, and Alfonso) for always encouraging me to do better in life, be my inspiration, and continue to provide your unconditional love and support. I would also like to thank the University of Sonora at Hermosillo, Dr. Bob Lippert, and Dinora Monroy for introducing me to Clemson and for their guidance during the application process to graduate school. I want to acknowledge The National Council of Science and Technology (CONACyT) in Mexico, whom supported financially most of my graduated studies at Clemson and played an important role in my education. Also, I greatly acknowledge the National Institutes of Health (NIH) for the funding provided for the data collection which this project used.

I cannot express enough thanks to my advisor Dr. Adam W. Hoover, which guidance and constant support made this project a reality. I will always appreciate your mentoring in the world of academic research and beyond. I want to acknowledge my committee members Dr. John N. Gowdy, Dr. Eric Muth, and Dr. Richard Brooks, which with their important feedback helped in the quality of this project. Also, I want to give a special thank you to Dr. Daniel L. Noneaker for the financial support and to Dr. Damon L. Woodard's advice during my time in Clemson.

I have been blessed with many friends that I got to know during my studies at Clemson and made this journey even more interesting. First, I would like to express my humble gratitude to those friends that stood by me in times of uncertainty and always found a way to lift me up: Cesar Castellon, Dorismel Diaz, Cecilia Funez, Samuel Zapata, Maciel Ugalde, Felipe Fernandez, Kryssia Diaz and Paulina Guerrero. My ECE colleagues and friends Salil Banerjee, Junphil Kwon, Anup

Sawant, Vinay Gidla, Harshad Gado, Matthew Pepper, Dule Shu, Ryan Mattfeld, and many more, I want to thank you for your help in projects and classes that we shared. My deepest respect to all the Latin community and the SACNAS Chapter at Clemson University from whom I learned much. Lastly, to Berenice Castaños, I will always treasure your kind words, prayers, friendship, and constant support in times of need, my greatest and sincere thank to you.

"Trust in the Lord with all your heart and lean not on your own understanding." (Proverbs 3:5)

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The primary measurements used in the study and treatment of obesity are energy intake (EI) and energy expenditure (EE). EI measures the amount of energy absorbed through food intake, and EE measures the energy cost of homeostasis (body maintenance) plus physical activities. At least in part, obesity results from a mismatch between EI and EE [79]. The units of EI and EE are calories (or joules). EI is a function of food intake, whereas EE is a function of resting metabolic rate, the thermic effect of food, and the thermic effect of physical activity.

EI measurement is classically concerned with determining a person's food intake and converting this information to energy and nutrient intake [8]. Paper-based methods to measure food intake include the 24-hour recall, food records, and food frequency questionnaires [168, 115, 50]. The 24-hour recall uses a directed interview to quantify the consumption of food and beverages of a single day [181]. Food records, such as a 7-day diet diary, require the subject to record time, type, and amount of food eaten [49]. Food-frequency questionnaires are designed to asses the habitual diet of a subject by asking the frequency with which food items are consumed [49]. Extensions to these classic tools include computer-based methods that can scan bar codes and record spoken descriptions of food intake over the course of a day [156]. Other methods include keeping a daily personal health record of food intake and daily exercise [182, 104] to estimate weight gain, weight loss, and nutritional information.

The accuracy of these methods is typically ascertained by comparing measurements of EI

against simultaneously measured EE. Accounting for weight gain/loss, EI should match EE, therefore any discrepancy may be considered error in the measurements. The most accurate method for measuring EE is the calorimetry chamber, in which a subject lives in a controlled environment. In this scenario, EI can also be accurately measured using bomb calorimetry for all foods given to a subject. These methods are expensive and obviously impractical for long-term and widespread use [158]. For free living, the most accurate method for measuring EE is doubly labeled water [159]. Typically used over a period of one week [132], the method measures the average metabolic rate using water in which hydrogen and oxygen are replaced by deuterium and oxygen-18. The loss of these isotopes is tracked through daily samples of saliva, urine, or blood. Disadvantages of this method include its cost and the technical expertise needed to take samples [150]; however, it remains the gold standard against which EI measurement tools are evaluated. It is well established that EI calculated from information provided by classic self-report methods show inconsistencies with measured EE. Food record, food-frequency questionnaires, and recall methods are designed to be self-administered [181]. People have a tendency to underreport, with estimates ranging from 10% to 50% [38]. This issue may be caused by several factors [149, 103, 83, 74, 34]. Using recall methods, people can forget what they consumed. Food-frequency tools require long lists of items, with usefulness somewhat determined by the overlap of the food list with typical eating habits. Food records require accurate self-estimation of portion sizes, but it has been shown that people tend to overestimate large portions and underestimate small portions. Another limitation of these tools is that they place a constant responsibility on subjects to accurately log the information [181]. Within the dietetics research community, there is a well-recognized need to develop new methods for measuring EI in free living that are more accurate, less expensive, and have less subject and experimenter burden [112, 169].

This dissertation investigates the use of wrist-worn accelerometers and gyroscopes to automatically recognize eating gestures. Previous work done by our research group studied the tracking of wrist motion as it relates to eating [55]. A method was developed to detect a pattern of wrist motion associated with the taking of a bite, defined as placing food or liquid into the mouth [57]. The method was shown to be accurate across a wide variety of foods, counting bites with a true positive rate of 86% and a positive predictive value of 82% [57]. Additional research showed that bites, automatically counted using this method, correlated with self-reported caloric intake at the meal level at 0.5 [151]. The work in this dissertation is concerned with improving upon this method to

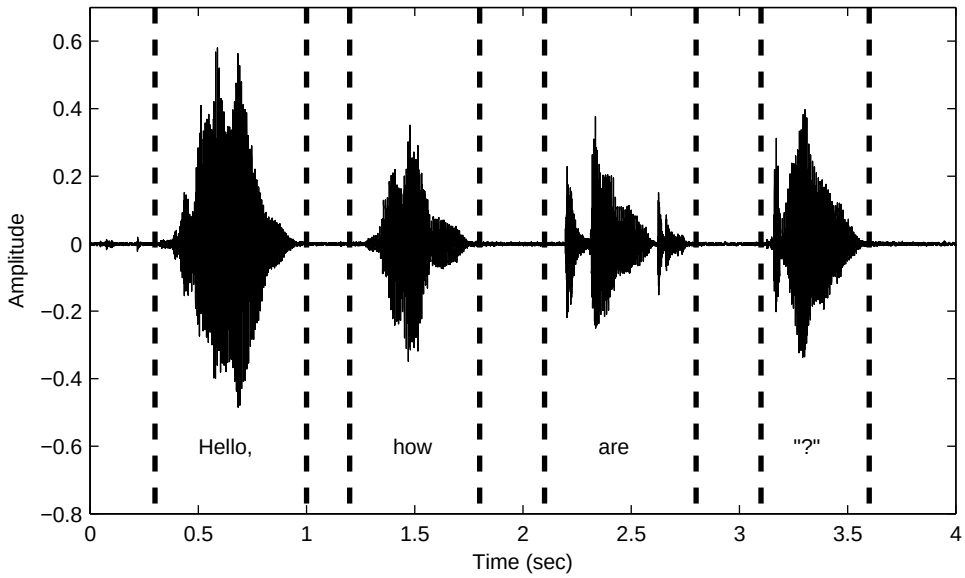


Figure 1.1: Speech message.

automatically measure EI. The original method treated all bites the same, regardless of context, and used a single pattern for detection. The proposed idea in this work is to study temporal sequencing as it relates to eating activities. By defining common activities during eating, we seek to determine if the recognition of previous activities can be used to improve the recognition of subsequent activities.

To begin studying this idea, first a "language of eating" is defined. Languages obey grammatical rules in order to make messages coherent; i.e., the ordered structure of the words is meaningful. Figure 1.1 shows an example from a speech signal. The recognition of each piece of signal can be undertaken independently, in order to determine the word. However, if the recognition results from the previous pieces of signal are known, this can constrain and improve the subsequent recognition of the next piece of signal. In this example, the recognition of the piece of signal labelled "?" can be improved by using the recognition results of the previous pieces of signal ("Hello" "how" "are"). For this example, it may be highly expected that the next piece of signal encodes the word "you". This dissertation pursues the same idea in the language of eating. The signal is obtained using accelerometers and gyroscopes to track wrist motion. The recognition of a piece of signal can be done independently, or can be augmented using recognition results from previous pieces of the signal. Figure 1.2 shows an example. If the actions "inactive" and "manipulate food" have previously been recognized, then it may be highly expected that the next action is "take a bite".

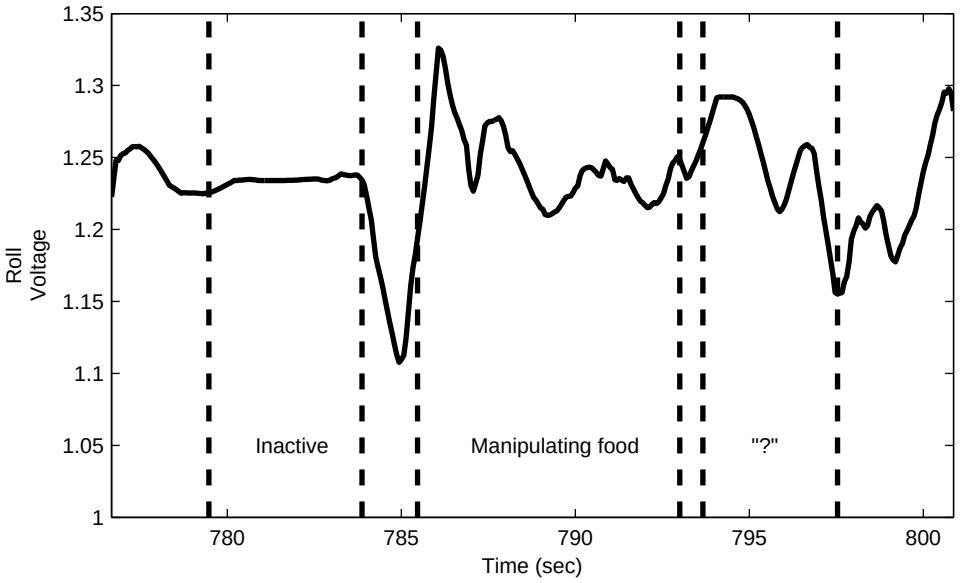


Figure 1.2: Eating activity signal (roll).

Our experiments use data recorded by a wrist-worn device mounted with gyroscopes and accelerometers MEMS sensors which track wrist motion. A sample of data collected from this device is shown in figure 1.3. The plot shows analog voltages of the accelerometers (AccX, AccY, and AccZ) and gyroscopes (Yaw, Pitch, and Roll) tracking wrist motion of a subject for approximately 48 seconds while eating. A collection of snapshots are shown on the top illustrating the activities during this period. From left to right, these activities are drinking, moving food items around the plate, executing a bite with and without utensils, and being in an inactive position. The shaded areas of the signal represent the length of the activity taking place. We extract features from these six signals and use them to build models for temporal analysis.

Among the tools for studying temporal sequencing of data, hidden Markov models (HMMs) are one of the most popular techniques. An HMM is a well-known probabilistic tool used in speech processing, which uses previously gathered data to make predictions of future information [136]. In this work, HMMs will be used to model the temporal sequencing of actions related to eating and to evaluate the utility of this idea for improving automated measurements of EI.

The following sections provide background information regarding the areas of study which touch upon the research discussed in this dissertation. Section 1.2 describes the problem of obesity. Section 1.3 gives an overview of mHealth systems and their importance. Section 1.4 discusses

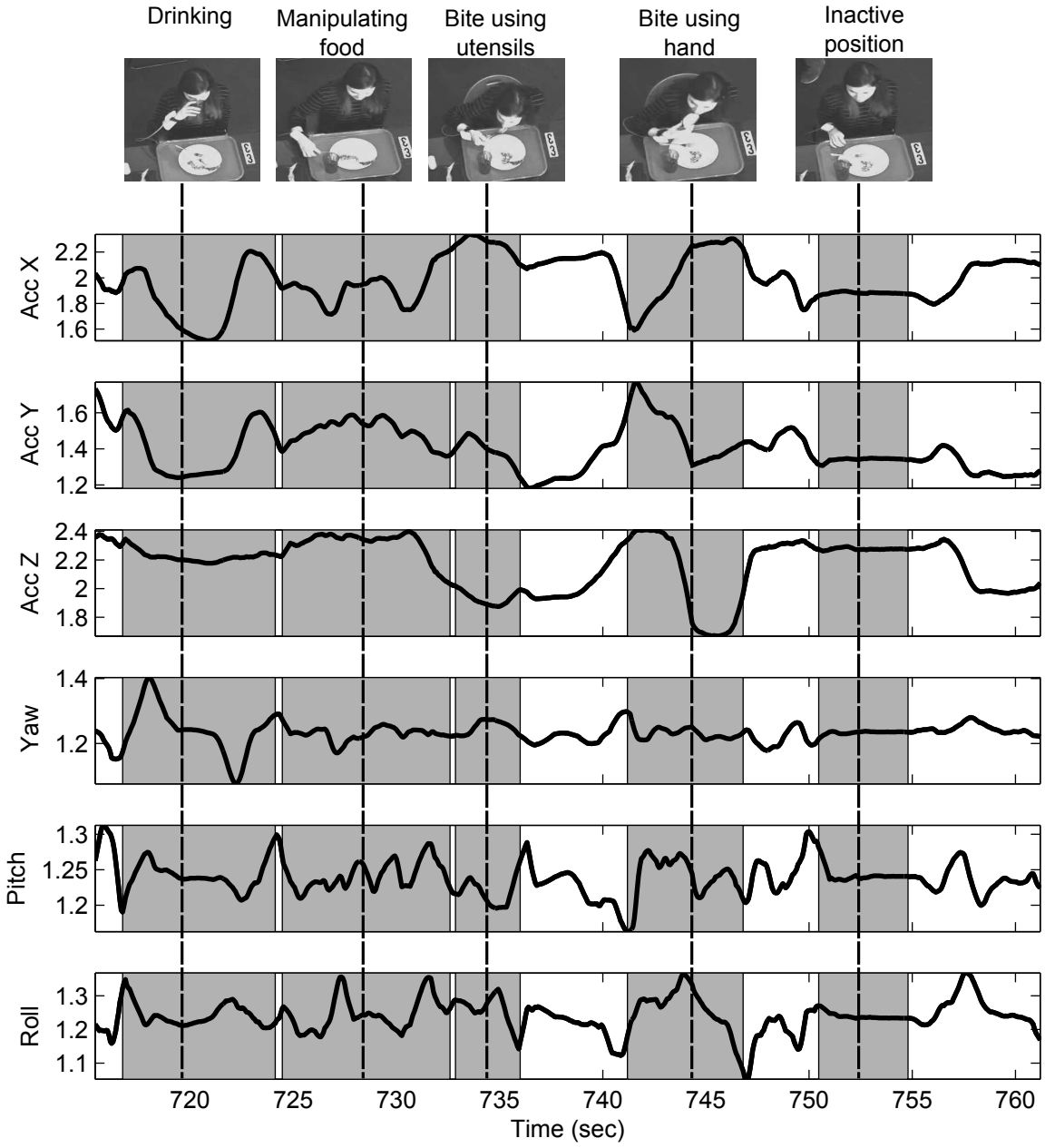


Figure 1.3: Eating activity sequence.

physical activity monitoring and the measurement of EE. Section 1.5 describes related work using body-worn sensors for automatic dietary monitoring. Section 1.6 briefly describes related work on gesture recognition. Section 1.7 gives an introduction to MEMS sensors, describes the type of sensors used in this work, and demonstrates raw data during eating activities. Section 1.8 describes eating patterns at three main levels (daily, meal, and microstructure). Section 1.9 explains the theory of hidden Markov models. Lastly, section 1.10 defines the novelty and specific contributions of this work.

1.2 Obesity

Obesity is a medical condition associated with abnormal or excessive fat accumulation. It is typically defined by body mass index (BMI) and is calculated by dividing the subject's weight (kilograms) by the square of height (meters). The World Health Organization defines overweight as a BMI$\geq$25, obese as a BMI$\geq$30, and morbidly obese as a BMI$\geq$40 [7].

Obesity is causing health damage to the population worldwide, and is considered to be a global epidemic [4, 35]. Obesity has been shown to correlate to several diseases [117, 68, 53, 74] including type 2 diabetes mellitus, gallbladder disease, coronary heart disease, high blood cholesterol level, high blood pressure, osteoarthritis, cardiovascular disease, several types of cancer (endometrial, postmenopausal breast, kidney, and colon), musculoskeletal disorders, sleep apnea, asthma, infertility, and depression. According to the World Health Organization, obesity has more than doubled since 1980 and 65% of the world's population lives in countries where obesity is among the leading causes of death [7].

The United States has seen an increase in the number of adults, adolescents, and children suffering from obesity. The National Health Nutrition Examination Surveys (NHANES) [5], in the years of 2009 and 2010, found that 35% of the adult population was obese. Additionally, 17% of adolescents and children were obese. These percentages represent over 78 million U.S. adults and approximately 12.5 million U.S. children and adolescents [122]. The "Healthy People 2012" initiative of the U.S. Department of Health and Human Services [6] set goals to reduce the prevalence of obesity among adults (ages 20 and over) to 15% and 5% among adolescents (ages 12-19). However, the obesity rate in children (ages 6-11) increased from 11% to 18% over the same time period. Recent studies [67] have concluded that if obesity trends were to continue based on 2010 levels, the

combined costs from medical expenditures over the next 2 decades would be $549.5 billion with a 33% increase in obesity prevalence and a 130% increase in severe obesity. By 2030 obesity rates would reach 42% and severe obesity would reach 11%.

There are three main strategies used in the treatment of obesity: surgery, pharmacology, and behavioral change. Surgical procedures, [121, 45] such as gastric bypass, gastric banding, vertical banded gastroplasty, and laparoscopic cholecystectomy have been shown to be effective for significant weight loss, but are only considered for those who are morbidly obese, have failed conventional methods for weight loss, and are good candidates from a psychological standpoint. Pharmacology methods [80] use medication to reduce appetite, decrease nutrient absorption, or increase thermogenesis (EE); some of these have shown negative side effects on the human body (e.g. valvular heart disease, cataracts, neuropathy, and hemorrhagic strokes). Behavioral weight loss methods [74, 34] involve increasing EE, decreasing EI, and altering behavior.

Recently, there has been an increased interest in involving an engineering perspective in the study of diet and obesity. Scientific research in dietetics is dependent on improving dietary assessment methodology [112]. The lack of accuracy inhibits dietetics practitioners from obtaining accurate information on which to base nutrition care plans. Computerized questionnaires using mobile devices and the Internet have improved data recording, but researchers have also worked on developing systems that can take advantage of sensors mounted on certain parts of the body [18, 144].

1.3 Mobile Health (mHealth)

Technology has revolutionized medical care in many ways. eHealth, short for electronic health, is concerned with bringing technology into hospital and supervised care settings [15]. It includes the use of modern information and communication technologies to meet the needs of citizens, patients, healthcare professionals, healthcare providers, and policy makers. Examples include bringing telemedicine into hospitals and managed healthcare, electronic medical records, allowing access to patient records by pharmacists, sharing information between clinicians, and even sharing information between same-site facilities. mHealth, short for mobile health, is a similar area of research but is concerned with using technology to empower self-care. Examples include self-administered diagnostic tests and daily health monitoring, with a focus on lowering costs and enabling individuals to better manage their own health. The goal of mHealth systems is to improve a subject's health by

monitoring their status, recognizing behaviors, diagnosing medical conditions, and providing interventions if necessary through the use of wireless portable devices capable of storing, processing, and retrieving real-time and non-real-time data from the users [9, 97]. The impact and effectiveness of mHealth systems is dependent on devices that provide connectivity and enable the user to become players in managing their own health and care. Wearable wireless sensors are expected to grow by 400 million units by 2014, while the number of operational mobile subscribers already surpassed the 5.2 billion mark in 2011 [9]. The market for mobile health applications is expected to grow by 6 billion US$ by the end of 2015 [141].

mHealth systems have enhanced monitoring and recording data for several health-related issues through the use of embedded sensors and applications in mobile devices, complemented with body-worn sensors. mHealth sytems have made many contributions to cardiovascular monitoring [133], diabetes [42], calorie intake monitoring [171], blood oxygen levels [23], blood sugar [12], and hypertension [43], among others. For example, Poh et al. [133] developed a system called Heartphones which measures bilateral volume pulse for cardiovascular monitoring using a reflective photosensor mounted on common earbuds. A custom application was built into a smartphone platform for processing data and displaying the resulting information to the user. Chomutare [42] worked on the development of a self-monitoring application for patients with type 2 diabetes. Alinker et al. [23] created a wrist-worn device which measured blood pressure, oxygen saturation, and ECG to monitor the status of high-risk cardiac/respiratory patients. The device has a communication interface which connects to a telemedicine center. Tsai [171] describes the development of a system to register and calculate caloric intake using an application on a mobile device which is connected to a server application. The server application reminds the user to update caloric information. Sung et al. [165] presents an ambulatory health monitoring system called LiveNet. This system has a PDA, a microcontroller board connected to a variety of sensors (activity detection, oxygen saturation, ECG, EMG, galvanic skin response, respiration, temperature, and blood pressure), and an application which registers data and processes it using machine-learning techniques. This system was built to be used in several clinical studies such as Parkison's disease monitoring, depression monitoring, epilepsy seizure detection, and general activity detection. In previous work by our research group, Dong [55] worked on an experiment in which accelerometers and gyroscopes were embedded in the iPhone 4 to recognize activity events during the day (ambulatory, sedentary, eating). Pantelopoulos [125] provides a broader survey on biosensor systems for use in health monitoring.

1.4 Physical activity monitoring

Several methods for measuring EE have been created, including behavioral observations [166, 173], self-report [89, 98], physiological markers [179, 161], motion sensors [180, 131, 82], and indirect calorimetry [107, 64]. This section focuses on methods that use body-worn sensors to monitor physical activity in order to measure EE. Physical activity is defined as body movement by skeletal muscles resulting in EE [37].

The most commonly used sensors are accelerometers and gyroscopes. Vathsangam et al. [175] used a triaxial accelerometer and a triaxial gyroscope mounted on the hip. Fourier analysis was used to examine the cyclic nature of walking and show its relationship with EE. They concluded that EE was most effectively predicted by utilizing both types of sensors instead of one alone. Badawi [25] implemented a system using a 3-axis accelerometer attached to the shoes of children. This device captured the physical activity and measured daily EE to provide nutrition recommendations. Plasqui et al. [132] predicted total EE and activity-related EE by using a 3-axis accelerometer placed on the lower back to record acceleration of the torso. Kinnuen et al. [94] used a one-dimensional acceleromenter mounted on a wrist-worn device to count hand movements and predict EE. When height and mass were included in the prediction model, the device showed a high correlation between the number of hand movements and EE. Foster et al. [71] developed an ankle-worn pedometer using a dual-axis accelerometer, and found a positive correlation between steps counted and walking EE. Lin [101] created a system using accelerometers on the wrist, waist, and ankle as well an ECG placed near the heart to monitor its electrical activity. The system combined the information gathered from both sensors and used EE regression models and neural networks. Sazonova [148] presented SmartShoe, a device that consists of five small pressure sensors embedded in key support locations of the insole and a heel-mounted 3D accelerometer (a total of 10 pressure sensors and 2 MEMS accelerometers) with the objective of estimating the body weight of a person and indirectly measuring EE.

Many studies have shown that self-reported estimates of EE suffer from bias [46], [108], [180]. Body-worn motion sensors provide more objective measurements with less user burden at less cost [60], and have become widely adopted in research and practice. However, the development of similar methods to measure EI using body-worn sensors has proven more difficult. The following section describes related work that uses body-worn sensors for the assessment of dietary behavior.

1.5 Automatic dietary monitoring systems

Automatic dietary monitoring (ADM) refers to the use of sensors placed on the body or in the environment to collect and process information to derive patterns about dietary behavior of users. ADM has the challenge of capturing and assessing several aspects of eating sessions, such as the time taken for food consumption, eating microstructure (chewing, oral bolus formation, tongue activity, and swallowing), meal composition, and the volume of food [18]. Currently, there is no device which can accomplish all these tasks. Much of the work on ADM has been concentrated towards collecting and analyzing data from different parts of the body. Amft [18] and Sazonov [142] have published reviews on ADM systems using body-worn sensors. The next three sections discuss previous works that monitored different parts of the body for chews, swallows, and gestures. The last section discusses the use of mobile cameras for capturing dietary intake. The work in this dissertation could be utlilized in many of these methods, where the study of the sequential nature of eating activities could improve recognition and measurement accuracy.

1.5.1 Chew-based systems

Research into automatic detection of chews uses microphones placed on the ear or throat to measure the sound created by chewing, or sensors placed on the jaw to monitor motion while chewing. Amft et al. [20] researched the detection of eating activities and classified four food types using chewing sounds collected by a microphone located inside the ear canal. In an extended work [16], the sensor was enhanced using a headphone-housed acoustic transducer to capture vibration and signal patterns under minimal ear occlusion conditions. Nishimura [120] used a Bluetooth headset and adapted an in-ear microphone to count the number of times food was chewed during a meal. Shuzo [154] combined a bone conduction microphone, a condenser microphone, and an IC recorder to record sound chews and classify activities such as eating, drinking, and speaking as well as provide chewing count. Sazonov [145] and Fontana [70] used a piezoelectric strain gauge sensor to capture motion of the lower jaw to monitor chewing. Liu et al. [102] proposed a wireless sensor which integrated a microphone and a camera (with a similar viewing angle to the subject's eye) to record real-time sounds (speech, eating, drinking, and others) and images of food while the chewing action was taking place. A restriction was that the user must have his or her food on a circular plate or bowl. Päßler et al. [128] proposed a work for automatic food intake recognition through adapting

Table 1.1: Summary of chew-based systems.

Year	Publication	Objective	Subjects	Features	Classifier	Metric(s)
2012	Liu et al. [102]	Recognize 4 types of sounds (speech, eating, drinking, others)	6	Energy features, spectral features,temporal features, color-histograms	Extreme Learning Machine	82.51%•/24.18%• (Eating/Drinking)
2012	Päßler et al. [128]	Adapt chewing sound models to classy 7 food types and 1 drink	2	MFCC and delta coefficients	GMM-HMM+MAP	71.3%•
2012	Sazonov et al. [145]	Detect periods of chewing and no chewing events	20	RMS, entropy, median, mean, max, zero-crossing, peaks, others (25-features)	SVM	80.98%▷
2012	Fontana et al. [70]	Discriminate between solid and liquid intake	7	Time and frequency domain filtered by two band-pass filters	SVM (Linear and RBF)	90.52%▷
2011	Cadavid et al. [36]	Discriminate chewing and no chewing events by explorong the visual quasi-periodicity of the mouth while eating	36	Power spectrum coefficietns from AAM parameters	SVM	93%•
2010	Amft et al. [16]	Classify 19 food type based on chew sounds	2	Spectral features	Naive Bayes	86.6%•
2010	Shuzo et al. [154]	Discriminate between 4 types of eating actions (eating a hard food, eating a soft food, drinking water, speaking). Also, count the number of mastications.	5	Barycentric frequency, max peak frequency, roll-off frequency, LPC coefficients, total power, ratios of divided section power of 1/3 octave band to the total power	KNN	70%•
2008	Nishimura et al. [120]	Count the number of chews	-	MFCC	One-vs-all LBG Codebook	98.07%•
2005	Amft et al. [20]	Recognize 4 food types	4	Zero-crossing rate, fluctuation of amplitude and spectrum, freq centroid, spectral roll-off, band energy ratios, 6 cepstral coefficients	C4.5 decision tree	80%-100%•

•-Accuracy, ▷- Avg precision and recall

sound models of chewing using the maximum a posterior estimation algorithm (MAP). MAP is a well known algorithm in speech processing which reduces the error caused by mismatch of model and features parameters [76]. This strategy was adopted because of the inter-individual differences of food intake sounds, causing a low recognition on new presented data as compared with the training data. A recent method to study chewing is one proposed by Cadavid et al. [36], which uses video information of a subject's face while eating. Features are extracted from the image sequences using a an active appearance model, which describes shape and appearance deformation around a given region [48]. Table 1.1 summarizes several characteristics of the studies mentioned here.

1.5.2 Swallow-based systems

Swallowing is an action inherently linked to food intake activities. Swallowing is often partitioned in three phases: oral preparation, the pharyngeal, and the esophageal phase [51]. The first phase has to do more with chewing while the other phases concern activation of muscles which move the food intake to the stomach. Researchers have monitored food intake using microphones on the ear and throat to capture the sounds produced by swallowing and others sensors such as a surface electromyograpgh to capture variations in the muscle patterns in the throat. Lopez-Meyer et al. [105] detected swallowing sequences through supervised and unsupervised methods, specifically SVM and KNN. Walker et al. [177] used an elastic band containing two surface microphones to record swallowing by mounting the sensor directly on the throat. Sazonov et al. [147] proposed a method to automatically detect swallowing, differentiate between solid and liquid food intake, and estimate ingested mass. Dong and colleagues [54] used a piezo-respiratory chest-belt, wrapped around the chest of the subject, to monitor breathing; future work with the device could encompasses the detection of swallows. Sazonov [143, 142] used a microphone positioned in the laeyngopharynx for swallowing detection, also a strain sensor implemented below the ear to detect chews and bites. Their work showed the ability to detect periods of food intake, differentiate between solid foods and liquids, and predict the mass of ingested food (solid and liquids). Päßler [129] recorded chewing and swallowing sounds by implementing an in-ear microphone in the outer ear canal of the right ear while a hearing aid package with a reference microphone was applied behind the outer ear to record environmental sounds. Sounds were used to detect activity of food intake and classify consumption of seven types of food and two type of beverages. Table 1.2 summarizes several characteristics of the studies mentioned here.

1.5.3 Gesture-based systems

Accelerometer and gyroscope sensors mounted on the body have also been explored for measuring food intake. Dong et al. [56] developed a wrist-worn device to track wrist motion and measure the number of bites taken during a meal. Additional research showed that bites, automatically counted using this method, correlated with self-reported caloric intake at the meal level at 0.5 [151]. Dong also developed a method to automatically detect eating activities during free living by continuously tracking wrist motion and recognizing a pattern indicative of a meal or snack [55].

Table 1.2: Summary of swallow-based systems.

Year	Publication	Objective	Subjects	Features	Classifier	Metric(s)	
2011	Walker et al. [177]	Discriminate swallow, vocal chord activation, clearing of throat, coughing	2	Energy and max absolute value inside window	Threshold scheme	-	
2011	Passler et al. [127]	Detect periods of food intake	40	Energy levels	Threshold scheme	83.3%$^{\bullet}$, 81.8%*	91.3%$^{\circ}$,
2010	Sazonov et al. [147]	Automatic detection of swallows	20	Mel-scale Fourier spectrum and wavelet packet decomposition	SVM	84.7%$^{\bullet}$	
2010	Lopez-Meyer et al. [105]	Discriminate swallow events	18	Absolute time difference between neighboring swallows	SVM and KNN	93.9%$^{\bullet}$, 98.5%$^{\diamond}$	92.4%†,
2009	Sazonov et al. [142]	Detect periods of food intake, deffirenciate solid and liquid intake, estimate mass intake	20	Average instantaneous swallowing frequency	Bayesian	>95%$^{\bullet}$, >83%$^{\bullet}$	>91%$^{\bullet}$,

$\bullet$-Accuracy, $\circ$-Precision, $\star$-Recall , $\dagger$-Sensitivity, $\diamond$-Specificity

A wrist-worn acceleration sensor was used by Amft [17] with the primary objective of detecting drinking activities, the container used, and the fluid level. Two drinking motions were studied, fetch and sip. Fetch describes a sequence of motion from picking up a container until placing it back in its original position, while sip describes a shorter gesture that only includes moving the hand to the mouth and back. They concluded that fetch and sip gestures could be spotted with similar accuracy, and sip is specific enough to indicate drinking in continuous data. Junker and Amft [85, 19] presented a recognition system that used five inertial sensors located on the wrists, upper arms, and upper torso. Their research describes motion gestures based on the particular utensil used, establishing four gestures (cutlery, drink, spoon, hands). Amft [22] later augmented the recognition by incorporating chewing and swallowing information through an ear-mounted microphone, and a collar combing a surface EMG and a stethoscope-like microphone as sensors. Table 1.3 shows a summary of these works.

1.5.4 Camera-based systems

Some researches have tried using computer vision algorithms coupled with mobile devices to measure EI. Chen [39] shows a technique to obtain food volume based on a image. After capturing the image the user must select a 3D shape model in which position, scale, and orientation parameters are optimized to obtain a volume measurement. Almaghrabi et al. [14] worked on a food recognition system which was coupled with nutrition tables to obtain energy intake estimation of food. The system requires the user to point at the food to start the process. A top and side view image of

Table 1.3: Gesture-based systems.

Year	Publication	Objective	Subjects	Features	Classifier	Metric(s)
2012	Dong et al. [57]	Count the number of bites	47	Roll motion velocity	Threshold scheme	85%†, 81%$^{\triangleright}$
2011	Dong et al. [58]	Automatic detection of eating activities	4	Variance of yaw, pitch, and roll velocities, bites per minute, occurrences of not detect bites over 1 minute	Threshold scheme	82%†, 70%$^{\triangleright}$
2010	Amft et al. [17]	Detect the sips in the drinking action	6	Sum of absolute amplitude, difference of begin and end of amplitude, number of zero-crossings, amplitude begin and end, sum of amplitude differences, mean, and variance	Feature similarity search	94%$^{\circ}$, 84%*
2009	Dong et al. [56]	Count the number of bites	10	Roll motion velocity	Threshold scheme	91%†
2008	Junker et al. [85]	Find sections with relevant gesture motion and classify 10 gestures; four gestures are related to eating (cutlery, drink, spoon, handled)	4	Pitch and roll angles from upper and lower arm, derivative and sum of acceleration signal of lower arm (along pitch angle), derivative and sum of rate of turn from lower arm (along roll angle)	GMM-HMM	57%$^{\circ}$, 80%*

$\circ$-Precision, $\star$-Recall , $\dagger$-Sensitivity, $\triangleright$-Positive predictive value (PPV)

the food are use to calculate the volume using area sizes. An extension of this work is presented by Villalobos et al. [176] in an application in a smartphone. Volume is used to compute calories using predefined nutritional tables. Wu and Yang [184] presented a work to recognize fast food using a web camera and estimate calories from a database of 101 types of fast foods. Video was taken avoiding the participant's face, conversations, and interruptions; pictures were also collected. Image processing was used to compute and classify fast food features. The resulting food classification was later associated to a nutritional table estimating calories based only on food appearance not food portions. Martin [111] describes a semi-automated system which requires the participation of the user to train the system for further estimation of food quantity based on pictures. In the training phase the user must use a reference card for calibration, manually select the area of each food type, and provide grams information. After training, the user may take pictures which automatically detects and classifies food and provides total grams as well. This information is sent to dietitians to recheck the information and adjust for errors manually. Other works for measuring energy intake from images can be found in [164, 186, 73, 167, 138, 95, 153, 190, 134]. In general these methods will require considerable more research before a fully functioning system can be developed that can automatically measure EI of any foods in any environment with little-to-no human assistance.

1.6 Gesture recognition

A gesture can be described as a configuration and/or movement of a part of the body (arms, hands, fingers, face, head, torso, pose, eye gaze) with the objective of expressing an emotion, intent or command. There are several devices used (both in research and commercially) to convert a gesture into digital information such as touch screens, 2-D and 3-D cameras, acoustic sensors, magnetic sensors, and mechanical sensors [27]. Gesture recognition has been applied to several problems, including sign language [124, 88], activity recognition [172, 157], virtual reality [44], and human-computer interaction [114, 130].

Much of the literature related to gesture recognition has been concerned with algorithms to detect, track and recognize actions in video sequences [172, 114, 124], but wearable sensors have also been explored. Specifically, accelerometers and gyroscopes are capable of determining the acceleration, position, and orientation of an object. Akl and his colleges [13] developed an accelerometer-based gesture recognizer of 18 traces using the Nintendo's Wii remote. Chen [40] also used this remote to develop 16 fundamental motion gestures which could represent digits. Bui [33] developed a glove with six accelerometers to classify 23 Vietnamese letters for sign language. Other studies motivated by sign language [96, 100] used surface EMG sensors with a 3-axis acceletometer mounted on the forearm and hand. Chen et al. [41] combined accelerometers and surface EMG to recognize 24 hand gestures based on wrist and fingers movements. Zhang [189] developed a hand gesture recognition device based on a Rubik's cube using EMG and accelerometers. The device defined three basic hand postures and six circular movements with a total of 18 gestures as control commands. Wang et al. [178] proposed a wireless digital pen with an accelerometer and used it to recognize digits and eight shape trajectories. Xu et al. [185] used a MEMS 3-axes acceleration sensing chip integrated with data management and Bluetooth wireless data chips to capture and classify 7 hand gestures (up, down, left, right, tick, circle, and cross). Other studies have used the accelereometers built into mobile devices to perform gesture recognition by moving the device in specific patterns [119, 91]. Dermitzakis [52] used a gyroscope-based system for upper-limb prosthetic gesture recognition. The system uses five sensors distributed along the right upper arm, the upper back and the head. Oh [123] developed a wand-like artifact built with accelerometers and gyroscopes with an objective of recognizing 10 characters and 3 gestures drawn in the air. Murao [116] presented a study on 27 gestures using accelerometers and gyroscopes and assessed the accuracy of the results

by altering the location of the sensors in an experimental device. The work in this dissertation bears some resemblence to research in gesture recognition, but the application domain of eating provides novel context for improving recognition accuracy.

1.7 Micro-electromechanical systems sensors (MEMS)

Micro-electromechanical systems (MEMS) are small (nano, micro, mili scale) integrated devices or systems that combine electrical and mechanical components. MEMS terminology is used in the U.S.; in Europe and Asia [84], microsystems technology (MST) and micromachines are sometimes used for the same purpose as MEMS. Maluf & Williams [109] describe MEMS as being simultaneously a toolbox, a physical product, and a methodology. MEMS provide techniques and processes to design and create miniature systems. The physical product is often designed for a specific task, combining functions of sensing and actuation. MEMS encompasses a plethora of sensors such as pressure sensors [146], optical sensors [99], thermal sensors [84], accelerometers [126], gear trains [77], mircodroplet generators [11], gyroscopes [140], strain gauges [84], microphones [187], micro-mirrors [93], probes [92], imagers [66], analyzers [72], global positioning system sensors [62], magnetometers [174], electrocardiography sensors [188], miniature robots [32], and others. The diversity and low power consumption of MEMS sensors makes them useful in various commercial (e.g. biomedical, automotive, and telecommunications) and military (e.g. navigation and weapons) applications.

The work in this dissertation uses MEMS accelerometers and gyroscopes. These measure linear acceleration and angular velocity. The combination of these measurements is sometimes called an inertial measurement unit (IMU) [24].

1.7.1 Accelerometers

Accelerometers follow the principle of the spring-mass system (figure 1.4) using Newton's second law of motion and Hooke's law. Newton's second law of motion states that if an object with mass m has experienced an acceleration a, then a force is acting on this mass. The force is proportional to the product of the mass and the acceleration, as seen in equation 1.1. In Hooke's law, an equation for the force applied to the spring is derived by considering the distance, Δx, of

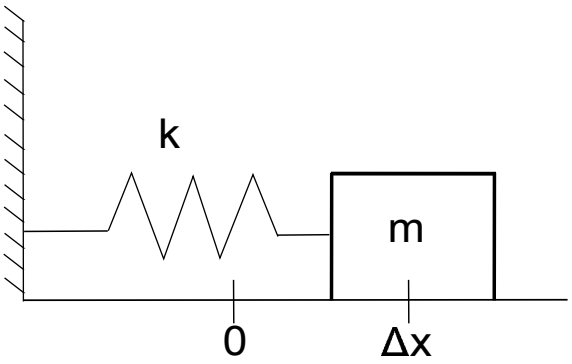


Figure 1.4: Spring mass model.

the spring displaced from its equilibrium position and a spring constant k, as seen in equation 1.2.

$$F = ma \tag{1.1}$$

$$F = -k\Delta x \tag{1.2}$$

The previous equations can be combined to provide a method for measuring acceleration using the linear displacement of the spring, as shown in equation 1.3. Negative acceleration refers to the compression of the spring. Ultimately, the acceleration measurement gets reduced to a linear displacement measurement.

$$a = -\frac{k}{m}\Delta x \tag{1.3}$$

1.7.2 Gyroscopes

Gyroscopes measure the angular velocity rate of a spinning mass to determine its orientation. Angular momentum is the principle that governs gyroscope operation and is represented by a vector quantity resulting from the product of the rotational inertia and rotational velocity of a particle about a particular axis. Equation 1.4 describes the angular momentum of a particle $\mathbf{L}$ with mass m, velocity $\mathbf{v}$, and distance from the origin $\mathbf{r}$.

$$\mathbf{L} = m\mathbf{r} \times \mathbf{v} \tag{1.4}$$

A MEMS gyroscope uses the Coriolis effect to compute a Force measurement. Equation 1.5 shows the parameters of the Coriolis equation, where $\mathbf{\Omega}$ are the angular velocities, $\mathbf{v}$ are the velocities of the particle in the rotating system, m is the mass of the particle, and $\mathbf{F}$ are the inertial forces.

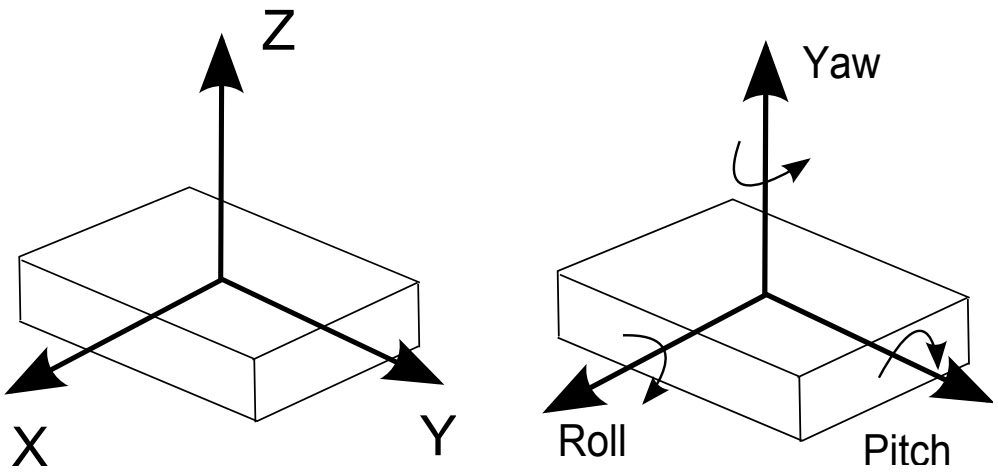
$$\mathbf{F} = -2m\mathbf{\Omega} \times \mathbf{v} \tag{1.5}$$

1.7.3 Wrist tracking data

In this work a custom wrist-worn device was built, containg gyroscopes and accelerometers to track the wrist's motion. The specific sensors used were STMicroelectronics LIS344ALH [1], LPR410AL [2], and LPY410AL [3] MEMS sensors. The device was used to record wrist motion data of humans during the consumption of a meal. A broader description of the development of the device can be found in the works of Dong [55] and Drennan [60]. A brief description of the sensors and the data recorded is provided next.

The LIS344ALH, is three-axis linear accelerometer used in the device. Figure 1.5a shows the three axes (X, Y, Z) that measure linear acceleration in the specified direction. The sensor has an element which displays a changing capacitance based on its acceleration and an integrated circuit interface which outputs three analog signals (voltages) for analysis. The linear acceleration is given in gravity units (g). The accelerometer does not directly sense gravity but rather a deviation from free fall, i.e. if the sensor is laying on a horizontal surface this will measure 0 g in the X and Y axis whereas the Z axis will measure 1 g. The sensor operates in a ± 2 g scale, thus gravity values of +2, +1, 0, -1, -2 represent voltages of 3.3, 2.31, 1.65, 0.99, 0.2 with a $\pm 5\%$ tolerance. Figure 1.6 shows an example of the signals provided by the 3-axis acceleremoter of a subject in the action of taking a drink.

MEMS gyroscopes were also used in the device. The LPR410AL gyroscope is used to capture roll and pitch and the LPY410AL gyroscope is used to capture pitch and yaw. One of the pitch signals is ignored. These gyroscopes include a sensing element composed of a single driving mass, kept in continuous oscillation and capable of reacting, based on the Coriolis principle, when an angular rate is applied. Figure 1.5b provides a visual representation of the axes used for estimating the angular rate of a rotational movement. The unit of measurement for angular velocity used is degrees per second (dps). Each gyroscope provides an output voltage signal, where 2.5 mV represents 1 dps. The signals provided by the sensors are combined with a 1.24 V DC signal to allow representation of



(a) Accelerometer X, Y, Z axis. (b) Yaw, pitch, roll axis.

Figure 1.5: MEMS sensors.

negative voltages. Figure 1.7 shows an example of the raw data of a subject in the action of taking a bite.

1.8 Eating patterns

The phrase "eating pattern" can refer to a variety of characteristics, including food consumption, beverage consumption, portion size, meal patterns and frequency, school meal participation and consumption, and dietary quality [118]. The concept of an eating pattern can be studied at different time scales, including the daily level, the meal level, and the microstructure level. Patterns at the daily level include the frequency of meals and the temporal distribution of eating events throughout the day. Meal level patterns include the quantity of calories consumed, nutrition information, and the pace at which a subject eats. The microstructure level is concerned with bite size portions, and patterns in chewing and swallowing during food consumption.

Some previous studies regarding eating patterns have shown a relationship between daily patterns and obesity. Ma et al. [106] found that adults who skip breakfast were more likely to be obese, including those which have breakfast and dinner away from home. People that had a greater frequency of eating episodes were less likely to suffer from obesity. Triches et al. [170] also found that children that consumed breakfast less frequently had a stronger likelihood of obesity. Other works have found evidence supporting the idea that regular consumption of breakfast, lunch, and dinner contributes to healthy weight and better intake of nutrients [118, 155, 183, 29, 90]. Gillman [78]

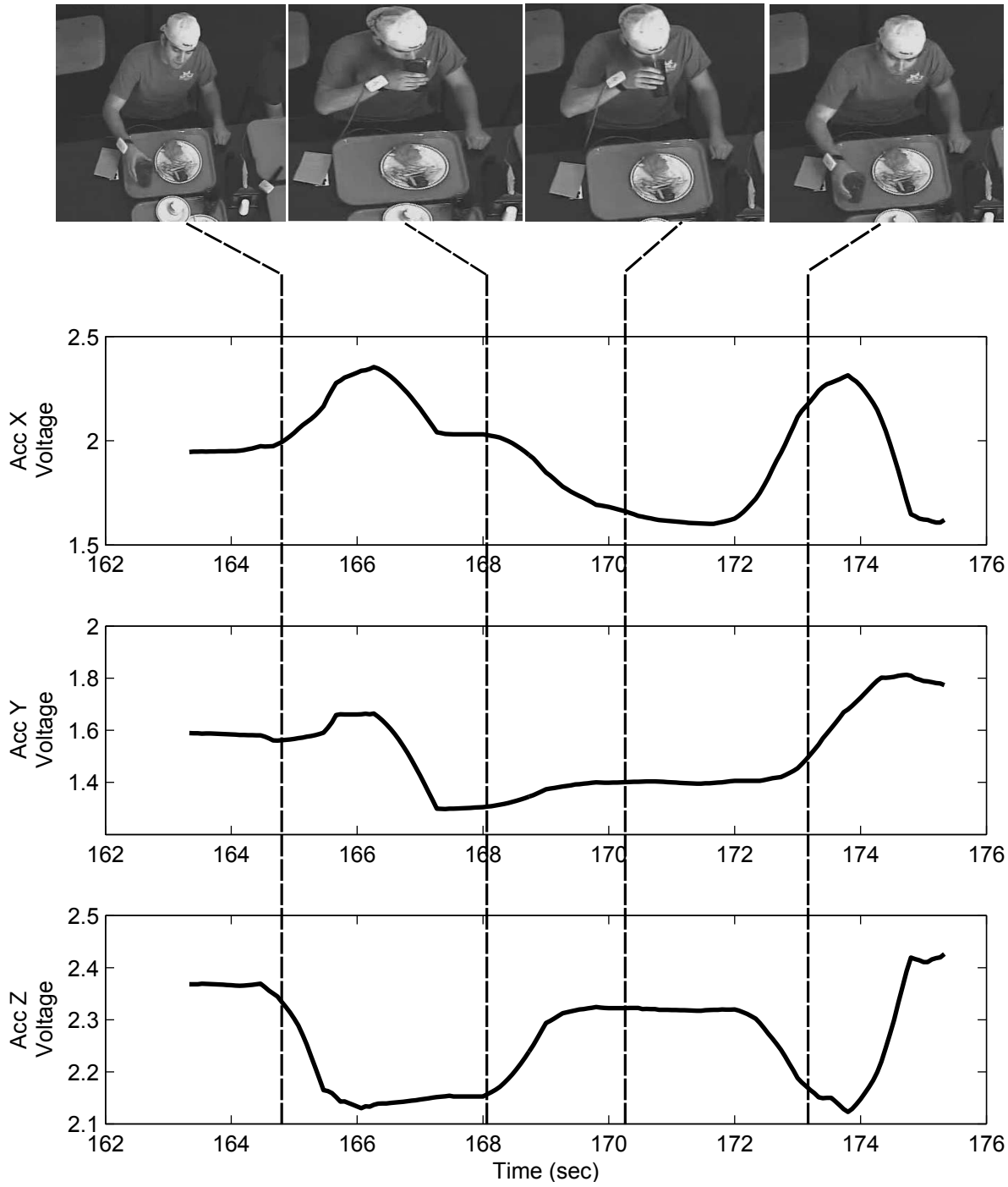


Figure 1.6: Accelerometer raw data.

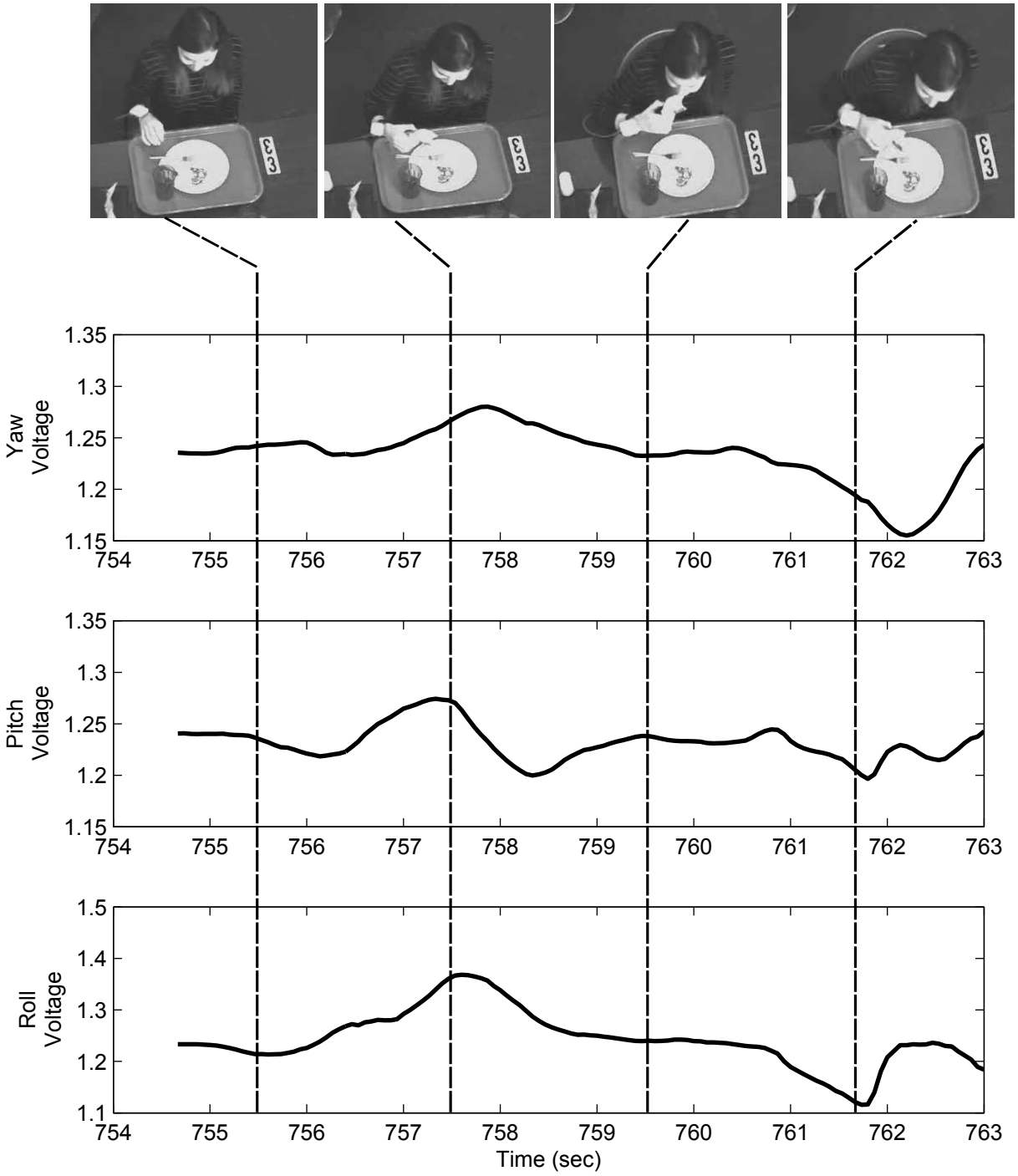


Figure 1.7: Gyroscope raw data.

observed that the amount of fruit and vegetables consumed was increased when young people had meals with their families often. Coon et al. [47] found that higher energy foods are consumed more frequently while watching television. Likewise, Putnam [135] observed that when people ate away from home, they were likely to consume more food or higher energy food (or both).

Researchers have also studied eating patterns at a meal level. Bowman [30] found that children ate more grains, fruits, and milk, when categorized as regular consumers of breakfast. Triches et al. [170], determined that low consumption of milk was significantly linked to obesity in children. Kant et al. [86] established that people which have the greatest variety of foods from all food groups have the most adequate nutrient intake. Some studies [65, 75, 163] have suggested that obese people tend to eat more rapidly, and that slowing the rate of consumption will help them eat less and be satisfied with less food. This idea that repid eating correlates to obesity has been challenged by the results of other studies. Kaplan [87] found that faster intake rates lead to more food consumption for both lean and obese individuals. Other studies have shown that increasing portion size also increases food intake for both lean and obese individuals [61, 113]. Wansink [31] studied the patterns of people with low and high BMIs in Chinese buffets. People with low BMIs tended to use smaller plates, browse the buffet more, leave more food on their plate, and place napkins on their lap. The individuals with high BMI tended to use bigger plates, immediately serve themselves, finish their portions, and place napkins on their lap less frequently.

Eating behaviors at the microstructure level has been studied. Spiegel [160] conducted a study which showed that bite size affected the extent to which food was chewed and the local rate of ingestion. According to the findings, subjects ate faster when they took larger bites, but slowing the ingestion rate by reducing bite size did not reduce the amount of food consumed. Kaplan [87] studied the effect of meal size, meal duration, bites per meal, and bites per minute in six lean and obese subjects, observing no significant difference in patterns among the two groups. Stellar [162] used a train gauge sensor implanted against the palate in a dental arch mounted on the back upper molars to find that chews and swallows can provide measures to the microstructure level of a meal, reflect hunger, and satiation. Bellisle [26] conducted a study of the effect of chewing and swallowing patterns based on the degree of acceptance of food in the mouth (palatability) using an oscillographic and video recordings. Ioakimidis et al. [81] collected data to describe chewing patterns using a magnetic jaw displacement detector. Frequency of chews varied significantly within the quartiles of chewing sequences, where more chews occurred in the first and fourth quartile.

The work in this dissertation focuses on eating patterns at the meal and microstructure level. Of particular interest is the idea that actions within a meal, or motions within an action, have some sequential dependence. The primary motivation for studying the sequential dependence is to improve automated recognition and measurement of the actions. However, these patterns could also be studied for any relation to BMI or other factors known to be related to obesity.

1.9 Hidden Markov Models

Markov processes and Hidden Markov models (HMM) are well-known probabilistic-based techniques used to model sequential information. They can be found in a wide variety of applications such as economics, biology, tracking systems, speech processing, image processing, and communication systems. Rabanier [136] and Bishop [28] provide thorough information on the fundamentals of general Markov processes and HMMs. Some of that information is reviewed here for background.

1.9.1 Markov models

A Markov model is a stochastic model which models temporal or sequential data, i.e., data that is ordered. The probabilistic framework of a Markov model provides a way to model the dependencies of current information (e.g. weather) with previous information, so each observation in the data sequence depends on previous elements in the sequence. For most applications the best prediction is based on the most recent data, rather then data in the distant past. Therefore, having a collection of data (or observations) $(x_1, ..., x_n)$ modeled by random variables $X_1, ..., X_n$, X_t could be dependent on X_{t-1}, X_{t-2}, ..., X_{t-m} for a fixed m. The simplest case is where $m = 1$, hence the prediction at a given time t is dependent on the information contained at time $t - 1$. This is the Markov property and is expressed by equation 1.6.

$$p(x_t|x_1, x_2, x_3, .., x_{t-1}) = p(x_t|x_{t-1}) \tag{1.6}$$

A Markov chain is a type of Markov model. This is governed by the Markov property, but it is more restrictive because the observations can only take on a finite set values s_j, for $j = 1, ..., N$, referred to as states. Thus, in a Markov chain, knowing the observation is equivalent of knowing the state as well. There exists probabilities that explain the relationship between states, i.e. the

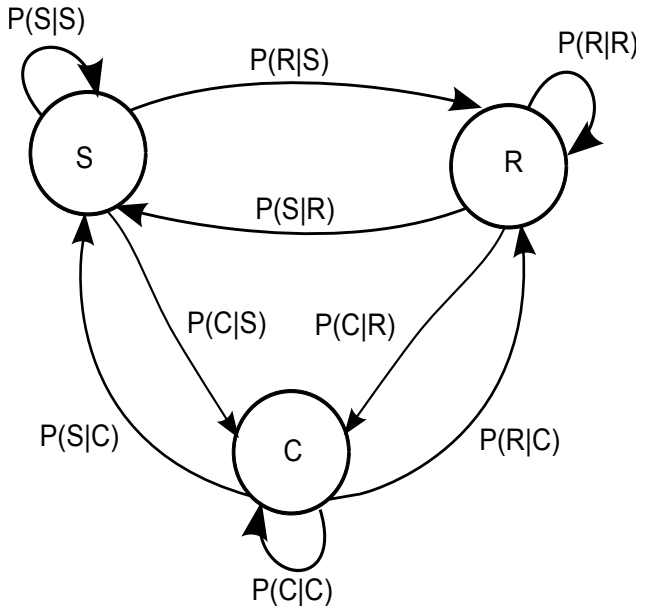


Figure 1.8: Markov chain weather model (C=cloudy, R=rainy, S=sunny).

probability of transitioning from one state to a different state or remaining in the same state. These transition probabilities can be calculated from training data. Figure 1.8 shows an example of a Markov chain used to predict weather based on past information. Three possible states are allowed in this model; sunny, rainy, and cloudy (S, R and C). Predicting the next day's weather can be accomplished through use of the Markov property, $p(s_i|s_{i-1})$.

The main disadvantage of the Markov Chain is that for many practical problems, it is not possible to perfectly observe the true state of the system. For example, in a GPS tracking system, the data is corrupted by noise, which might cause variation in the representation of the true position. Because systems contain this hidden information, they must be modeled differently to accurately predict information. The systems can be modeled in two parts; the hidden and the observed part. The hidden portion is modeled using hidden (latent) variables. This leads to the definition of hidden Markov models.

1.9.2 Hidden Markov models

An HMM is a stochastic model in which the states of the model are hidden and each state emits an output which can be observed. In an HMM the outputs of a state are not deterministic (as in the case of a Markov chain). Rather, the observations are associated with a probabilistic function

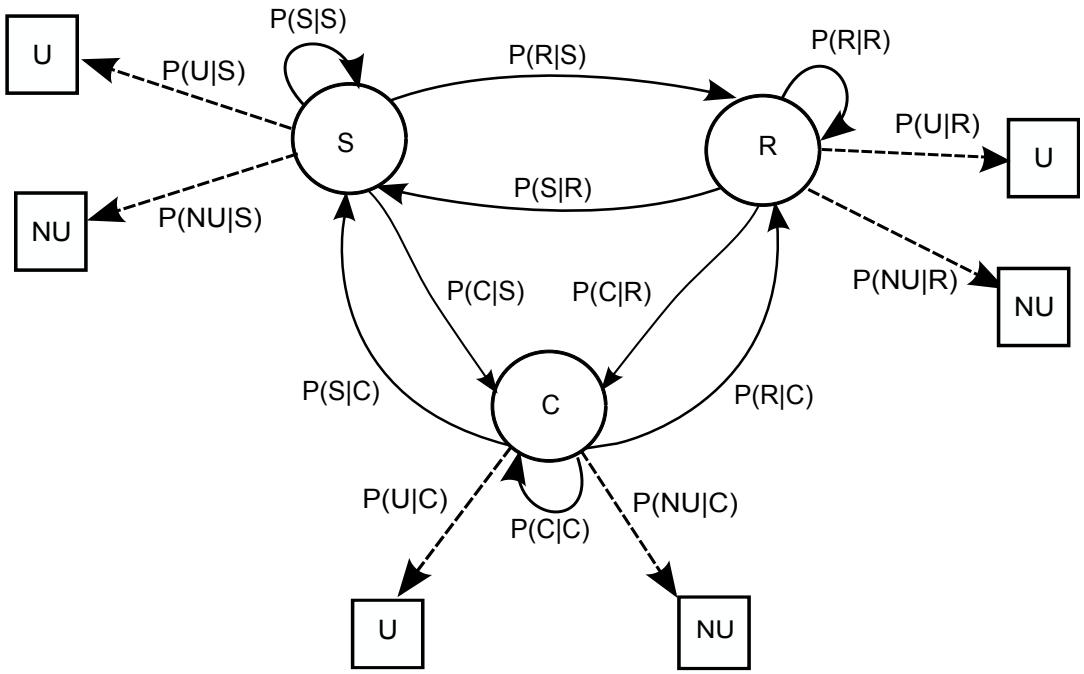


Figure 1.9: HMM weather model (C=cloudy, R=rainy, S=sunny; U=umbrella, NU=no umbrella).

given a state s_j. Extending the last Markov chain example, an example HMM can be constructed by assuming that we do not have the ability to view the state of the weather and predictions about the weather are based on observing whether a subject is carrying an umbrella or not. Figure 1.9 shows the HMM for the weather model, where given the state of the weather, there is a probability function associated with observing an umbrella (U) or not (NU).

Concretely, in an HMM there exists an observation sequence $O = \{o_1, o_2, ..., o_T\}$ and a state sequence $Q = \{q_1, q_2, ..., q_T\}$. An observation o_t can be discrete or continuous and an element of the state sequence q_t represents any of the N possible states, i.e. $q_t = s_j$. The parameters that govern an HMM are derived by computing the probability of the state sequence given a set of observations. Baye's theorem can be used to describe this in terms of its likelihood function, prior probabilities, and normalization factor as shown in equation 1.7.

$$p(q_1, q_2, ..., q_T|o_1, o_2, ..., o_T) = \frac{p(o_1, ..., o_T|q_1, ..., q_T)p(q_1, ..., q_T)}{p(o_1, .., o_T)} \tag{1.7}$$

Using the Markov property and the assumption that observations are independent from each other,

the terms in the numerator can be rewritten as in equations 1.8 and 1.9.

$$p(q_1, ..., q_T) = \prod_{t=1}^{T} p(q_t|q_{t-1}) \tag{1.8}$$

$$p(o_1, ..., o_T|q_1, ..., q_T) = \prod_{t=1}^{T} p(o_i|q_t) \tag{1.9}$$

Then equation 1.7 is proportional to the form presented in equation 1.10.

$$p(q_1, q_2, ..., q_T|o_1, o_2, ..., o_T) \propto \prod_{i=1}^{T} p(q_t|q_{t-1}) \prod_{i=1}^{T} p(o_t|q_t) \tag{1.10}$$

The last equation provides a set of conditional probabilities which mark the basis for the parameters used in a HMM: transition probabilities, emission (observation) probabilities, and initial state probabilities.

1.9.2.1 HMM parameters

The common notation for an HMM is given by $\lambda = \{A, B, \pi\}$.

1. Initial state distribution $\pi = \{\pi_j\}$ describes the probabilities of starting the observation sequence in state s_j. Equation 1.11 defines π_j. The sum of these should add to 1 and $\pi_j \geq 0$.

$$\pi_j = p(q_1 = s_j), \quad j = 1, .., N \tag{1.11}$$

2. State transition probability distribution $A = \{a_{ij}\}$ is a matrix containing the probability of transitioning between states. The notation a_{ij} represents a transition from state s_i to state s_j and is defined by equation 1.12. In a HMM, $a_{ij} \geq 0$ and all transitions probabilities from state s_j should add to 1, i.e. $\sum_{j=1}^{N} a_{ij} = 1$.

$$a_{ij} = p(q_t = s_i|q_{t-1} = s_j), \quad i, j = 1, ..., N \tag{1.12}$$

3. Emission probability distribution $B = b_j(o_t)$ describes the probabilistic function of the observations given state s_j, This function is shown by equation 1.13 and $b_j(o_t) \geq 0$. This probability

function is described by the observations (discrete or continuous).

$$b_j(o_t) = p(o_t|q_t = s_j), \quad j = 1, ..., N \tag{1.13}$$

If the observations are discrete, the problem can be simplified by constraining an observation o_t to a finite set of values $V = \{v_1, .., v_W\}$, shown by equation 1.14. A formal way to develop discrete outputs is by using vector quantization methods to build code books. Discrete probability functions govern the probability of an observation being observed at time t for a given state. For example, a die has 6 sides with outcomes 1 trough 6, while another die has 12 sides with seven sides having an outcome of 1 and just one side for each outcome 2 through 6. In this example, each die represents a different state in the experiment and the probability mass function is different for the possible outcomes given the state.

$$b_j(v_k) = p(o_t = v_k|q_t = s_j), \quad k = 1, ..., W \tag{1.14}$$

A continuous representation of the observations has the advantage of better capturing the underlying statistical model. Here, $b_j(o_t)$ takes on the form of a probability density function (pdf). The most common pdf used for a HMM is the Gaussian density. Equations 1.15 and 1.16 show the single and multivariate Gaussian distribution, respectively.

$$b_j(o_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(o_t - \mu_j)^2}{2\sigma_j^2}\right) \tag{1.15}$$

$$b_j(o_t) = \frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(o_t - \mu_j)'\Sigma_j^{-1}(o_t - \mu_j)\right) \tag{1.16}$$

In general, mixture of Gaussian densities are used to model continuous observations. This means that M number of Gaussians collectively model the density. Equation 1.17 describes the probability density function $b_j(o_t)$ based on M Gaussians; where, c_m is a weighting value with $\sum_{m=1}^{M} c_m = 1$ and $b_{jm}(o_t) = \mathcal{N}(o_t; \mu_{jm}, \Sigma_{jm})$. The parameters of the Gaussians can be computed using maximum likelihood (ML) estimation through the expectation-maximization (EM) algorithm. More details on

Gaussian mixture models (GMM) can be found in Reynolds et al. [139] and Bishop [28].

$$b_j(o_t) = \sum_{m=1}^{M} c_m b_{jm}(o_t) \tag{1.17}$$

1.9.2.2 Fundamental problems in HMMs

HMMs can be used in a number of ways to solve problems. The following briefly outlines the most common types of problems for which HMMs are used.

Observation sequence evaluation

One of the primary tasks when using a HMM $\lambda = \{A, B, \pi\}$ is evaluating a sequence of observations $O = \{o_1, ..., o_T\}$ or $p(O|\lambda)$ to determine how well a model predicts a given observation sequence O. Thus, if there are L HMMs each must be evaluated in order to choose the most appropriate model λ_l, for $l = 1, ..., L$.

An observation sequence O depends on the state sequence $Q = \{q_1, ..., q_T\}$ of an HMM λ_l. Therefore, the probability of a state sequence generating an observation sequence can be written as shown in equation 1.18.

$$p(O|Q, \lambda_l) = \prod_{t=1}^{T} p(o_t|q_t, \lambda_l) = b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdot ... \cdot b_{q_T}(o_T) \tag{1.18}$$

Depending on the model λ_l, the probability of a state sequence is expressed in equation 1.19.

$$p(Q|\lambda_l) = p(q_1|\lambda_l) \prod_{t=2}^{T} p(q_t|q_{t-1}, \lambda_l) = \pi_{q1} \cdot a_{q_1 q_2} \cdot a_{q_2 q_3} \cdot ... \cdot a_{q_{T-1} q_T} \tag{1.19}$$

Finally, the evaluation of an observation sequence has the form shown in equation 1.20.

$$p(O|\lambda_l) = \sum_{Q} p(O|Q, \lambda_l) p(Q|\lambda_l) = \sum_{q_1, ..., q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) ... a_{q_{T-1} q_T} b_{q_T}(o_T) \tag{1.20}$$

This approach has a drawback in that it must consider all possible state sequences for the observation sequence evaluation. The Forward-Backward algorithm is typically used to overcome this problem by acknowledging that there are redundancies in the calculations. The algorithm consists of two similar recursive algorithms which compute $p(O|\lambda_l)$. In the forward case, we have a

parameter α which represents the probability of the partial observation sequence $o_1, .., o_t$ for state s_j at time t as shown in equation 1.21.

$$\alpha_t(j) = p(o_1, .., o_t, q_t = s_j|\lambda_l) \tag{1.21}$$

The algorithm is described as follows:

1. Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

2. Recursion:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1}), \quad 1 \leq i, j \leq N, \quad 1 \leq t \leq T$$

3. Termination:

$$p(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

A graphical representation of the forward algorithm using only two states is shown in figure 1.10. The nodes represent states, the arrows are called arcs and represent the product of the transition probability with the emission probability, and incoming arcs to a node represents summation. At $t = 1$ the algorithm initializes by computing α_1 for all states represented as nodes in the second column. Then, α_1 gets propagated and multiplied by the value of the probabilities in the arcs to form α_2. Subsequently, all the values of α are computed by propagating previous values, until reaching the last observation at time T.

The backward algorithm calculates the probability $p(O|\lambda)$ by starting at the end of the observation sequence at time T and moving in reverse towards $t = 1$. The parameter β calculates the partial probability of observations $o_{t+1}, .., o_T$ for state s_j at time $t+1$ as shown in equation 1.22.

$$\beta_{t+1}(j) = p(o_{t+1}, .., o_T, q_t = s_j|\lambda_l) \tag{1.22}$$

The algorithm is described as follows:

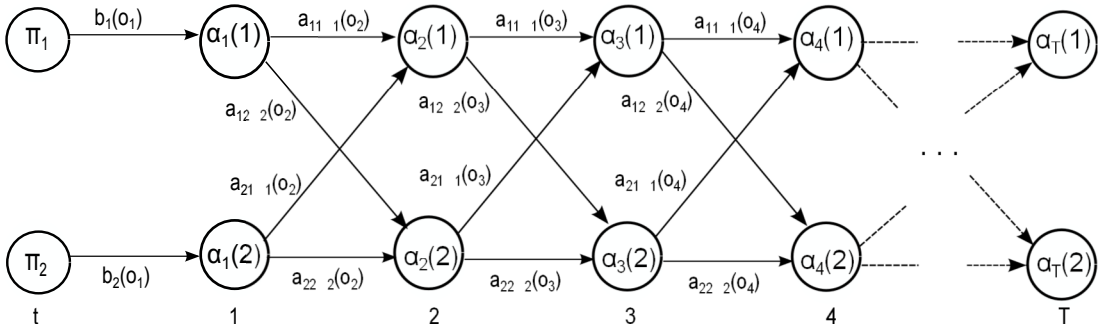


Figure 1.10: Forward algorithm.

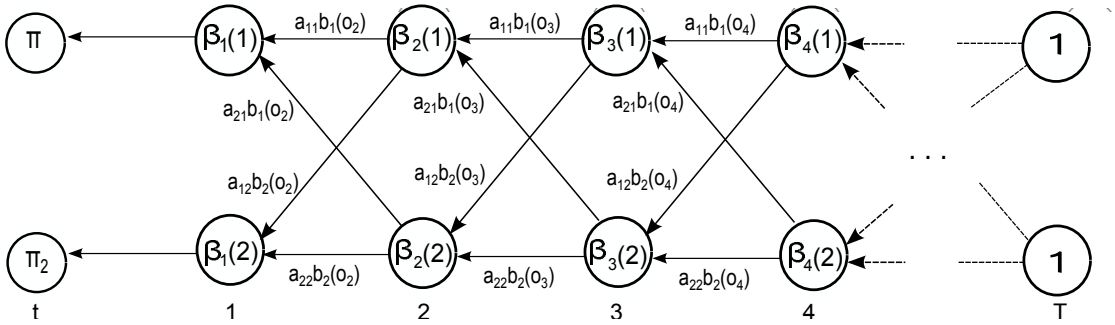


Figure 1.11: Backward algorithm.

1. Initialization:

$$\beta_T(j) = 1,\ t = T,\ \ 1 \leq i, j \leq N$$

2. Recursion:

$$\beta_t(j) = \left[\sum_{i=1}^{N} \beta_{t+1}(i) a_{ij}\right] b_j(o_{t+1}),\ \ 1 \leq j \leq N,\ 1 \leq t < T$$

3. Termination:

$$p(O|\lambda) = \sum_{i=1}^{N} \pi_i \beta_1(i)$$

A similar graphical representation for two states for the backward algorithm can be found in figure 1.11.

State sequence decoding

The objective of decoding is to discover the hidden state sequence that most likely describes a given observation sequence. One solution to this problem is to use the Viterbi algorithm, which finds the single best state sequence for a given observation sequence. The Viterbi algorithm is a trellis algorithm which is very similar to the forward algorithm, except that the transition probabilities are maximized at each step rather than summed. A parameter δ is used as the probability of the most probable state path for the partial observation sequence as shown in equation 1.23.

$$\delta_t(i) = \max_{q_1,q_2,\ldots,q_{t-1}} p(q_1 q_2 ... q_t = s_i, o_1, o_2, ..., o_t|\lambda) \tag{1.23}$$

The Viterbi algorithm is defined as follows:

1. Initialization

$$\delta_1(i) = \pi_i b_i(o_1),\ 1 \leq i \leq N$$

$$\psi_1(i) = 0$$

2. Recursion:

$$\delta_t(j) = \max_{1\leq i\leq N} \left[\delta_{t-1}(i)a_{ij}\right] b_j(o_t), \quad 2 \leq t \leq T,\ 1 \leq j \leq N$$

$$\psi_t(j) = arg \max_{1\leq i\leq N} \left[\delta_{t-1}(i)a_{ij}\right], 2 \leq t \leq T,\ 1 \leq j \leq N$$

3. Termination:

$$P^* = \max_{1\leq i\leq N} \left[\delta_T(i)\right]$$

$$q_T^* = arg \max_{1\leq i\leq N} \left[\delta_T(i)\right]$$

4. Optimal state sequence backtracking:

$$q_t^* = \psi_{t+1}\left(q_{t+1}^*\right), \quad t = T-1, T-2, ..., 1$$

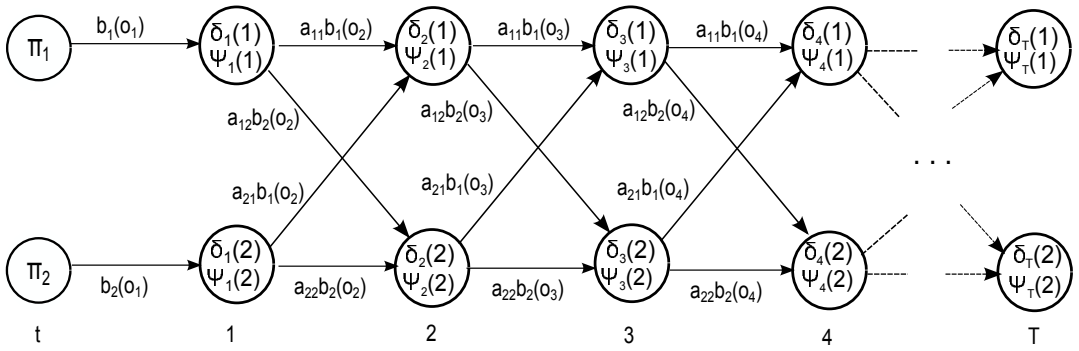


Figure 1.12: Viterbi forward step.

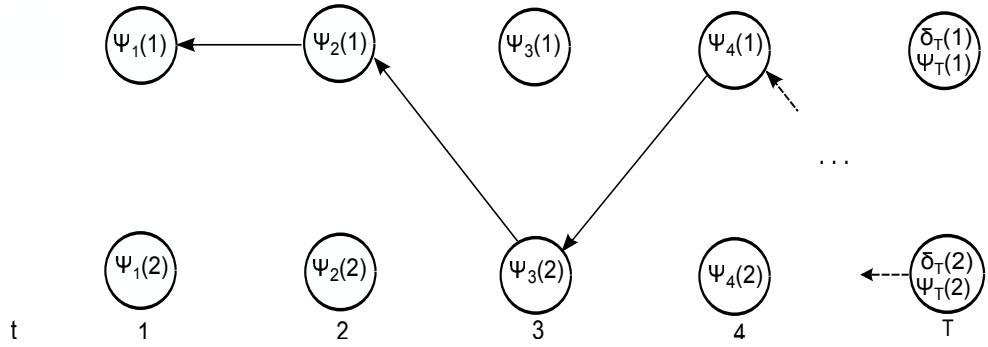


Figure 1.13: Viterbi backward step.

Backtracking allows the best state sequence to be found from the back pointers stored in the recursion step, but it should be noted that there is no easy way to find the second best state sequence. Figures 1.12 and 1.13 shows the two-state example showing the forward and backtracking steps, respectively.

Parameter estimation

Parameter estimation is the process used to train an HMM given a set of observations. This process results in an estimation of the parameters π, A, and B. This problem is difficult in part because the parameters cannot be estimated directly. Instead, by using the data, the problem is reformulated to find a set of parameters that maximizes the probability of the observation sequence as in equation 1.24.

$$arg\max_{\lambda} p(O|\lambda) \tag{1.24}$$

This can be solved by using the maximum likelihood estimation (MLE) technique. Concretely, the Baum-Welch algorithm is a special case expectation-maximization (EM) method which

iteratively improves the likelihood of $p(O|\lambda)$. This algorithm re-estimates the parameters of the model $\hat{\lambda}$ by using the EM algorithm and finding local maxima as seen in equation 1.25.

$$p(O|\hat{\lambda}) \geq p(O|\lambda) \tag{1.25}$$

An intermediate step is defined to describe the reestimation of the HMM parameters. This parameter (equation 1.26) is $\epsilon_t(i,j)$ which is defined as the probability of being in state s_i at time t and state s_j at time $t+1$ given a model and the observation sequence.

$$\epsilon_t(i,j) = p(q_t = s_i, q_{t+1} = s_j|O,\lambda) \tag{1.26}$$

Using the forward and backward variables, equation 1.26 can be rewritten as shown in equation 1.27. The denominator defines $p(O|\lambda)$.

$$\epsilon_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{i=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)} \tag{1.27}$$

The probability of being in state s_j at time t for a given set of the observation (O) is defined as $\gamma_t(i)$ as shown in equation 1.28.

$$\gamma_t(i) = \sum_{j=1}^{N}\epsilon_t(i,j) \tag{1.28}$$

A method for reestimating the parameters of the HMM can be accomplished by using equations 1.27 and 1.28. This method can be seen as counting event occurrences. Equation 1.29 records the number of times state s_j occurs at time $t = 1$. Equation 1.30 records the ratio between the number of times a transition from state s_i to state s_j happens and the expected number of total transitions from state s_i. Equation 1.31 is the ratio of the expected number of occurrences of state s_j when observing symbol v_k to the total number of occurrences in state s_j.

$$\bar{\pi}_i = \gamma_1(i) \tag{1.29}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \epsilon_t(i,i)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{1.30}$$

$$\bar{b}_j(v_k) = \frac{\sum_{\substack{t=1 \\ s.t. o_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{1.31}$$

1.10 Novelty

The primary contribution of this work is the study of the sequential predictiveness of activities related to eating to improve the automatic recognition of unrestricted eating gestures. Some previous works have studied the recognition of individual gestures related to eating [17, 85], but to our knowledge we are the first to propose studying the sequential dependencies across a larger scale of time.

In chapter 2, we undertook the task of identifying the gestures related to eating activities during the consumption of a meal. An inter-rater reliability test was conducted to determined the stability of the definitions of the proposed gestures.

In chapter 3, the sequential dependency of eating gestures is studied. Here, higher-order HMMs are converted into equivalent first-order to model several levels of history from gestures sequences, i.e. that previous knowledge of eating gestures can help predict the upcoming gesture.

Chapter 4 considers the problem of automatically classifying and segmenting wrist motion data for the purposes of recognizing eating gestures. These ideas again explore the utility of sequential dependencies, in this case to guide the segmentation process.

Although this work is primarily concerned with using the idea of sequential predictiveness to improve the recognition and measurement of EI, it might also prove useful in studying the characteristics of people with respect to eating habits. It may be that the methods developed herein could prove useful in identifying behaviors associated with increased EI or other criteria associated with obesity.

Chapter 2

Definition and Assessment of Wrist Motions Related to Eating Activities

2.1 Introduction

The motivating goal of this work is to measure eating activities by tracking wrist motion. To achieve this goal, a set of gestures are defined that comprise common eating activities. A large number of specific activities are possible, such as cutting solid food with an utensil, extending and contracting the arm for liquid intake, grabbing a napkin to wipe the mouth (or hands), stirring with a spoon, etc. However, because the goal is automated recognition, it is desirable to seek as small a set as possible. In this chapter, a study was undertaken to determine which gestures are suitable to describe eating activities throughout a variety of subjects. By visually inspecting the participants while eating, a set of four basic activities was found to appear frequently during the consumption of a meal: rest, utensiling, bite, and drink. The gestures for these activities were defined based on the intention of the eater at the moment of wearing the wrist tracker device. Using these definitions, three meals were manually labeled by five human raters. The data provided by the raters served to explore the stability of the gesture definitions through an inter-observer reliability evaluation. The results indicated that these gestures were consistently recognizable by human raters and thus could

serve as a base language for the study of automated recognition.

2.2 Methods

2.2.1 Data

The data used for this study was recorded in a large data collection supported by the National Institutes of Health via grant 1R41DK091141-A1. The Harcombe Dining Hall of Clemson University was the location for data collection. The facility seats up to 800 guests and provides a wide range of foods and beverages, allowing people to build their own meal. For example, some of the choices include omelets, sandwiches, pizza, pasta, fruits and vegetables, meat cuts, deserts, juices, milk, sodas, teas and coffee. The foods are served in a wide variety of containers, including plates, bowls, wraps, pouches, trays, cartons, cups, and glasses, and they are consumed using a variety of utensils including forks, knives, spoons, and fingers.

Inside the dining hall, an instrumented table was prepared to record data simultaneously from four participants. Four digital cameras in the ceiling (approximately 5 meters height) were used to record the participant's mouth, torso, and tray while meal consumption. Also, a custom wrist-worn device was used to record the motion in the participant's wrist during his or her meal consumption. Wrist motion was sampled at a frequency of 15 Hz. Each participant had his device wired to a laptop where data was stored. A scale was located under the subject's tray to monitor food weight while eating. Figure 2.1 shows a picture of the instrumented table.

Participants wore the device on their dominant hand. There were no restrictions on foods or eating style during data collection. Participants were allowed to eat as normally as possible, including natural movements unrelated to eating (e.g. conversations with people, using a phone, gesturing, scratching, etc.). Subjects were free to build their own meal, as well as drink any type of beverage using the container of their choice. Figure 2.2 shows an image of four subjects eating at the instrumented table.

The full data set is composed of 276 subjects, including 131 males and 145 females subjects ranging age from 18 to 75. Body mass index (BMI) ranged from 17.4 kg/m^2 to 46.2 kg/m^2. Ethnicity is predominantly Caucasian (192), but also includes African-American (27), American Indian or Alaska Native (2), Asian or Pacific Islander (29), Hispanic (11), and Other (15). The data has been used to evaluate the accuracy of the original bite counting method developed by our group

Figure 2.1: Instrumented table.

Figure 2.2: Eaters on instrumented table.

Table 2.1: Participant's information.

Participant	Sex	Age (years)	BMI (kg/m^2)	Ethnicity
1	Female	21	27.5	Caucasian
2	Male	21	22.6	Caucasian
3	Male	20	21.7	Caucasian
4	Female	20	20.7	Caucasian
5	Male	18	21.3	Caucasian
6	Female	29	27.9	Caucasian
7	Female	24	19.2	Caucasian
8	Female	23	24.6	Caucasian
9	Male	43	24.6	Caucasian
10	Male	50	22.3	Caucasian
11	Female	20	20.4	Caucasian
12	Female	57	27.4	African-American
13	Female	29	23.1	Caucasian
14	Female	40	27.3	Caucasian
15	Male	21	28.5	Caucasian
16	Female	23	25.7	Hispanic
17	Male	20	21.6	Caucasian
18	Female	43	24.9	Caucasian
19	Female	20	23.3	Caucasian
20	Male	26	35.1	Caucasian
21	Male	54	20.1	Caucasian
22	Female	26	26	Caucasian
23	Male	23	28.6	Caucasian
24	Male	26	25.2	Caucasian
25	Male	28	28.3	Caucasian

[**?**], and to compare a measure of EI derived from automated bite count against self estimates [**?**]. For the work in this dissertation, a subset of 25 meals were chosen randomly from the full dataset. This smaller sample was used due to the amount of time needed to manually review the videos of the eaters and hand-label all eating activities. The selected data set consisted of 11 male and 9 female subjects. Table 2.1 shows sex, age, BMI, and ethnicity information for the 20 participants. Also, table 2.2 describes the meal information, including the duration, food type, drinks, utensils, and containers used in the meal for each of the participants.

2.2.1.1 Preprocessing

Accelerometer data (AccX, AccY, AccZ) and gyroscope data (Yaw, Pitch, Roll) were smoothed using a Gaussian-weighted window defined by equation 2.1. Here R_t is the raw data and S_t is the

Table 2.2: Food list per subject.

Subj.	Time (min)	Food	Drink	Utensil	Containers
1	12.8	Stir fry vegetables	Water	Hand, fork	Mug, plate
2	7.7	Chicken sandwich, french fries	Tea	Hand	Mug, plate
3	16.1	Refried beans, shoestring french fries, popcorn chicken, taco	Tea	Hand, fork	Mug, plate
4	23.8	Popcorn chicken, shoestring french fries, banana	Water	Hand, fork	Glass, plate
5	8.8	Stew beef, seasoned dry limas, steamed california blend vegetables, white rice	Water	Hand, fork	Glass, plate
6	16.8	African spiced sweet potato, blackened tilapia, seasoned corn, sauteed tomatoes and zucchini	Sweet tea	Hand, fork	Mug, plate
7	20.3	Bread, salad, wild rice	Sprite zero	Hand, fork	Mug, plate
8	10.2	Ice cream, cupcake, custom fruit bowl	Water	Hand, spoon	Glass, plate, bowl
9	15.4	Eggplant and broccoli pizza, hamburger, pepperoni pizza, shoestring french fries	Sweet tea	Hand, spoon	Glass, plate
10	14.6	Black beans and rice, sauteed pollock	Kiwi juice	Hand, fork	Glass, plate
11	11.7	Garlic bread sticks, grilled italian sausage with onions and peppers, rotini with marinara	Water	Hand, fork	Glass, plate
12	25.2	Cupcake, veggie indian currry, brownie	Lemonade	Hans, fork	Glass, plate
13	25.1	Eggplant parmesan, salad bar, sweetzza cinnamon pecan, sweetzza apple	Unsweet tea	Hand, fork	Glass, plate
14	30.9	Portobello sandwich, salad	Sweet tea	Hand, fork	Glass, plate, bowl, mug
15	4.3	Pasta tour of italy	Powerade	Hand, fork	Glass, plate
16	17.6	Buffalo tenders, salad bar	Cranberry juice water mix	Hand, fork	Glass, plate
17	14.2	BBQ brisket and kaiser roll, cereal corn pops, hash sweet potato and bacon	Apple juice and water	Hand, fork, spoon	Glass, plate, bowl
18	21.0	Peanut butter chocolate fudge, salad bar, spinach and cheese quiche	Water	Hand, fork	Glass, plate, bowl
19	13.8	Homestyle chicken sandwich, salad bar	Sweet tea	Hand, fork	Glass, plate, bowl
20	6.6	Fish, mac and cheese	Sweet tea		
21	41.3	Pasta tour of italy, salad bar, bread, blueberry crobbler, brownie	Sweet tea	Hand, fork	Glass, plate, bowl
22	18.0	Pepperoni pizza, spice pork and vegetable, sweetzza chocolate peanut butter	Diet coke	Hand, fork	Glass, plate
23	10.0	Bread sticks, desert pizza, ziti	Mellow yellow	Hand, fork	Glass, plate
24	18.9	Asian vegetables, salad, wasabi potatoes	Coke	Hand, fork	Glass, plate
25	13.9	Shoestring french fries, homestyle chicken sandwich, salad bar	Diet coke	Hand	Glass, plate

smoothed data at time t. For the sensors we used, the best results were obtained using $N = 15$ (1 second) and with $\sigma^2 = 10$. Figure 2.3 shows and example of the six signals after smoothing.

$$S_t = \sum_{i=-N}^{0} R_{t+i} \frac{\exp\left(\frac{-t^2}{2\sigma^2}\right)}{\sum_{x=0}^{N} \exp\left(\frac{-(x-N)^2}{2\sigma^2}\right)} \tag{2.1}$$

2.2.2 Eating gestures

One significant problem when performing eating activity recognition is defining a set of "gestures" that describe the individual eating activities. We crated our definitions based on the concept of discernible user intent. The subject's intent is determined by observing the hand wearing the device. The duration of an action lasts from when the intent can first be observed, until that intent has ended. Our group engaged 8 people in repeatedly observing videos and manually segmenting and labeling them with different sets of gesture names over the course of several months. During this period of time, some gestures were combined and others were considered but removed to their limited occurrences. The criteria for defining a gesture was refined several times with the goal of being as objective and repeatable as possible. Ultimately, the group settled on four gestures related to eating: *rest*, *utensiling*, *bite*, and *drink*. All other gestures for which intent was not defined, including both eating and non-eating activities (e.g. gesturing while talking, cleaning with a napkin, waving at a friend, etc.), are referred to as *other*. We developed a definition for each word consisting of four parts: a) the description of the activity, b) the start time of the activity, c) the end time of the activity, and d) particular events that should be included or excluded from the word label. Table 2.3 provides these details for our four words.

2.2.2.1 Ground truth

A custom tool was developed for reviewing the recorded data to label segments with the gesture definitions. The tool was coded using Microsoft Visual C. Video and sensor information were synchronized and displayed as shown in figure 2.4. Time navigation was performed using the keyboard to move forwards, backwards, play and pause. Labeling was done manually by looking at the intent of the instrumented hand of the eater in the video. A word could be labeled by enclosing it within a box using specific keys in the keyboard. A completed label is marked by a colored box

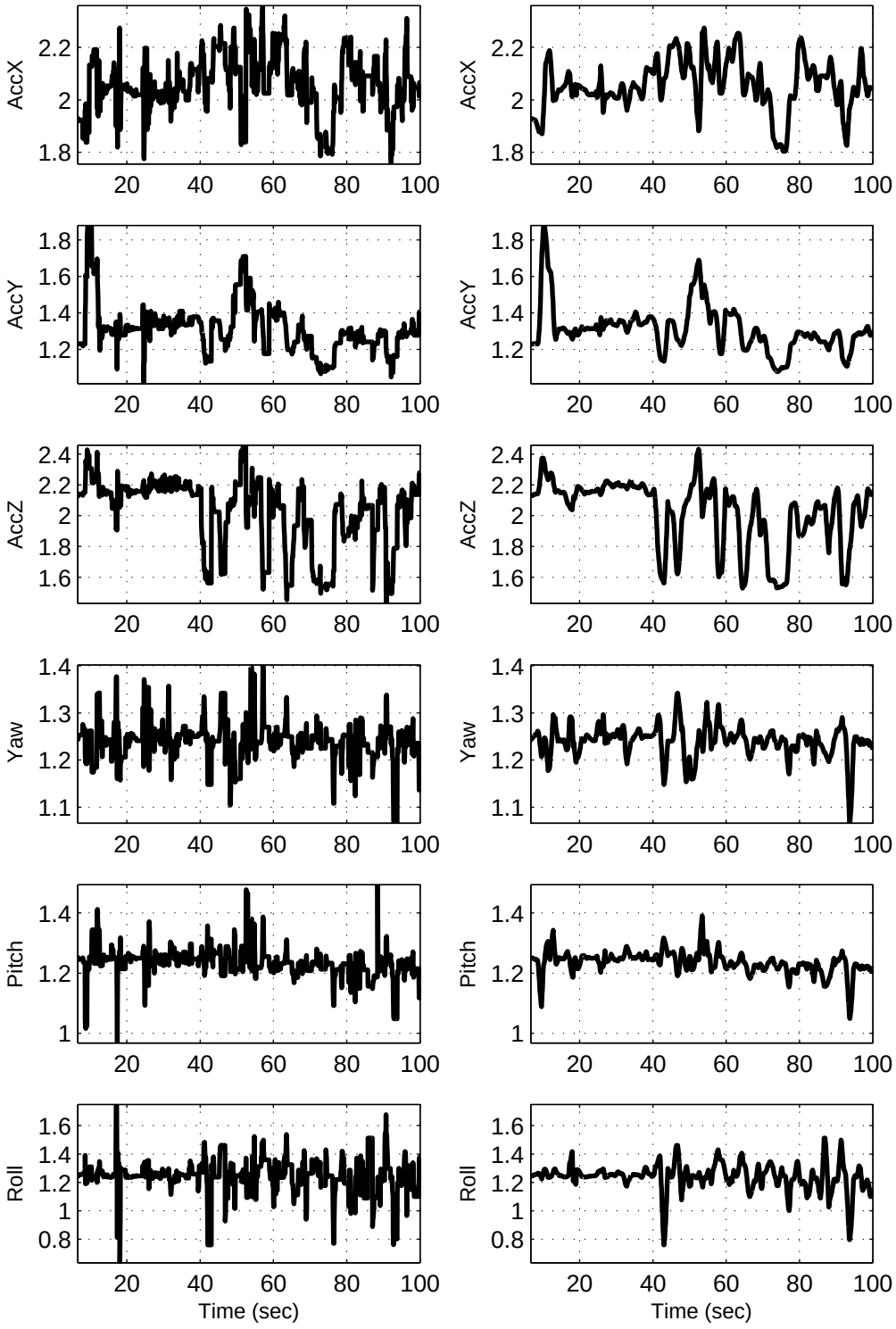


Figure 2.3: Raw (left column) and smooth (right column) data.

Word	Details
Bite	(a) The subject puts food into their mouth. (b) Begins when a hand or utensil starts moving towards the mouth. (c) Ends when the hand or utensil finishes moving away from the mouth (d) If a bite is interrupted, the word label should start after the interruption, when motion towards the mouth resumes. Bites need not begin and end at the plate. Motion towards and away from the mouth should define the boundaries; with food consumption taking place in between.
Rest	(a) The subjects dominant hand has little or no motion. (b) Begins when subjects hand stop moving. (c) Ends when subjects hand begins moving again (and moves for at least one second) (d) A rest may include brief periods of motion (less than one second) such as posture adjustments. Rests should not include motions with other intent, whether eating related or not. (ex: gesturing while talking, cleaning with a napkin, or waving to a friend).
Utensiling	(a) The subject uses a utensil or their hand to manipulate, stir, mix or prepare food(s) for consumption. (b) Starts when utensil or hand moves towards food with intent to manipulate. (c) Ends when manipulating has finished. (d) This includes moving food around the plate, dipping foods in sauces, cutting foods, etc.
Drink	(a) The subject puts beverage into their mouth. (b) Starts when a hand begins moving a beverage towards the mouth. (c) Ends when the hand has finished moving away from the mouth. (d) Each individual sip should be a different drink (if multiple sips are taken).

Table 2.3: Eating gesture definitions.

and a legend used to discriminate between labels as shown at the bottom of figure 2.4. This figure served as a guide for raters and with the set of instructions listed below (numbers correspond to indicated positions in figure 2.4):

1. This value indicates which index you are currently watching (which frame of the video is displayed).

2. This box allows you to jump to any previously labeled gesture by clicking on the word and index. If you click on a label you must click over the signal plot to keep on labeling.

3. This is an indicator telling you which gesture you most recently labeled.

4. These windows provide instructions for the tools usage.

5. This feature allows you to jump to an index by typing the number of the index you would like to move to and clicking the Jump button.

6. This is an example of the tool's usage: from the instructions, you can see that bite labels can be created using Q. When you enter a Q for the first time, you begin creating your bite label. The label begins where your cursor currently is located and the outline of a box is drawn to indicate the minimum amount of time required for the word action to take place. When this is created, you now must complete your label by moving the video forward (using the video controls) and entering the same label button (in this case, a Q for a bite) to complete labeling the word. When this is done, the box will be filled in like the bite labels shown before and after the box circled above. Gestures should be labeled according to the word definitions provided. Note: if you want to label is shorter than the minimum box you can ignore it.

7. If you had mistakenly chosen an incorrect gesture and you have not finish labeling, you can press ESC button to erase the current word selection.

A minimum time requirement for each gesture was determined and displayed for the rater to help guide the label-making process. One meal file was randomly selected and five raters were asked to label the complete meal using the definitions in table 2.3. Later, matched labels were compared for rest, utensiling, bite, and drink gestures. Table 2.4 shows statistics for the gestures obtained from the five raters. Finally, it was determined that a bite required at least 1 second to perform and utensiling, drink, and rest each required a minimum of 2 seconds to perform.

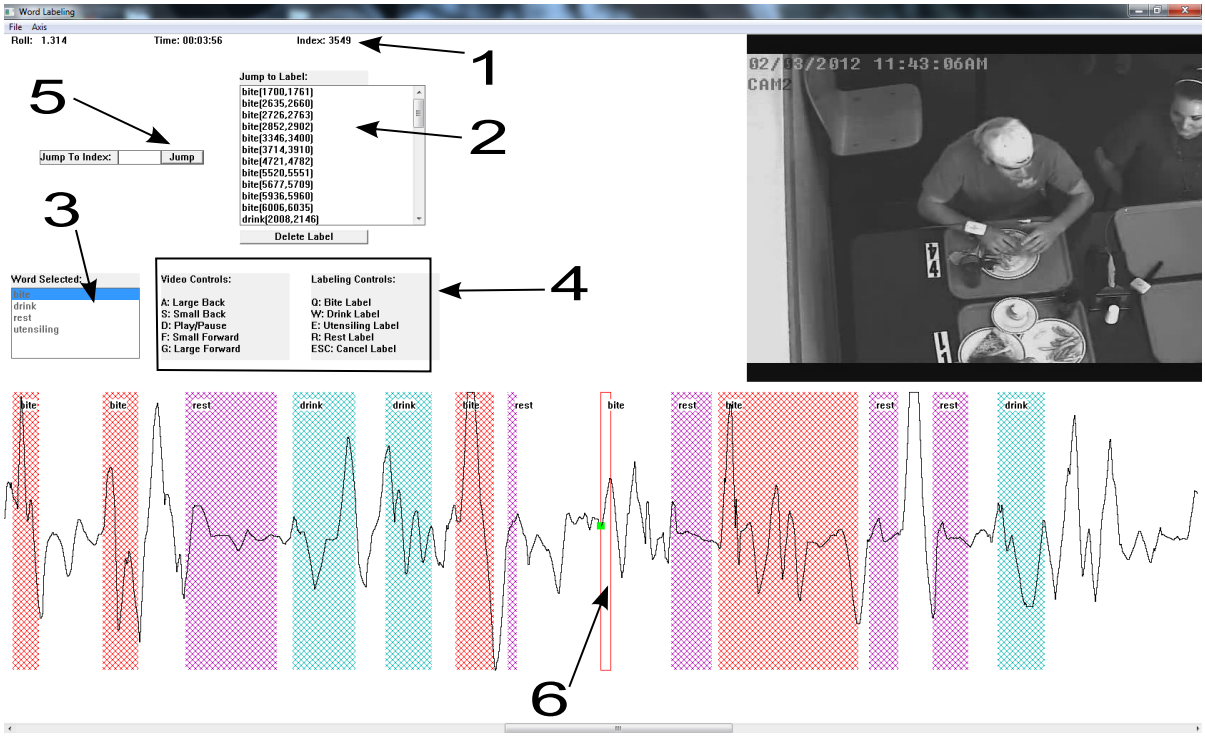


Figure 2.4: Labeling software.

Table 2.4: Word statistics based on one meal.

Gesture	Min Time (sec)	Max Time (sec)	Average Time (sec)	Std. Dev Time (sec)
Rest	1.9	42.9	10.1	7.9
Utensiling	2.1	50.1	8.9	6.1
Bite	1.1	11.4	2.9	1.3
Drink	2.1	15.1	6.0	2.8

Each rater was given a final set of instructions for labeling the gesture definitions provided to them. The list is given below:

1. Carefully read the provided gesture definitions.
2. Usually at the beginning of the meal the participant is answering questions from an interviewer, this time can be ignored.
3. If the subject simultaneously eats and answer questions or has been interrupted while eating to answer questions by an interviewer, you may label the appearing gestures.
4. Time used to put the device on and take it off should be ignored.
5. Labeling starts when subject is about to start eating.
6. When labeling, a tolerance of 0.5 sec (7 or 8 samples) before starting the action and after finishing the action is permitted.
7. If an action is shorter than the initial bounding box then such action can be discarded.

Five raters were involved for ground thruthing the data. These raters labeled three meals separately and these were used to evaluate the inter-rater reliability as explained in section 2.2.2.3. Subsequently, a single rater hand labeled 17 meals and two other raters labeled 4 meals each to complete a set of 25 labeled meals.

2.2.2.2 Other gestures

After labeling the data, there were short periods of unlabeled time between words created by the transition from one gesture to another (e.g. utensiling to bite). We refer to these as “gaps”. Because our sampling rate is 15Hz, some of these gaps could be as small as 67ms. We do not consider

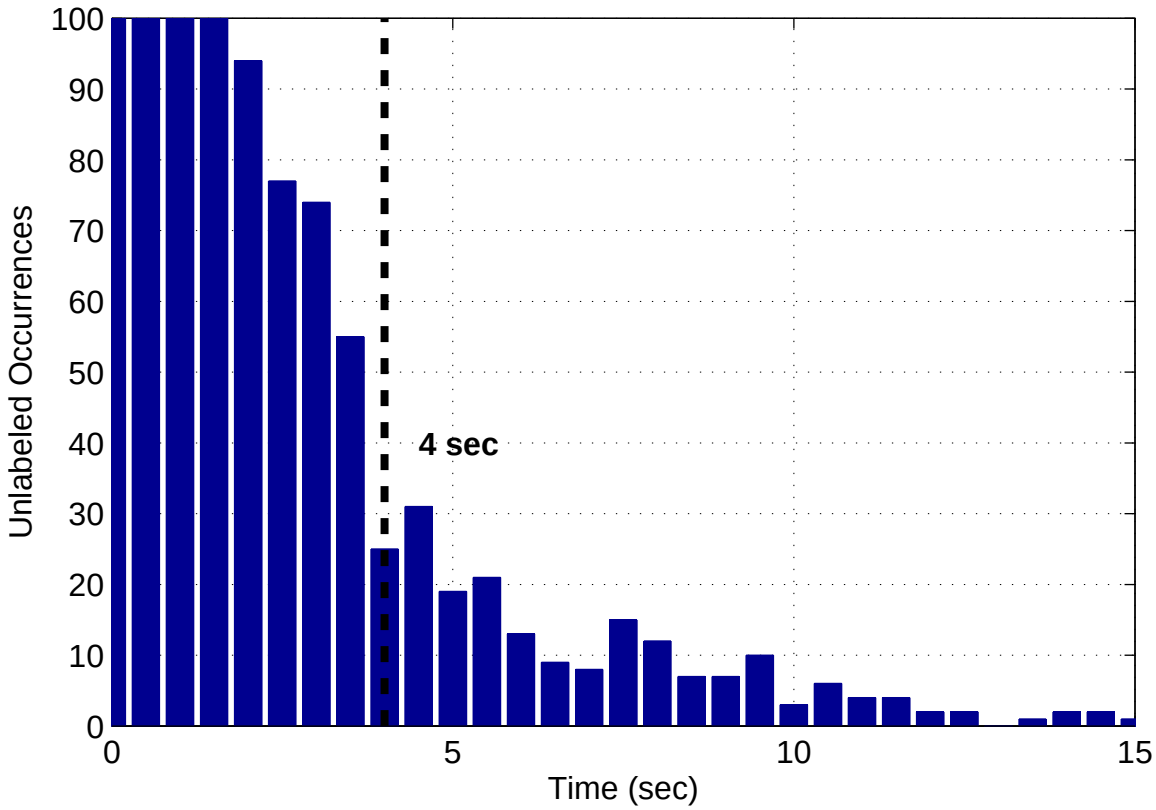


Figure 2.5: Histogram of other gestures for 20 meals between 0 to 15 sec.

these gaps to be the same in nature as larger sections of unlabeled data that correspond to other activities. We therefore devised a strategy to remove the gaps from consideration. After labeling the 20 meals a histogram of the unlabeled sections was studied. Figure 2.5 shows the distribution of unlabeled sections of data from the 25 meals for periods of up to 15 sec. The figure shows a knee in the curve at 4 sec (dashed line) which we used to indicate where the true distribution for other unlabeled activities overlaps the gap distribution. Based on this analysis, we discarded unlabeled sections of data that were less than 4 seconds, and identified unlabeled sections longer than 4 seconds using the gesture *other*.

Table 2.5 shows the total count and total time of gestures from the 25 meals that includes all unlabeled segments. Table 3.1 shows the final gesture count and word duration after removing gaps. A total of 86.9% of the data was labeled using our 4 gesture language, while 13.1% was considered to be *other* activities.

Table 2.5: Total gestures and time before removing gaps.

Gesture	Total Count	Total Time (min)
Rest	582	80.6
Utensiling	700	84.7
Bite	1039	52.3
Drink	155	18.6
Other	2216	78.1
Total	4692	314.4

Table 2.6: Total gestures and time after removing gaps.

Gesture	Total Count	Total Time (min)
Rest	582	80.6
Utensiling	700	84.7
Bite	1039	52.3
Drink	155	18.6
Other	310	44.5
Total	2785	280.8

2.2.2.3 Inter-rater reliability

In order to determine the stability of our definitions and the ground truthing process, three meals were chosen randomly and labeled independently by five raters. A "meta" ground truth was created for each meal by taking the majority vote of the five raters for every unit of time (67 msec, from the 15Hz rate of recording). After comparing the meta GT against labeled sequences, the total labeled time was used as the parameter to estimate the percentage of the meal activities that could be described using our 4 gestures. Additionally, the creation of meta GT provided insight regarding how well raters agreed when labeling. Finally, a confusion matrix was computed from the three meals to provide insight into the level of ambiguity of word definitions.

The three chosen meals have a total recording time of 12.7, 13.3, and 20.3 minutes, respectively. The time needed for the participant to put on or take off the wrist-worn device was considered as dead time, i.e., this amount of time was removed from the total time. After removing dead time the respective times were 12.4, 12.3, and 17.7 minutes for the three meals.

Tables 2.7, 2.8, and 2.9 list how often each rater used each of the gestures in each meal. In each cell are two values; the first value indicates how often the rater used the gesture and the second value indicates the total time (as percentage of the total meal time) that the rater used the

gesture. From these tables, it can be observed that there is more variability between raters in total gesture count than in total time labeled with each gesture. Therefore we use total time as a metric for further evaluation.

Rater	Rest Gest.(%Time)	Utensiling Gest.(%Time)	Bite Gest.(%Time)	Drink Gest.(%Time)	Other Gest.(%Time)
Rater 1	14(8.4)	30(18.1)	45(27.1)	4(2.4)	73(44.0)
Rater 2	20(11.0)	34(18.7)	39(21.4)	2(1.1)	87(47.8)
Rater 3	22(13.0)	46(26.6)	37(21.9)	2(1.2)	63(37.3)
Rater 4	4(2.6)	24(15.8)	43(28.3)	5(3.3)	76(50.0)
Rater 5	14(9.2)	42(27.5)	37(24.2)	2(1.3)	58(37.9)

Table 2.7: Number of gestures and total percentage labeled time for meal 1 per rater.

Rater	Rest Gest.(%Time)	Utensiling Gest.(%Time)	Bite Gest.(%Time)	Drink Gest.(%Time)	Other Gest.(%Time)
Rater 1	23(12.8)	17(9.5)	50(27.9)	5(2.8)	84(46.9)
Rater 2	29(13.5)	31(14.4)	47(21.9)	5(2.3)	103(47.9)
Rater 3	33(17.2)	42(21.9)	49(25.5)	5(2.6)	63(32.8)
Rater 4	10(6.0)	19(11.4)	49(29.3)	5(3.0)	84(50.3)
Rater 5	36(15.7)	47(20.5)	47(20.5)	5(2.2)	94(41.0)

Table 2.8: Number of gestures and total percentage labeled time for meal 2 per rater.

Rater	Rest Gest.(%Time)	Utensiling Gest.(%Time)	Bite Gest.(%Time)	Drink Gest.(%Time)	Other Gest.(%Time)
Rater 1	30(13.5)	35(15.8)	44(19.8)	15(6.8)	98(44.1)
Rater 2	34(13.6)	35(14.0)	44(17.6)	15(6.0)	122(48.8)
Rater 3	35(16.5)	37(17.5)	44(20.8)	15(5.7)	84(39.6)
Rater 4	14(6.8)	30(14.6)	44(21.4)	15(7.3)	103(50.0)
Rater 5	39(18.8)	36(17.3)	43(20.7)	15(7.2)	75(36.1)

Table 2.9: Number of gestures and total percentage labeled time for meal 3 per rater.

A collective meta GT sequence was calculated for each meal. Figure 2.6 shows a graphical description of the process. The meta GT is built by majority vote of the label used by the five raters. Voting is done independently at each time unit. If all individual rater labels are different, then the meta GT is left unlabeled. The result of the majority voting scheme is shown at the bottom of figure 2.6 and is identified as the meta GT of the meal. The complete labeling from each rater and the meta GT are shown for all 3 meals in figure 2.7. Visually it can be seen that raters agreed fairly consistently. Table 2.10 summarized the meta GT for the three meals in terms of the total number of gestures as well as the total percentage of labeled time for each word.

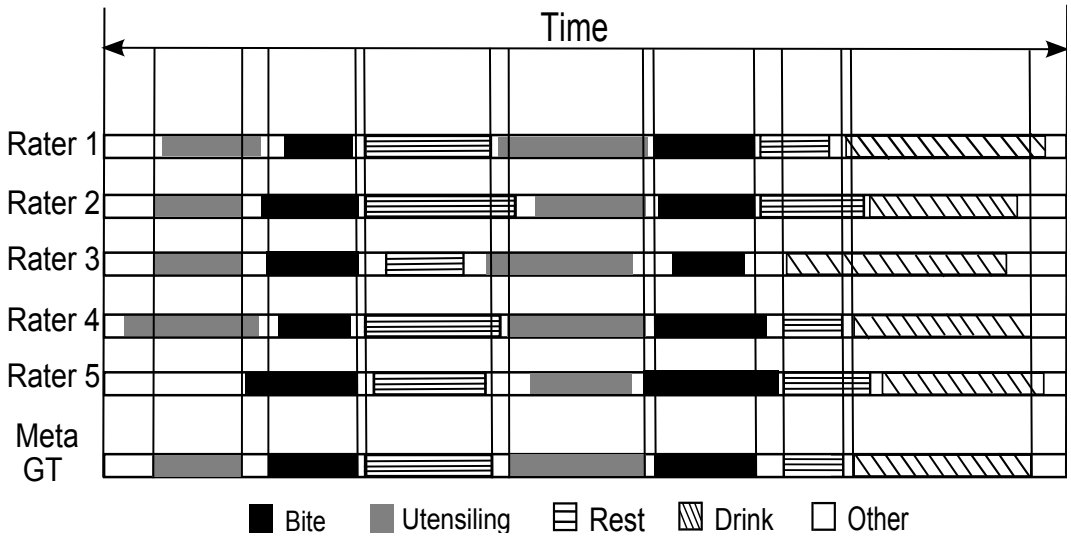


Figure 2.6: Ground truth sequence.

Table 2.10: Meta GT sequence.

Meal	Rest Gest.(%Time)	Utensiling Gest.(%Time)	Bite Gest.(%Time)	Drink Gest.(%Time)	Other Gest.(%Time)
1	18(20.4)	35(39.0)	43(11.0)	3(4.1)	87(25.5)
2	28(25.0)	36(25.9)	50(14.8)	5(3.4)	102(30.8)
3	30(20.6)	38(28.6)	45(13.2)	15(9.0)	103(28.6)

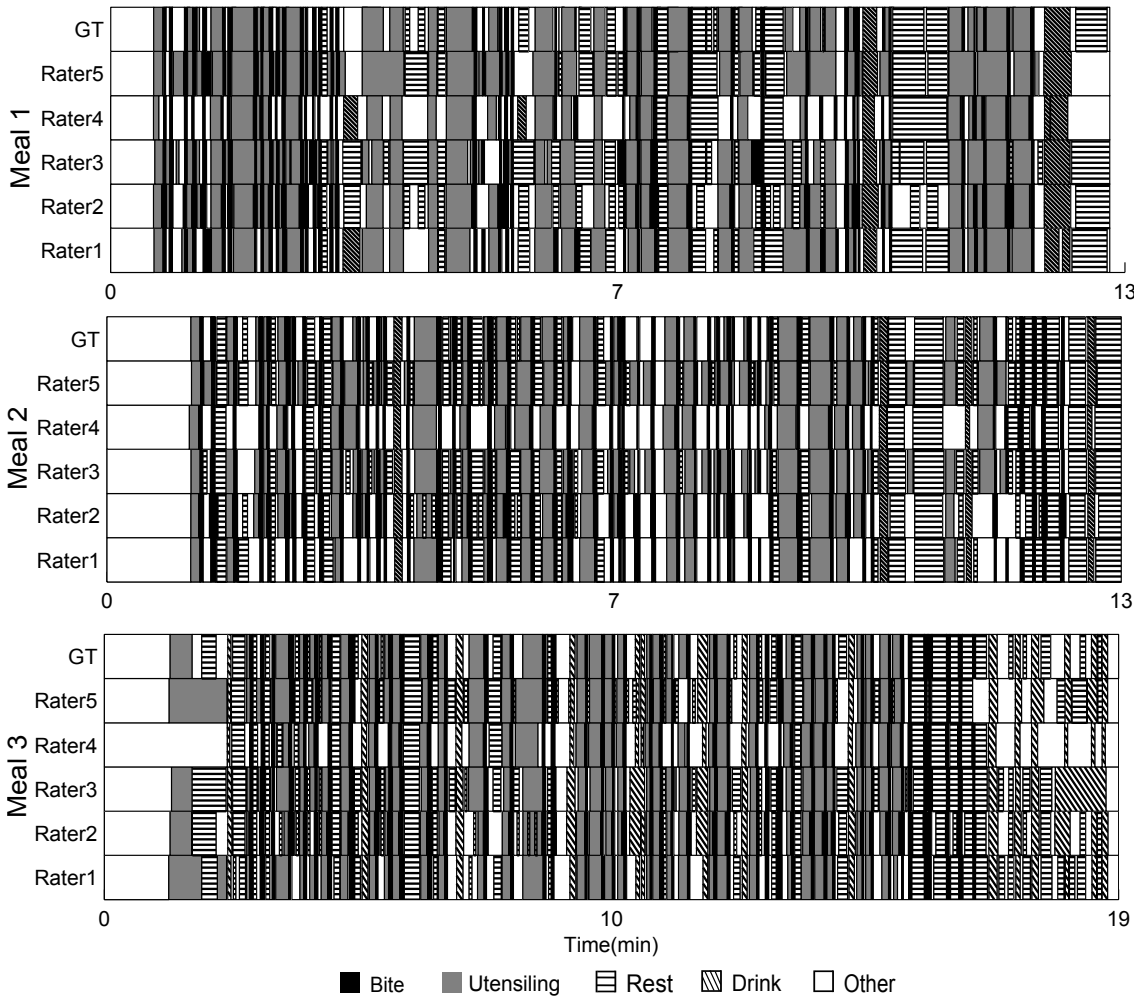


Figure 2.7: Meta ground truth from multiple raters.

2.3 Results

To evaluate agreement quantitatively, individual rater labelings were compared against the meta GT. The percentage of time that the participants agreed over the course of each meal was 89.6%, 93.6%, 89.1%, for meal 1, meal 2, and meal 3, respectively. The average inter-rater agreement across the three meals was 90.7%. Additionally, a confusion matrix was computed to further determine the level of confusion between gestures. The later evaluation was done independently for every unit time, as above. The process begins by computing individual confusion matrices for each rater given a meal. The word label for every unit of time or sample of the GT sequence is compared against the label in the rater sequence, and the frequency of matching or confusing the sample with each other gesture is recorded, as in equation 2.2.

$$C_{R_k}(i_n, j_n) = C_{R_k}(i_n, j_n) + 1 \tag{2.2}$$

Here, $C_{R_{mk}}$ represents the confusion matrix of a rater for given a meal, where $m = 1, 2, 3$ (meal index), $n = 0, .., N-1$ (sample index of the meal), $k = 1, ..., 5$ (rater index), and i, $j \in$ {Rest, Utensiling, Bite, Drink, Other}. Second, rater's confusion matrices for a given meal are combined by using equation 2.3.

$$C_{M_m} = \sum_{k=1}^{5} C_{R_k} \tag{2.3}$$

Lastly, the confusion matrices from each of three meals are combined into a single confusion matrix C, as seen in equation 2.4. This matrix captures the total presence of each word in the three meal and it quantifies the percentage of samples confused with other words. Table 2.11 shows the result of the confusion matrix C of the three meals combined.

$$C = \frac{\sum_{m=1}^{3} C_{M_m}}{5(N_{M_1} + N_{M_2} + N_{M_3})} \times 100\% \tag{2.4}$$

2.4 Conclusion

Four eating activities were defined and an other category is considered to model all activities which do not belong to our four main eating activities. To this point the study has been developed at

Table 2.11: Inter-rater confusion matrix (units are % time during 3 meals).

Gestures	Rest	Utensiling	Bite	Drink	Other
Rest	13.1	0.2	0.1	0.2	2.9
Utensiling		20.0	0.5	0.0	2.4
Bite			8.2	0.0	1.0
Drink				4.0	0.5
Other					38.9

the meal level, therefore sequential dependency of other daily activities are not included. Inter-rater reliability was found to be 90.7% showing that our definitions are fairly objective. Although four words is a fairly limited set, we found that they comprised on average 86.9% of the time during a meal. A confusion matrix shows that the greatest ambiguity resides in the remaining 9.3% of other activities, suggesting that additional words may be useful (e.g. cleaning hand or mouth with napkin).

Chapter 3

Recognition of Eating Gestures Using Inter-Gesture Sequential Dependencies

3.1 Introduction

This chapter considers the problem of recognizing eating gestures by tracking wrist motion. Eating gestures are activities commonly undertaken during the consumption of a meal, such as sipping a drink of liquid or using utensils to cut food. Each of these gestures causes a pattern of wrist motion that can be tracked to automatically identify the activity. Previous works have studied this problem at the level of a single gesture. In this chapter, we demonstrate that individual gestures have sequential dependence. To study this, three types of classifiers were built: 1) a K-nearest neighbor classifier which uses no sequential context, 2) a hidden Markov model (HMM) which captures the sequential context of sub-gesture motions, and 3) HMMs that model inter-gesture sequential dependencies. We built first-order to sixth-order HMMs to evaluate the usefulness of increasing amounts of sequential dependence to aid recognition. On a dataset of 25 meals, we found that the baseline accuracies for the KNN and the sub-gesture HMM classifiers were 75.8% and 84.3%, respectively. Using HMMs that model inter-gesture sequential dependencies, we were able to increase accuracy to up to 96.5%. These results demonstrate that sequential dependencies

exist between eating gestures and that they can be exploited to improve recognition accuracy.

3.2 Methods

3.2.1 Data

The 25 meals described in table 2.2 were hand labeled using the custom tool depicted in figure 2.4. Table 3.1 lists the total number, average time and cumulative duration of gestures in the data set. This data was used for training and testing of the classifiers described next.

Table 3.1: Occurrences, average time per gesture, and cumulative time of gestures.

Gesture	Occurrences	Avg. Time (sec)	Total Time (min)
Rest	582	8.3	80.6
Utensiling	700	7.3	84.7
Bite	1039	3.0	52.3
Drink	155	7.2	18.6
Other	310	8.6	44.5

3.2.2 Classifiers overview

Before introducing our new classifier, we describe two simpler classifiers used to establish baseline accuracy. This provides a reference for evaluating the improvement obtained by our method. Figure 3.1 illustrates the windows of time over which features are calculated for each classifier. The first classifier is a K-nearest neighbor (KNN) and calculates features across an entire gesture. The second classifier is an HMM and calculates features across sub-gesture periods of time. This is the classic gesture spotting approach. It is assumed that the second classifier will perform better than the first due to its use of sequential context within a gesture. The third classifier is also an HMM but it calculates features across an entire gesture, modeling inter-gesture sequential dependencies. Using the first two classifiers, each gesture is recognized independently, while the third classifier uses the context of one or more previous gestures to improve recognition accuracy. The following sections provide the details for each classifier.

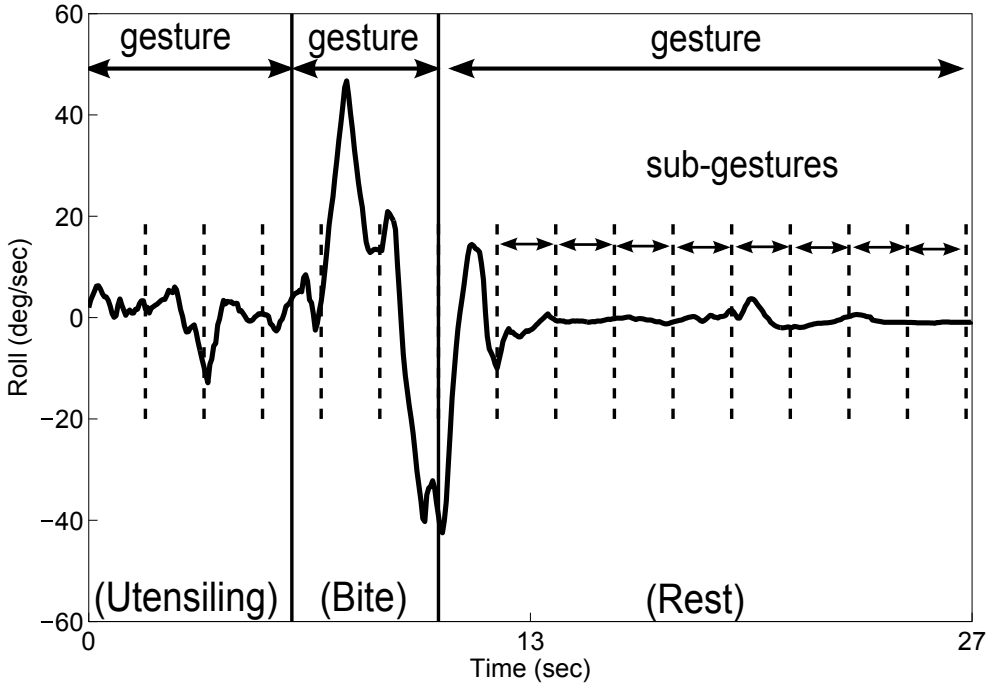


Figure 3.1: Periods of time used to calculate features for the classifiers.

3.2.3 K-nearest neighbor

For our first baseline we used a KNN classifier [137]. Each vector in a training data set is associated with a class label. The process of classification places an unlabeled sample x in the feature space among the training data. The classification searches for the K closest labeled samples to x. The number of occurrences of each label are calculated among the K closest samples. The label with the largest number of occurrences is assigned to x. Features are normalized by computing the Z-norm. Euclidean distance was used for feature comparison.

We calculated the following features across the duration of a gesture: the total motion in each of the 6 axes, the 15 correlations between all pairs of axes, and the ratio of rotational motion to linear motion (called manipulation in [59]). This feature set was reduced to those most useful for classification using a feature forward selection method [10]. This is an iterative process that begins by selecting the single feature that by itself provides the highest classification accuracy. Subsequently, all the remaining features are paired with the first selected feature to find the pair that provides the highest classification accuracy. This process continues until adding another feature results in a negligible increase in classification accuracy. This feature selection process was performed for

Table 3.2: KNN features.

No.	KNN Features
1	Manipulation
2	Amount of motion of AccX
3	Amount of motion of AccY
4	Amount of motion of AccZ
5	Amount of motion of Pitch
6	Correlation(AccX,AccZ)
7	Correlation(AccY,AccZ)
8	Correlation(Yaw,Pitch)
9	Correlation(Yaw,Roll)
10	Correlation(AccX,Yaw)

$K = 1, 3, ..., 19$, using 18 meals for training and 7 meals for testing. This process yielded $K = 7$ and the 10 features shown in table 3.2.

3.2.4 Sub-gesture HMMs

For our second baseline we used HMMs to model the temporal sequencing of the sub-components of each gesture. States represent gesture fragments (see Figure 3.1). Five HMMs were built, one for each of the four defined eating gestures and one for the other category. The HMMs were designed with a left-to-right architecture with skip using an HMM toolbox for MATLAB[1]. For observables, a set of features were computed using a sliding window of 0.5 seconds with 50% overlap. The features included the mean, standard deviation, and slope for each of the six axes. Emission probabilities were calculated by modeling features as continuous observations using Gaussian mixture models (GMMs), where the model order M is the number of Gaussians. We calculated GMMs using the expectation maximization algorithm [139]. The GMM for sub-gesture γ is given by equation 3.1, where c_{γ_m}, μ_{γ_m}, and Σ_{γ_m} represent the mixture weight, mean, and covariance matrix of the m^{th} Gaussian, respectively. In this work, only diagonal covariance matrices were used.

$$G_\gamma = \{c_{\gamma_m}, \mu_{\gamma_m}, \Sigma_{\gamma_m}\}, \text{ where } m = 1, ..., M \tag{3.1}$$

For model parameter selection, we tested several numbers of states (3 to 15) and Gaussians (1 to 10) using a 5-fold cross validation, selecting 13 states and 5 Gaussians as the best parameters for

[1] `http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html`

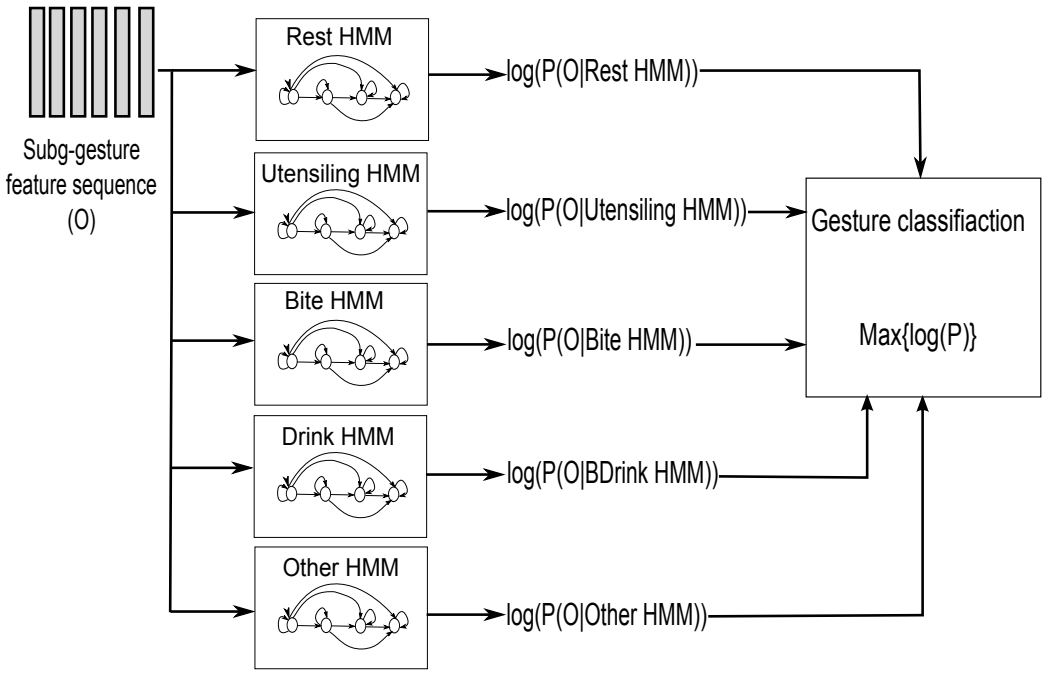


Figure 3.2: Sub-gesture HMM gesture recognition.

each gesture. For recognition, the Viterbi algorithm was used to calculate the log probability of each of the five sub-gesture HMMs given the observed feature sequence. The HMM with the highest log probability determined the classification of the gesture, as shown in figure 3.2.

3.2.5 Gesture-to-gesture HMMs

The purpose of this classifier is to capture sequential context between gestures; in other words, to use the history of one or more preceding gestures to improve the recognition of the current gesture. States represent whole gestures. Figure 3.3 shows an HMM that uses a history of 1 gesture, which we denote as HMM1. Each state corresponds to one of the five gestures (rest, utensiling, bite, drink, and other). This HMM is ergodic, meaning every state can be reached by any other state.

To incorporate additional history we first build a conditional higher-order HMM and then convert it to an equivalent first-order HMM [110]. We denote HMMn as an nth-order HMM that uses the previous n gestures for sequential context. For example, HMM2 has the capability to recognize state q_t based on the previous states q_{t-1} and q_{t-2}. Figure 3.4 shows an example second-order HMM for only two of our gestures (for clarity), bite (B) and utensiling (U). The temporal

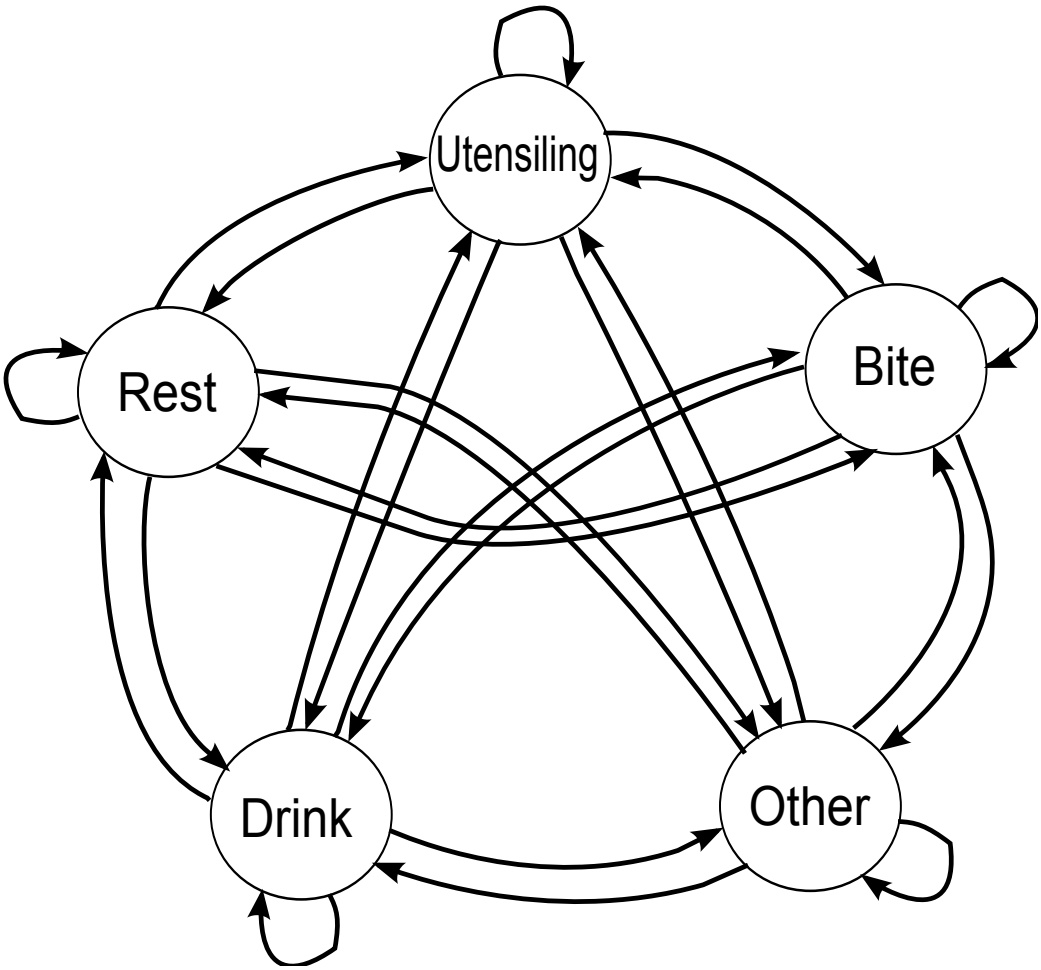


Figure 3.3: HMM1 uses a history of 1 gesture to help recognize the current gesture.

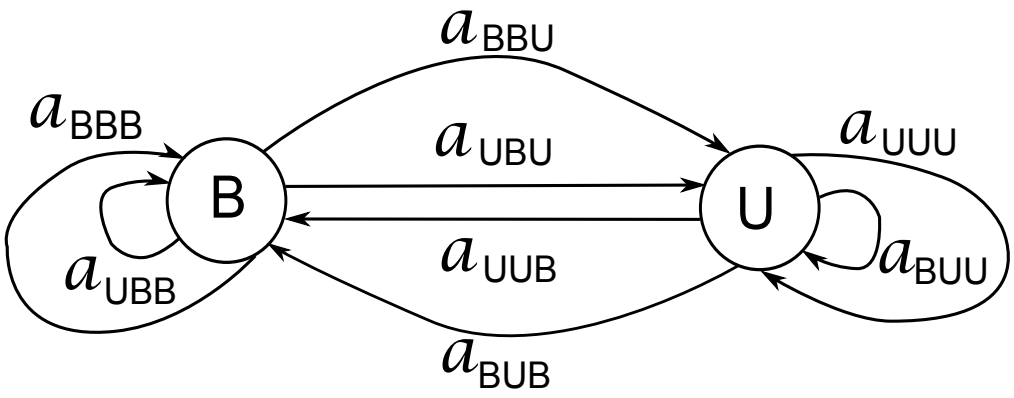


Figure 3.4: Example second-order HMM of bite (B) and utensiling (U). A transition probability a is conditional and has three subscripts indicating the gesture history at times $t-2$, $t-1$, and t.

sequencing is left-to-right. For example, a_{UBU} indicates a state transition where the previous two states were U then B, and the next state is U at times $t-2$, $t-1$, and t, respectively. Figure 3.5 shows the equivalent first-order HMM. Each state models a sequence of two gestures. The state notation (e.g. BB) shows the memory (left-to-right) of the most recently seen gestures. Logically, transitions between some states are impossible. For example, the state BU cannot transition to BB because the former's most recently recognized gesture is U, which does not match the memory of the latter. As another example, figure 3.6 shows the same part of HMM3 after it has been reduced to first-order. The state and transition notation reads left-to-right and defines the memory of the most recently seen gestures. In general, any HMM of order n (HMMn) can be converted into a first-order equivalent.

For observables we used the 5 log probailities obtained from the sub-gesture HMMs described in section 3.2.4. Emission probabilities were calculated by modeling these observables using GMMs as in equation 3.1. We selected the number of Gaussians by performing a 5-fold cross validation for $M = 1, ..., 20$ in HMM1, choosing $M = 7$.

The transition probabilities $a_{\alpha\beta...\phi\omega}$ are equivalent to $P(\{q_t = \omega\}|\{q_{t-1} = \phi, .., q_{t-n-1} = \beta, q_{t-n} = \alpha\})$, where $\alpha, \beta, ..., \phi, \omega \in$ {rest, utensiling, bite, drink, other} and $\alpha\beta...\phi\omega$ represent the states visited (from left to right) at times $t-n$, $t-n-1$, ..., $t-1$, and t, respectively. Thus, for a

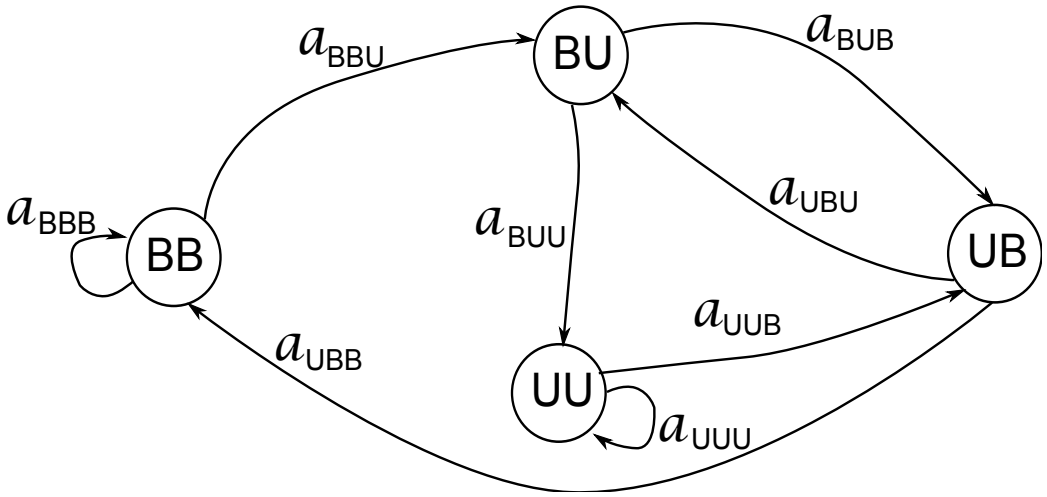


Figure 3.5: Equivalent first-order HMM of figure 3.4. The state notation (e.g. BB) shows the memory (left-to-right) of the most recently seen gestures. A transition probability a has three subscripts indicating the gesture history at times $t-2$, $t-1$, and t..

given set of data, these can be calculated as

$$a_{\alpha\beta\ldots\phi\omega} = \frac{\begin{array}{c}\text{Total \# of transitions from gesture sequence} \\ \alpha\beta\ldots\phi \text{ to gesture } \omega\end{array}}{\text{Total \# of } \alpha\beta\ldots\phi \text{ gesture sequences}} \tag{3.2}$$

Similarly, the prior probabilities can be calculated from a given set of data as

$$\pi_{\alpha\beta\ldots\phi} = \frac{\text{Total \# of } \alpha\beta\ldots\phi \text{ gesture sequences}}{\text{Total \# of n-gesture sequences}} \tag{3.3}$$

The Viterbi decoding algorithm outputs the most probable n-gesture $\alpha\beta....\phi$ state sequence Q, but only ϕ from each state is retained in the output. The other portions of each state are the running memory and are redundant.

A drawback of this approach is that the number of states and transition probability matrix grow exponentially as the order grows. In our case, as we increase the model order, the total number of states are 5^n and the transition probability matrix contains 5^{2n} elements. Although not all transitions are logically possible due to the constraint of the history of gestures, there are no fully-empty columns or rows in the transition matrix. Due to software limitations we were only able to test this approach for our data up to $n = 6$.

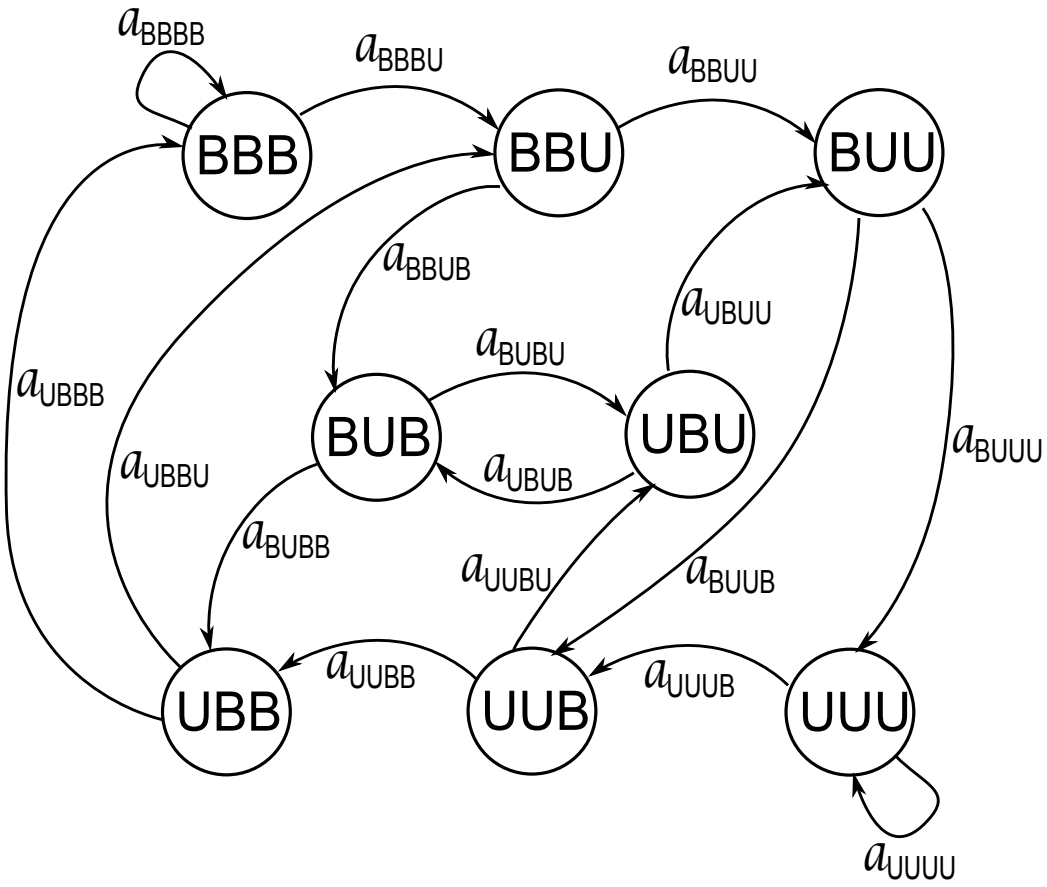


Figure 3.6: Equivalent first-order HMM from a third-order HMM for bite (B) and utensiling (U). The state and transition notation reads left-to-right and defines the memory of the most recently seen gestures.

Table 3.3: Transition probabilities.

From \ To	Rest	Utensiling	Bite	Drink	Other
Rest	0.072	0.338	0.364	0.093	0.133
Utensiling	0.134	0.007	0.811	0.007	0.040
Bite	0.309	0.364	0.141	0.022	0.165
Drink	0.253	0.226	0.137	0.151	0.233
Other	0.285	0.256	0.292	0.167	0.000

3.3 Results

The transition probabilities found for HMM1 are shown in table 3.3. The amount of sequential dependence between gestures can be seen in entries with values larger or smaller than 0.2. For example, the likelihood of transitioning from utensiling to bite is 81.1%. The tables for HMM2 to HMM6 are too large to display easily.

All classifiers were trained and tested using leave-one-out cross validation. Training the sub-gesture and gesture-to-gesture HMMs consisted of calculating GMM values (equation 3.1). Training the KNN consisted of populating the feature neighbor space with the training data for purposes of calculating nearest neighbors for the test data.

For all classifiers, accuracy was measured as the total percentage of time in all the meals that gestures were labeled correctly. Table 3.4 presents a summary of the accuracy achieved by each of the classifiers. For the baseline classifiers, the sub-gesture HMMs performs better than the KNN with an 8.5% improvement. Our gesture-to-gesture classifier that incorporates gesture history shows further imporvement, from 3.4% to 12.2% across HMM1 to HMM6. For example, HMM3 uses a history of 3 gestures and achieved 89.6% accuracy while HMM6 uses a history of 6 gestures and achieved 96.5% accuracy.

Table 3.5 shows the results broken down by gesture type. It can be seen that gesture history improves the recognition accuracy of every gesture type except drinks. Table 3.6 shows the confusion matrix for HMM6 for each gesture. The most confusion occurs between drink and rest gestures, and utensiling and other gestures. This is likely due to the similarities of motions between these gesture types. Drink gestures can have pauses that can be confused with brief periods of rest, and both utensiling and other gesture types encompass a wide variety of motions.

We conducted an additional test of our classifier to determine if the most recent gesture history is the most important. This was done also to provide assurance that our results are not

Table 3.4: Total recognition accuracy.

Method	Accuracy (%)
KNN	75.8
Sub-gesture HMM	84.3
HMM1	87.7
HMM2	88.0
HMM3	89.6
HMM4	92.2
HMM5	94.6
HMM6	96.5

Table 3.5: Recognition accuracy for each gesture.

Method	Rest (%)	Utensiling (%)	Bite (%)	Drink (%)	Other (%)
KNN	88.8	76.8	84.3	71.5	31.7
Sub-gesture HMM	91.7	83.2	86.9	86.5	56.0
HMM1	93.5	87.5	93.0	75.1	63.1
HMM2	93.8	87.9	92.8	75.5	64.6
HMM3	94.3	89.5	92.8	77.5	69.0
HMM4	96.8	92.6	95.1	79.2	75.5
HMM5	97.7	94.4	96.2	82.7	82.8
HMM6	99.3	97.3	98.5	82.0	88.4

Table 3.6: HMM6 confusion matrix.

Classifier / Actual	Rest (%)	Utensiling (%)	Bite (%)	Drink (%)	Other (%)
Rest	99.3	0.3	0.1	0.0	0.3
Utensiling	1.2	97.3	0.3	0.0	1.3
Bite	0.5	0.2	98.5	0.0	0.8
Drink	8.5	1.9	3.0	82.0	4.6
Other	2.0	7.5	0.9	1.2	88.4

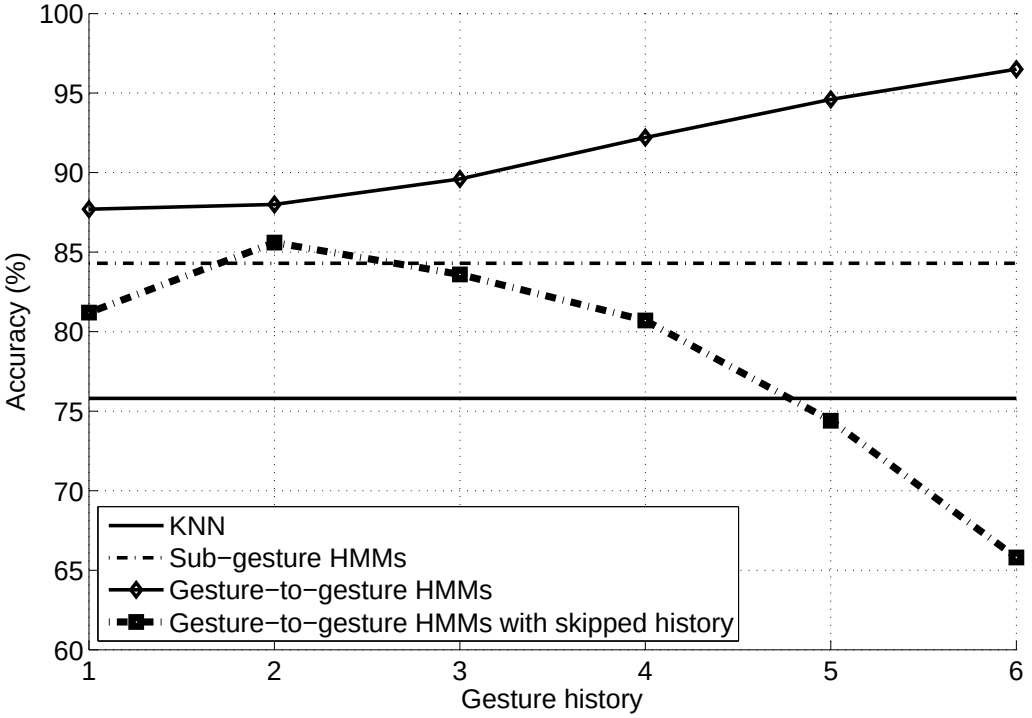


Figure 3.7: Classifier accuracy with increasing gesture history and increased skip in history.

an effect of over-fitting. We constructed an HMMn-skip for each value of n, where the gesture history captured in HMMn skipped over the most recent n gestures and instead used the history of the n gestures preceding them. For example, HMM1-skip models history by skipping over the most recent gesture and instead uses the identity of the one preceding it; HMM2-skip models the history by skipping over the two most recent gestures and instead uses the sequenece of the two preceding them; etc. Formally, equation 3.2 was modified to skip n gestures between $\alpha\beta...\phi$ and ω. Figure 3.7 shows the accuracies of all the classifiers as the amount of history and skipped history is increased. The KNN and sub-gesture HMMs have constant accuracy because they do not incorporate gesture-level history. Our gesture-to-gesture HMM increases in accuracy as the history of gestures is increased. The HMMs with skip perform much worse. This indicates that the most recent gesture history is in fact the most relevant.

3.4 Conclusions

This work demonstrates that the recognition of eating gestures can be improved through knowledge of the sequential dependence of individual gestures. We believe this is due to the patterned nature of activitues during eating. For example, a common pattern is to use utensils to prepare a bite of food, consume the bite of food, and then rest hands while masticating and swallowing. Another common pattern is to intersperse drinks with food bites. Our HMMs capture this "language of eating" and use it to improve recognition accuracy. The results in figure 3.7 show evidence supporting this conclusion. A skipped history of 2 gestures shows improvement over a skipped history of 1 gesture, but additional skipped history shows a continually decreasing accuracy. This is likely due to HMM2-skip capturing some of the cyclical phrasing of gesture patterns, such as utensil-bite-utensil-bite, where the skip of 2 matches the repetition in the phrasing. We assume that other HMMn-skip classifiers capture less phrasing and hence show increasingly worse results. A phrase-level study of the sequential context during eating is a topic for future work.

Methods for activity recognition that are developed in a controlled setting will potentially be brittle in a natural setting. This problem affects all studies where some instrumentation is necessary to record behavior. In our case, our data was collected in a cafeteria which is as natural a setting as possible where wrist motion trackers and discreetly positioned video cameras could be used to record eating. Participants selected their own foods from everything available in the cafeteria and consumed them however they wished, with no instructions on how or what to eat. This is arguably more natural than asking participants to conduct a sequence of scripted gestures in a lab or asking participants to eat a small set of controlled foods (as in [85]). Although our data set was acquired from 25 people each eating a meal, the total gestures recorded numbered 2,786 (see Table 3.1). We believe this captures sufficient variety in pace and style of each gesture type that the improvement we found in recognition is not brittle.

Other limitations of this work include the number of activities modeled and the use of manually segmented data. We found that 15.9% of the total time during meals is comprised of other activities, such as wiping hands on a napkin or gesturing while talking. Increasing the number of activities may affect overall classifier accuracy. However, we still expect that modeling sequential dependencies would produce an improvement in accuracy. For our experiments, we used manually segmented data in order to determine the impact of sequential dependence modeling on classifier

accuracy independent of possible segmentation errors. In the future we plan to explore automated segmentation methods. It may be that sequential dependencies can be exploited to improve automated segmentation as well as classification. Finally, it should be noted that other classifiers besides HMMs can model sequential context such as conditional random fields and dynamic Bayesian networks. The comparison of these classifiers on the problem of eating gesture recognition is a topic of future work.

The method presented in this chapter could be used to improve the accuracy of automated methods that track wrist motion to monitor the number of drinks [17] or bites [57] taken by a person. Our method could also potentially be used to analyze eating habits of individuals that may correlate with variations in energy intake. For example, slowing the pace of eating has been found to be associated with decreased intake during a meal [152]. It may be that other activity patterns have similar associations. In this case, a tool that automatically measures these patterns could prove useful in diagnosis and behavior treatment. These topics are subjects for future work.

Chapter 4

Automatic Classification and Segmentation of Eating Gestures

4.1 Introduction

In this chapter, we explored methods to recognize wrist motion gestures related to eating activities from unsegmented data. In the last chapter, we studied the performance of recognizing these gestures using hand segmented data; in practical use the data must be automatically segmented. Since the beginning and ending points of a gesture are not known, we fragmented the wrist motion signal into frames and classify each frame as belonging to a gesture. Figure 4.1 shows the wrist motion signal fragmented into frames using a 0.5 sec window with a 50% overlap (dashed rectangles). Features are computed using the six motion signals inside the window. Using our classifiers, the frames are labeled either as rest, utensiling, bite, drink or other. Figure 4.2 depicts the labeling of the frames, where U is utensiling, B is bite and R is rest. Finally, consecutive frames pertaining to the same gesture are grouped together, thus segmenting the wrist motion signal as shown in figure 4.3.

For classifiers, the same techniques studied in the previous chapter were used again: KNN, sub-gesture HMMs, and gesture-to-gesture HMMs. In the case of the sub-gestures HMMs, segmentation and classification of wrist gestures is accomplished by connecting the individual HMMs together. For the gesture-to-gesture HMMs, they use the output from the sub-gesture HMM as a preliminary

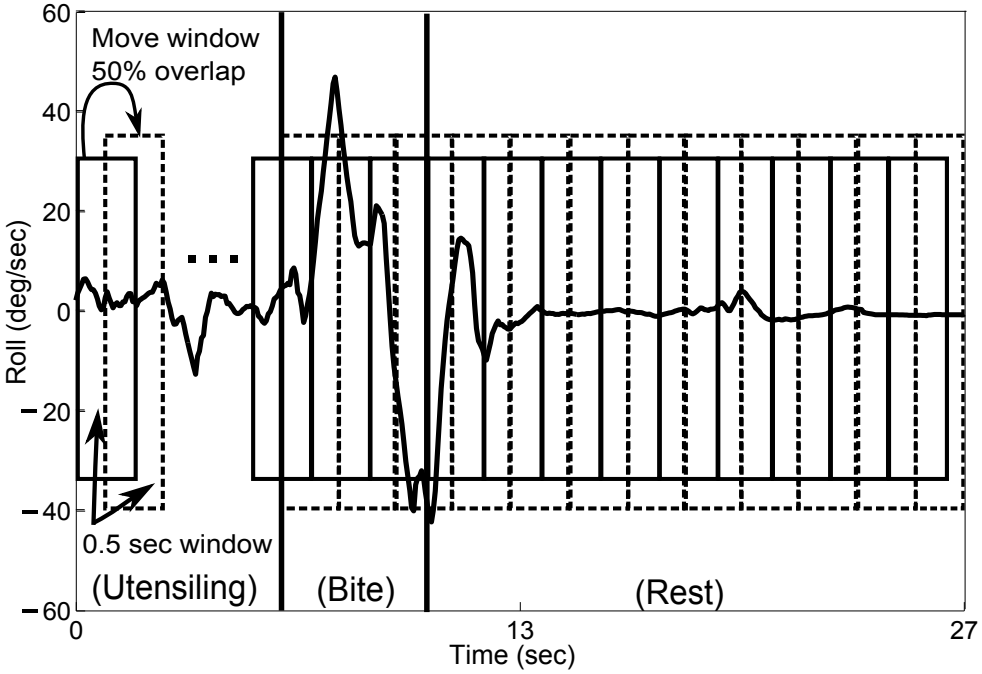


Figure 4.1: Fragmenting the wrist motion signal into frames.

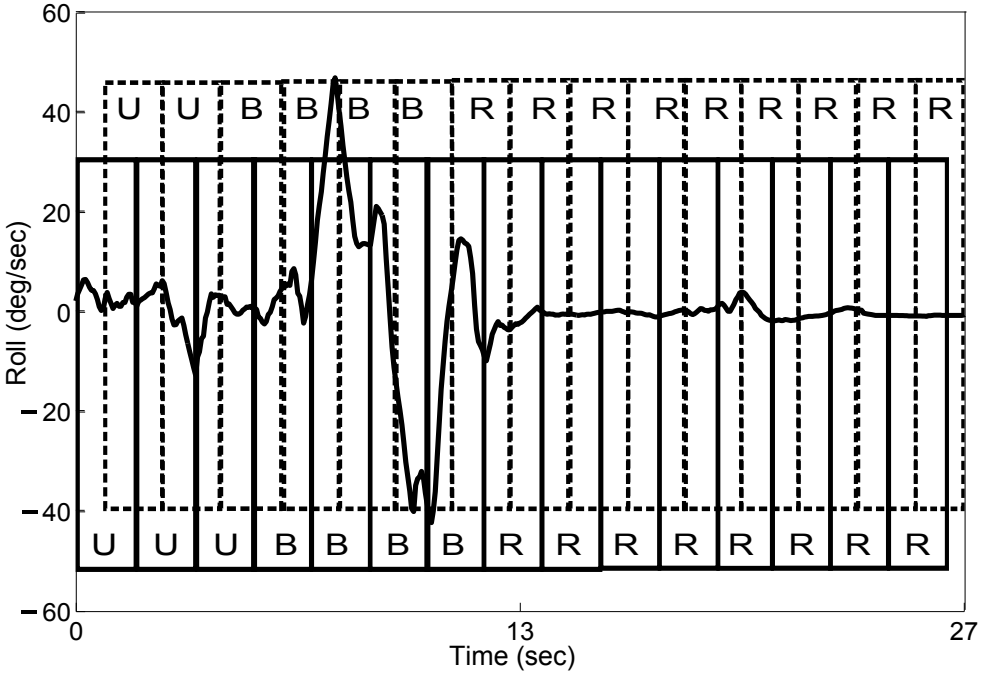


Figure 4.2: Recognizing gestures for every frame.

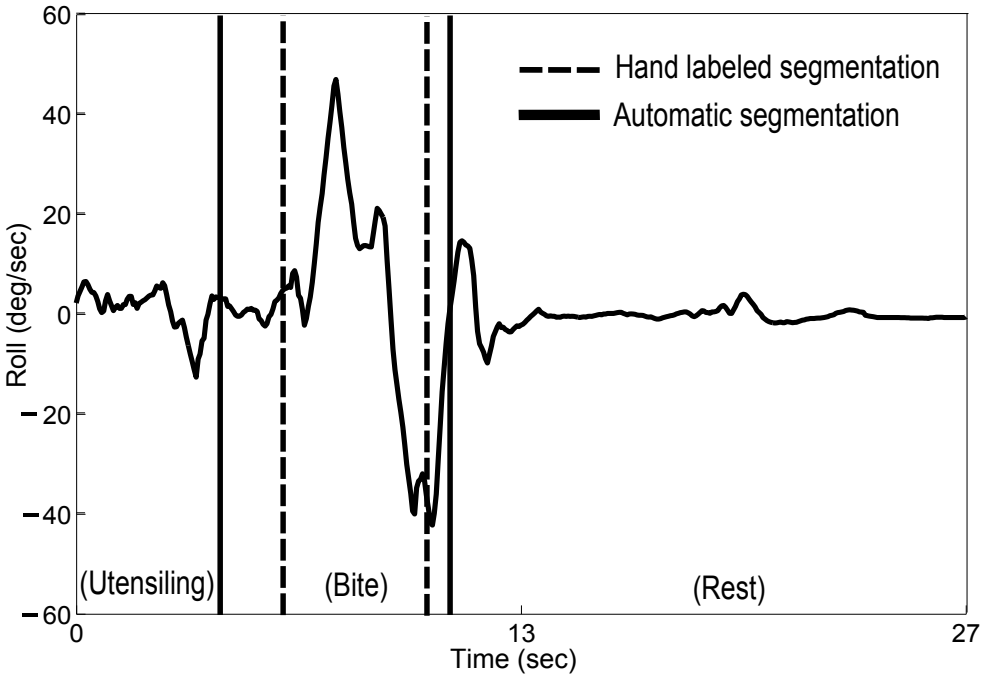


Figure 4.3: Grouping frames into gestures to segment wrist motion.

segmentation that is later reclassified using gesture-to-gesture sequential dependency analysis. For evaluation, the segmented output from the three methods is compared with the hand labeled data. The baseline algorithms achieved a 59.9% and 71.1%, respectively. The gesture-to-gesture HMMs showed an accuracy of 90.0%.

4.2 Methods

4.2.1 Training Data

In chapter 2, ground truth (GT) sequences of eating gestures were established using manual labeling of 25 meals. In this GT there are gaps which are unlabeled sections less than 4 sec that represent transitions between gestures with no real intention; these could be as small as 66.7 msec. The segmentation task of wrist motion data will consist of analyzing a small section of the wrist motion signal to determine which gesture corresponds to that particular segment. So, from a segmentation perspective, it is not necessary to distinguish these gaps from actual unlabeled sections with real intention, i.e. the other gesture. Hence, the hand labeled gesture sequences are preprocessed to

eliminate these gaps as shown in figure 4.4. Gaps are located in a meal and are labeled using the labels of the gestures located in the extremes of the gap. Half of the gap is filled using the gesture label on the left side of the gap and the other half is labeled using the gesture label on the right of the gap. If the gap is only one sample point, then this is labeled with the gesture label on the left of the gap. These labels are used for training purposes for the classifiers. It is important to note that the original hand labeled GT is still used to evaluate segmentation and classification performance.

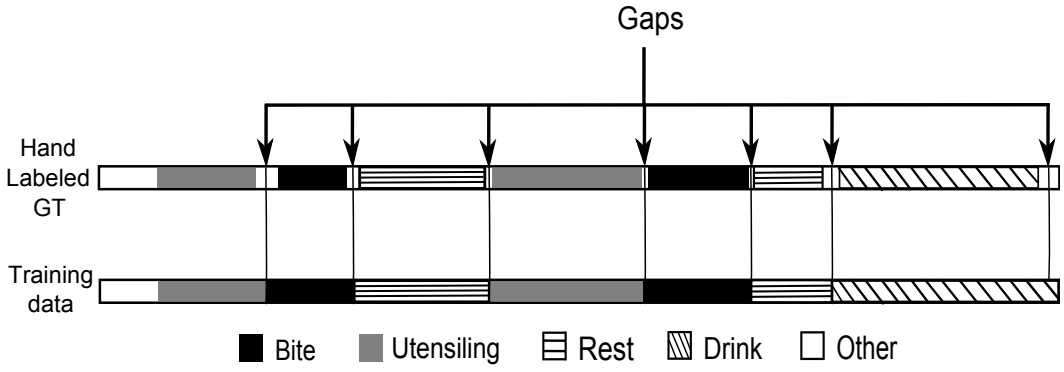


Figure 4.4: Smoothing gaps for training data.

4.2.2 KNN segmentation and classification

The goal is to segment an unlabeled continuous wrist motion signal and classify segments into one of five gestures: rest, utensiling, bite, drink, and other. For the KNN segmentation and classification, features from table 3.2 are computed on the accelerometers and gyroscopes signals using a frame window of 0.5 sec with 50% overlap. Each one of these frames are labeled using the training data of figure 4.4. If a frame overlaps two gestures, it gets assigned to the gesture it overlaps the most. Using this classifier, segmentation and classification happen simultaneously. Features are normalized using the Z-norm and Euclidean distance is use as measure of distance. Using the 25 meals, a five fold cross validation was performed for $K = 1, 3, 5, ..., 101$, selecting $K = 45$.

4.2.3 Sub-gesture HMMs segmentation and classification

Sub-gestures HMMs are interconnected to create a single HMM to classify and segment simultaneously the frames of a continuous wrist motion signal. Figure 4.5 shows the implementation

where each box represents a sub-gesture HMM. The last state has a transition probability to the first state of each sub-gesture HMM. The value of these transitions are $\frac{1-p}{5}$, where p is the self transition probability of the last state in a subg-gesture HMM. In this manner the probability of transitioning from one sub-gesture HMM to another is the same.

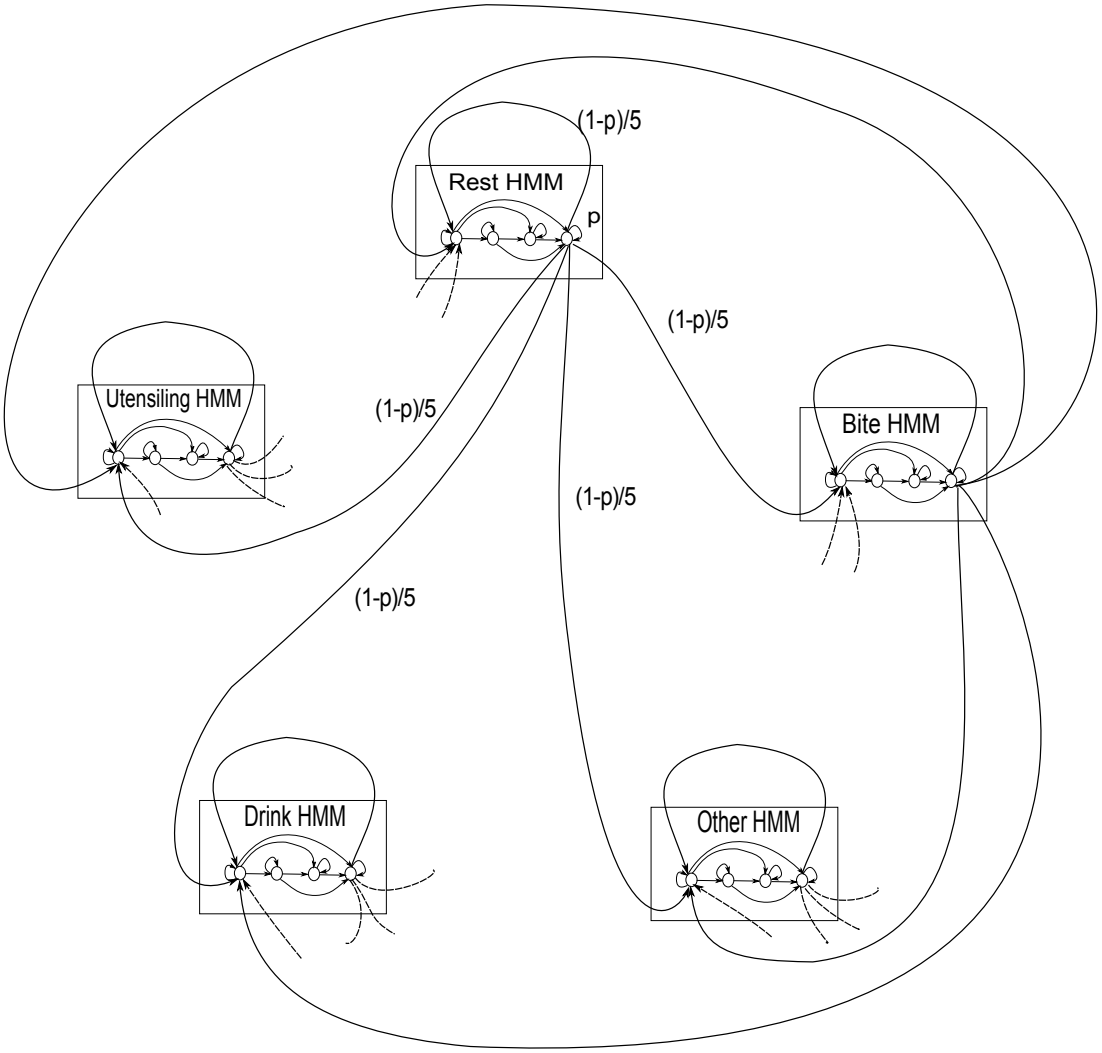


Figure 4.5: Interconnection of sub-gesture HMMs. Each box represents a left-to-right with skip sub-gesture HMM. The last state of each HMM is connected to the first state of the other HMMs. Equal probabilities are assigned to these transitions based on the probability p of the last state of the HMM, i.e. $\frac{1-p}{5}$.

The complete process for the implementation of figure 4.5 involves computing the features from the wrist motion signals, training the transition probabilities for each sub-gesture HMM, and lastly connecting each sub-gesture HMM to all the other sub-gesture HMMs. As in the KNN,

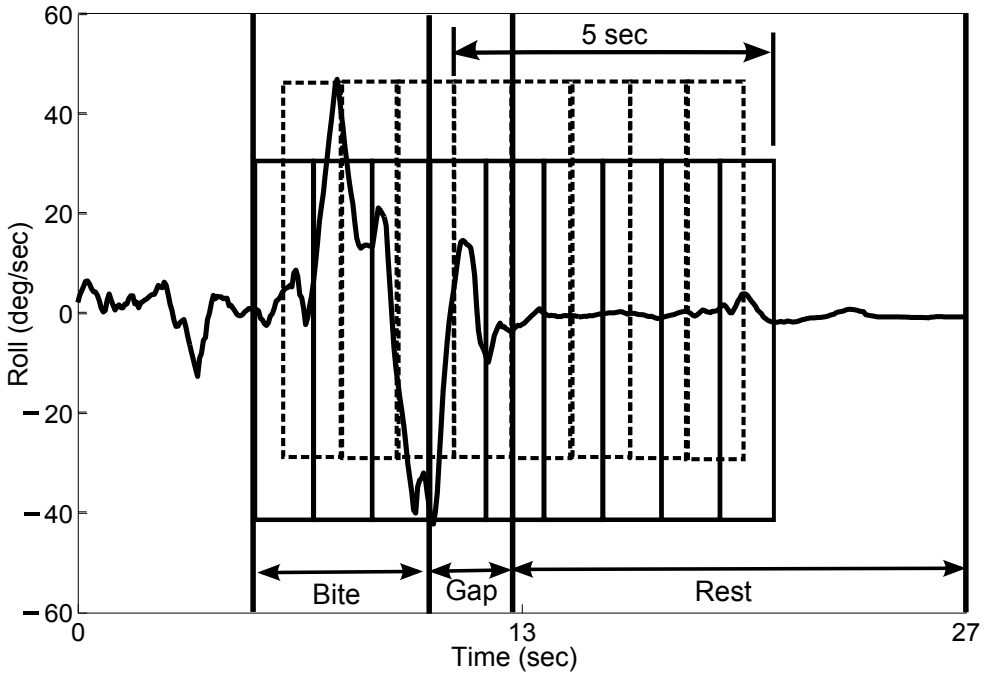


Figure 4.6: Training bite sub-gesture HMM.

features are calculated using a frame window of 0.5 sec with 50% overlap. The features in each frame include the mean, standard deviation, and slope for each of the six axes. Frames are labeled in the same manner as in the KNN to train the sub-gestures HMMs. Because each sub-gesture HMM is trainined independently, the transition probability for the last state of each sub-gesture HMM has a self transition with probability of one. This means that no knowledge of subsequent activities is modeled. To address this issue, when training a sub-gesture HMM, all frames corresponding to a particular gesture are gathered with 10 additional frames which represents 5 sec after the gesture. This additional information is included in the training so that the last state resembles the transition and beginning of the upcoming gesture as shown in figure 4.6.

The last state is then removed (the sink state with the single transition probability of 1) and all transitions connecting from the other preceding states to the last are modified. Figure 4.7 shows an example for the rest sub-gesture HMM. For simplicity the HMM is illustrated consisting of four states. The upper box represents the HMM built for the rest gesture after initial training. The bottom box shows the final sub-gesture HMM for rest after removing the last state and readjusting the transition probabilities. Finally, the sub-gestures HMMs are connected together as explained

earlier.

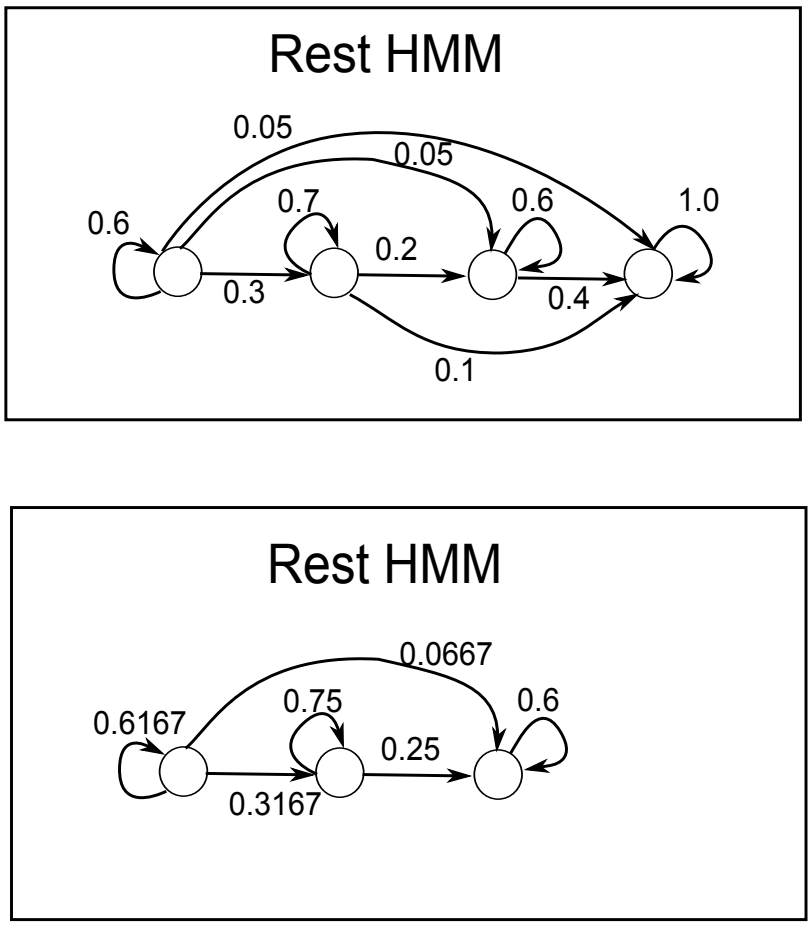


Figure 4.7: Preprocessing of sub-gesture HMMs.

To process a continuous unlabeled wrist motion signal, the features from the frames are used as input into the fully connected sub-gesture HMM and the Viterbi algorithm is used to find the most likely state path. Since specific states represent a gesture, the output sequence path has to be mapped into one of five gestures to be able to compare with the hand labeled GT. For example, for simplicity, if each sub-gesture HMM has three states then there are 15 states in the fully connected sub-gesture HMM. States 1 to 3 represent the rest gesture, state 4 to 6 represent the utensiling gesture, states 7 to 9 represent the bite gesture, states 10 to 12 represent the drink gesture, and states 13 to 15 represent the other gesture. Figure 4.8 shows an example of the Viterbi state sequence output and how frames are grouped to finalize the segmentation of gestures.

A five fold cross validation was performed using the 25 meals to obtained the number of

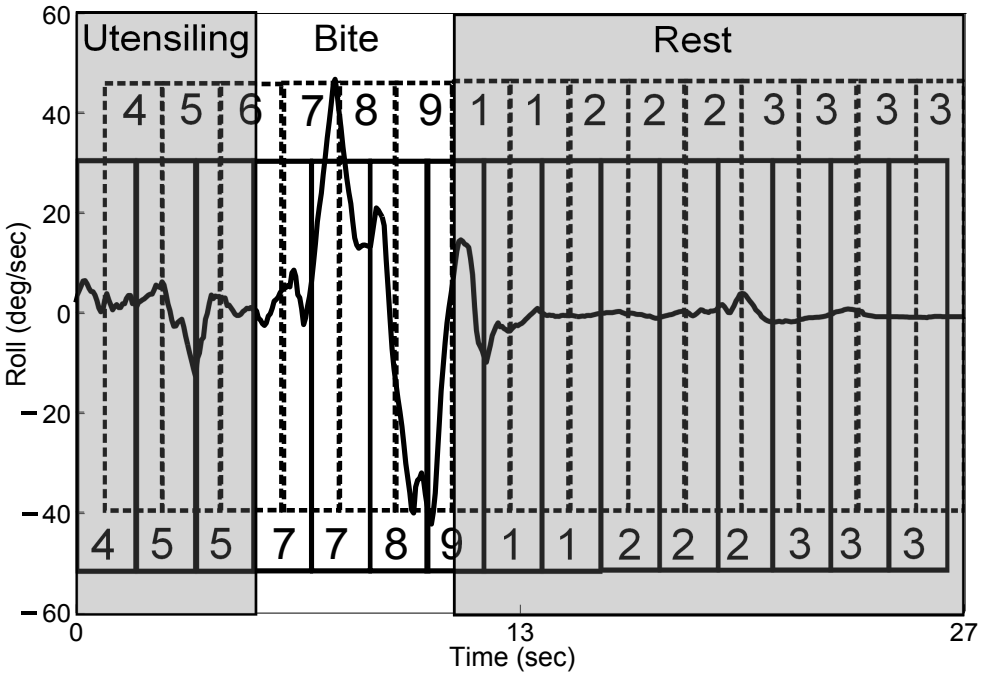


Figure 4.8: Grouping state output sequence to segment wrist motion signal.

states for the sub-gesture HMMs and the number of Gaussians to model each state. The validation tested the range of states ($Q = 3, 4, 5, ..., 20$) and the range of Gaussians ($M = 1, ..., 10$), selecting 17 states and 10 Gaussians as the best parameters.

4.2.4 Gesture-to-gesture HMMs segmentation and classification

For this classifier, the sub-gesture HMM segmentation is used as an initial segmentation and classification. The gesture-to-gesture HMM calculates new features across each entire segment and then re-classifies them using inter-gesture sequential dependencies. Although this will not substantially change the segmentation, in can change it somewhat in cases where a segment is relabeled to the same label as a preceding or succeeding segment, which in effect also changes the final segmentation.

For training purposes, the original hand labeled GT for each meal was used. Figure 4.9 shows the process of building the training data using the hand labeled GT and the segmentation sequence from the sub-gesture HMM. The latter is used to propose an initial segmentation of gesture sequences which are relabeled using the information on hand labeled GT. The bottom line shows

the new training data that will be use in the gesture-to-gesture HMMs. It can be observed that a single gesture in the hand labeled GT can be split into two or more in the same period of time in the new training data.

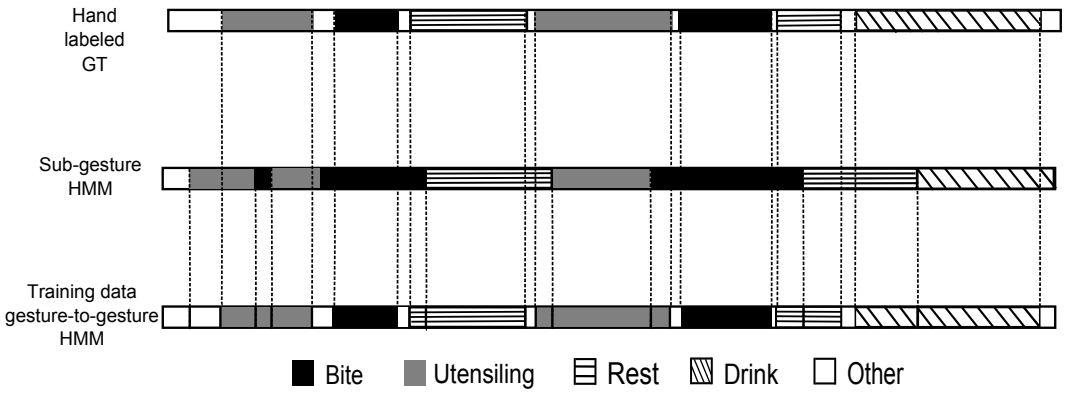


Figure 4.9: Training data for gesture-to-gesture HMM.

The classification is a two-step process. First, using the new training data from figure 4.9, features (mean, standard deviation, and slope) are calculated on each gesture segment through an overlapping 0.5 sec window. These features are used to model the states in a sub-gesture HMMs; 13 states and 5 Gaussians are used as in the previous chapter. After modeling, the features corresponding to a gesture segment are fed into each of the sub-gesture HMMs as shown in figure 3.2. Second, the log probability score for each gesture segment is output by each model and are collected to form a 5-element feature vector. These features are used to model the states in the gesture-to-gesture HMMs as in figures 3.3, 3.5, and 3.6; 7 Gaussians per state are used as before. Also, the transition probabilities and prior probabilities for all gesture-to-gesture HMMs. This is because in the new training data has a different gesture sequence as to compare with the hand labeled GT, i.e. one rest segment could be two or more segment in the same period of time. Hence, transitions from gesture to gesture may increase or decrease. Lastly, the Viterbi algorithm is used as described in the previous chapter to find the most probable path. The final evaluation is done by comparing the Viterbi state sequence and the hand labeled GT.

4.3 Results

The transition probabilities found for HMM1 are shown in table 4.1. The training data shows over-segmentation, i.e. there are two or more segments for one gesture segment in the hand labeled data. Some gestures of rest, utensiling, drink, and other tend to be long, so more transitions to themselves will occur, thus having an increase in the transition probability. We can see that phenomenon in the table as compared to table 3.3. Even with over-segmentation, we can still see consistencies in the transitions from a gesture to another gestures as in transitioning to bite from utensiling and transitioning to rest from bite.

Table 4.1: Transition probabilities.

From \ To	Rest	Utensiling	Bite	Drink	Other
Rest	0.311	0.241	0.290	0.071	0.087
Utensiling	0.094	0.379	0.483	0.005	0.040
Bite	0.316	0.304	0.235	0.017	0.128
Drink	0.198	0.150	0.092	0.401	0.159
Other	0.137	0.104	0.126	0.084	0.549

All classifiers were trained and tested using leave-one-out cross validation. For all classifiers, accuracy was measured as the total percentage of time in all the meals that gestures were labeled correctly. Table 4.2 presents a summary of the accuracy achieved by each of the classifiers. For the baseline classifiers, the sub-gesture HMMs performs better than the KNN with an 11.2% improvement. Our gesture-to-gesture classifier that incorporates gesture history shows further imporvement, from 6.8% to 18.9% across HMM1 to HMM6.

Table 4.3 shows the results broken down by gesture type. For utensiling, bite, and other,

Table 4.2: Total recognition accuracy.

Method	Accuracy (%)
KNN	59.9
Sub-gesture HMM	71.1
HMM1	77.9
HMM2	78.7
HMM3	79.5
HMM4	81.4
HMM5	84.5
HMM6	90.0

Table 4.3: Recognition accuracy for each gesture.

Method	Rest (%)	Utensiling (%)	Bite (%)	Drink (%)	Other (%)
KNN	89.7	68.0	56.5	37.2	4.2
Sub-gesture HMM	85.7	73.9	75.3	85.5	28.1
HMM1	84.3	84.3	78.9	77.7	49.3
HMM2	84.3	85.1	79.6	81.4	50.4
HMM3	85.3	85.5	79.6	84.6	51.4
HMM4	87.0	84.2	81.6	90.9	58.4
HMM5	89.7	87.0	85.4	87.9	65.4
HMM6	95.3	91.9	89.6	87.2	76.7

the gesture-to-gesture HMMs show cumulative improvement of recognition as each level of history is incorporated. In the case of drinks, the improvement peaks at a history of 4, which is consistent with recognition of gestures using hand labeled data in the previous chapter. In the case of rest, the KNN and HMM5 are comparable but HMM6 shows even better accuracy, which suggests that modeling a history greater than 5 preceding gestures the rest gesture can be recognize better within the framework of fragmenting the wrist motion signal into small overlapping frames.

A confusion matrix of the gestures was created based on HMM6. Table 4.4 shows the confusion matrix for each gesture. The greatest confusion for the word rest is utensiling, which may be because small tremors and posture adjustments are allowed in the rest definition. Thus during the preliminary segmentation these get separated from a rest period. The utensiling and bite shows the greatest confusion between each other. We believe this is because utensiling can have several hand motiones that are similar to bites, hence causing the confusion. In the case of a drink gesture, there can be a brief rest period during the action of liquid intake (while the container is held to the mouth), which may cause over-segmentation of a drink gesture with a rest in-between. The other category shows a fairly even confusion between rest, utensiling, and bite. The other gestures have duration greater than 4 sec, among these gestures the average time is 8.6 sec and the maximum other gesture is 42 sec. Since the other gesture is an catch-all gesture this gives the possibility that several pieces get confused with the other gestures.

Table 4.4: HMM6 confusion matrix.

Actual \ Classifier	Rest (%)	Utensiling (%)	Bite (%)	Drink (%)	Other (%)
Rest	95.8	1.7	1.3	0.2	1.1
Utensiling	2.4	91.9	4.3	0.0	1.4
Bite	1.6	4.5	91.5	0.2	2.2
Drink	6.2	0.9	2.3	88.9	1.7
Other	7.0	7.7	6.7	0.0	78.6

4.4 Conclusions

In this chapter, we automatically classified and segmented eating gestures from a continuous wrist motion signal. The process involved fragmenting the wrist motion signal into small overlapping frames where features were computed and fed into the classifiers. The classifiers labeled each frame as a specific gesture. The segmentation of eating gestures is finalized by grouping consecutive frames that pertain to the same gesture. These gesture sequences were compared with the GT hand labeled data and the accuracy was calculated on the total time gestures matched. Three classifiers were studied: KNN, sub-gesture HMM, and gesture-to-gesture HMM. The first two are baseline algorithms; after classifying and segmenting, these showed an accuracy of 59.9% and 71.71%, respectively. The gesture-to-gesture HMM incorporates knowledge of the inter-gesture sequential dependencies. The accuracy obtained with this method ranged from 77.9% to 90.0%. Using hand labeled data of the 25 meals, the highest accuracy of recognizing eating gestures through the gesture-to-gesture HMMs was 96.5%. The unsegmented case does 6.5% worse than the hand labeled data, nevertheless it can be observed that by adding sequential dependence knowledge in both cases the recognition of eating gestures is improved.

There are some limitations to this work. The gesture-to-gesture HMMs are dependent of the segmentation output provided by the sub-gesture HMMs implemented here. Once gesture sequences are available, these are relabeled using the hand labeled GT data to use as training data. Hence, there are several stages the gesture sequence must go through, these may delay a system to provide feedback in real-time. Another drawback is that the automatic classification and segmentation in this work are offline methods which need the complete wrist motion signal.

From this work we conclude that the segmentation and recognition of the our eating gestures from a continuous wrist motion signal can be accomplished. The performance increases if knowledge

of the history of actions is incorporated into the classifier. More work will move towards minimizing the stages for the gesture-to-gesture HMM for practical implementations. Also, the process will need to be adjusted to be online to present immediate feedback from the device to the user.

Chapter 5

Conclusions

This work is motivated by recent advances in body sensing and mobile health technology that have created new opportunities for empowering people to take a more active role in managing their health [97]. We are researching methods to build a watch-like tool that tracks wrist motion to help automatically monitor dietary intake [56, 58, 57, 59]. Obesity now afflicts one in three adults and one in six children in the United States [69, 122] and has been recognized as a major health problem in need of new tools [63, 112, 169]. Self-monitoring of dietary intake, body weight and physical activity have been consistently found to be associated with successful weight loss and maintenance [34]. However, food diaries and other tools currently used for monitoring dietary intake require users to manually estimate and record energy intake, making them prone to error and difficult to use for long periods of time [168]. Body-worn sensors offer the opportunity to automatically track dietary intake [22, 144, 129].

Prior work by this research group has demonstrated methods for detecting periods of eating (i.e. meals, snacks) by tracking wrist motion continuously all day [57, 59], and for detecting and counting bites taken during a meal [56, 58]. This dissertation considers the problem of recognizing eating gestures by tracking wrist motion. Eating gestures are activities commonly undertaken during the consumption of a meal, such as taking a bite of food, sipping a drink of liquid, or using utensils to cut food. Each of these gestures causes a pattern of wrist motion that can be tracked to automatically identify the activity. Previous works have studied this problem at the level of a single gesture [21, 17, 85]. This work demonstrates that individual gestures have sequential dependence. For example, the action of using utensils to cut food is typically followed by the action of taking a bite

of food, which is typically followed by the action of resting the wrist during mastication of the food. We have developed hidden Markov models (HMMs) that model these sequential dependencies, and demonstrate that they improve recognition accuracy compared to simpler classifiers.

In chapter 2 we identified the a set of gestures related to eating activities. For this work we defined four actions related to eating: *rest*, *utensiling*, *bite*, and *drink*. All other actions for which intent was not defined, including both eating and non-eating activities (e.g. gesturing while talking, cleaning with a napkin, waving at a friend, etc.) are referred to as *other*. The definitions of these are based on discernible user intent, i.e. the subject's intent is determined by observing the hand wearing the device. The duration of an action lasts from when the intent can first be observed in the synchronized video, to when that intent has ended. Using three meals and five raters, an inter-rater reliability study was conducted to determine the stability of the gesture definitions and hand labelings, finding 90.7% total time agreement. Although four words is a fairly limited set, we found that they comprised on average 86.9% of the time during a meal.

In chapter 3, using hand labeled gesture sequences from 25 meals, we demonstrated that individual gestures have sequential dependence. Three types of classifiers were built: 1) a K-nearest neighbor classifier which uses no sequential context, 2) a hidden Markov model (HMM) which captures the sequential context of sub-gesture motions, and 3) HMMs that model inter-gesture sequential dependencies. We built first-order to sixth-order HMMs to evaluate the usefulness of increasing amounts of sequential dependence to aid recognition. We found that the baseline accuracies for the KNN and the sub-gesture HMM classifiers were 75.8% and 84.3%, respectively. Using HMMs that model inter-gesture sequential dependencies, we were able to increase accuracy to up to 96.5%. These results demonstrate that sequential dependencies exist between eating gestures and that they can be exploited to improve recognition accuracy. Among the limitations to this point of our work include the number of activities modeled and the use of manually segmented data. We found that 15.9% of the total time during meals is comprised of other activities, such as wiping hands on a napkin or gesturing while talking. Increasing the number of activities may affect overall classifier accuracy. However, we still expect that modeling sequential dependencies would produce an improvement in accuracy.

In chapter 4 we were able to automatically classify and segment our wrist motion gestures from unsegmented data. Initially, a wrist motion signal is fragmented into overlapping frames computing features in this period of time. Each frame was classified as part of a particular gesture.

The final gesture sequence segmentation is achieved by grouping consecutive frames pertaining to the same gesture. We used the three classifiers from chapter 3 (KNN, sub-gesture HMMs, and gesture-to-gesture HMMs). In the case of the sub-gestures HMMs, segmentation and classification of wrist gestures is accomplished by connecting the individual HMMs together. For the gesture-to-gesture HMMs, it uses the output from the sub-gesture HMM as a preliminary segmented data. These segments are relabeled using the hand labeled data and serve as input. The final classification follows the two-step process explained in chapter 3. For evaluation, the segmented output from the three methods is compared with the hand labeled data. Our baseline algorithms achieved a 59.9% and 71.1%, respectively. The gesture-to-gesture HMMs showed an accuracy of 90.0%. Working with unsegmented data does 6.5% worse than the hand labeled data, nevertheless it can be observed that by adding sequential dependence knowledge in both cases the recognition of eating gestures is improved.

5.0.1 Future work

This dissertation has shown the existence of sequential dependence among wrist motion gestures during the consumption of a meal. The results obtained in this work can point several directions for future research projects. The analysis of gestures may help identify if a lower or higher calorie meal is being consumed. For example, eating a salad may have a faster movement from the wrist when collecting food from the plate or bowl, and a steak may have less movement since the eater needs time to masticate and swallow the portions. At the same time, knowing the type of meal could help to automatically adjust the parameters for calorie intake and provide immediate feedback to the user. Another application could be for studying eating habits in humans. This could benefit people that suffer from obesity or some types of eating disorders by providing an objective and automated measure of within-meal eating patterns.

Appendices

Appendix A

Curriculum Vitae

Raul I. Ramos-Garcia

E-mail: gramos@clemson.edu, raul_r64@hotmail.com

Areas of Interest

Signal Processing, Body Worn Sensors, Embedded Systems, Computer Vision and Pattern Recognition, Machine Learning, Biometrics, Speech and Speaker Recognition.

Education

- **Ph.D. in Electrical Engineering** May 2014
  Clemson University, Clemson, SC, USA
- **B.S. in Electronic Engineering (Highest Distinction)** December 2006
  Instituto Tecnologico Superior de Cajeme, Obregon, Sonora, Mexico
- **Computer Technician (Highest Distinction)** June 2001
  Centro de Bachillerato Tecnologico Industrial y de Servicios No. 64, Navojoa, Sonora, Mexico

Professional Experience

- **Research Assistant, BMW Manufacturing, Co.**, Spartanburg, SC, USA May 2013 - May 2014
  - Project development *"Visual Inspector Assistant"*. Developed image processing and computer vision algorithms (C/C++, Win32 API, OpenCV) to detect and provide immediate feedback of defects on cars for BMW Plant Spartanburg, SC, US. This project was invited to be presented at the BMW Plant Munich Innovation Fair and at the BMW FIZ - Forschungs- und Innovationszentrum in Munich, Germany. Other main collaborators are: Dr. Adam Hoover (Clemson University), Dr. Jörg Schulte (BMW Plant Spartanburg), Junghil Kwon (Clemson University).
- **Internship, Federal Electricity Commission (CFE)**,
  Obregon, Sonora, Mexico August - December 2006
  - Project development *"Automatic Aperture Detection for Primary Equipment"*. This project consisted on design of software and hardware interface using Visual Studio and micocontrollers (PIC) with the objective to sense and monitor electric outage in several sectors in Obregon City. This project was presented at the 2006 CFE Entreprenuer Event held in the city of Hermosillo, Sonora (Mexico). The project was developed for the Distribution Engineering Department.

- Project support and improvement *"CFEACTIVAME"*. This was an existence project designed for Client Support Department. My contribution was to perform improvements in the assembly code for the microcontrollers to extend the capabilities of the system.

- **Internship, GE Security Navojoa Plant**, Navojoa, Sonora, Mexico May - August 2005
  - Project development *"Control of Tools and Equipment"* and *"Register of Anti-static Protection"*. These projects consisted of a software designed for the log of in and out of tools and anti-static protection garments provided to task forces workers. These projects were designed for the Quality Department.
  - I also contributed to the development, inspection, and testing of reference boards for the Electric Testing Department.

Teaching Experience

- **Clemson University**, Clemson, SC, USA
  Graduate teaching assistant:
  - ECE 202 Electric Circuits I Fall 2011
  - ECE 307 Basic Electrical Engineering Fall 2011, Spring 2012
  - ECE 467/667 Intro. to Digital Signal Processing Fall 2012
  - ECE 309 Electrical Engineering Laboratory I (3 sections) Spring 2013
- **Universidad de Sonora**, Navojoa, Sonora, Mexico
  Lecturer for the following undergraduate courses:
  - 6989 Statistics I Spring 2008
  - 7794 Fluid Mechanics Spring 2008
  - 8733 Systems Development II Spring 2008
  - 8730 Computational Logic Spring 2008
  - 8791 Applied Informatics II Spring 2008

Publications

Refereed Journal Publications

[J01] **Ramos-Garcia, R.I.**, Muth, E.R., Gowdy, J.N., Hoover, A.W., *"Improving the Recognition of Eating Gestures Using Inter-Gesture Sequential Dependencies"*, in IEEE Journal of Biomedical and Health Informatics (under review, reference number: JBHI-00061-2014).

Conference Proceedings (Reviewed)

[C01] **Ramos-Garcia, R.I.**, Hoover, A.W., *"A Study of Temporal Action Sequencing During Consumption of a Meal"*, in the proc. of the ACM Conf. on Bioinformatics, Computational Biology, and Biomedical Informatics, Washington, DC, September 22-25, 2013.

Other Scholarly Publications

[P03] **Ramos-Garcia, R.I.**, Hoover, A.W., *"A Study of Sequential Actions at the Meal Level"*. Graduate student poster session, 2013 SACNAS National Conference "Strengthening the Nation Through Diversity, Innovation and Leadership in STEM", San Antonio, TX, October 3-6, 2013.

[P02] **Ramos-Garcia, R.I.**, Gowdy, J.N., *"Development of an Eye Tracker Using the Kalman Filter with Outlier Removal for Use in a Visual Speech Recognition System"*. Graduate student poster session, 2012 SACNAS National Conference "Science, Technology, and Diversity for a Healthy World", Seattle, WA, October 11-14, 2012.

[P01] **Ramos-Garcia, R.I.**, Shu, D., *"Vision Based Vehicle Tracking Using a High Angle Camera"*. Technical report, Clemson University, Clemson, SC, 2010.

Papers Presented/Simposia/Invited Lecturers/Professional Meetings/Workshops

[A08] **Ramos-Garcia, R.I.**, Antani, K., *"The Ultimate Success Factor: Quality Research by Clemson University"*. BMW FIZ - Forschungs- und Innovationszentrum, Munich, Germany. (April 8-9, 2014).

[A07] Schulte, J., **Ramos-Garcia, R.I.**, *"Visual Inspector Assistant"*. Innovationsmesse, BMW Plant Munich, Munich, Germany. (April 4, 2014).

[A06] **Ramos-Garcia, R.I.**, *"Examples of hidden Markov models applications"*. Guest lecturer in Analysis of Robotic Systems (ECE 854) taught by Dr. Adam W. Hoover at The Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. (November 19, 2013).

[A05] **Ramos-Garcia, R.I.**, *"An introduction to hidden Markov models (HMM)"*. Guest lecturer in Analysis of Robotic Systems (ECE 854) taught by Dr. Adam W. Hoover at The Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. (November 14, 2013).

[A04] **Ramos-Garcia, R.I.**, *"Isolated word recognizer using HTK"*. Guest lecturer in Digital Processing of Speech Signals (ECE 846) taught by Dr. John N. Gowdy at The Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. (November 2012).

[A03] **Ramos-Garcia, R.I.**, *"Examples of hidden Markov models applications"*. Guest lecturer in Analysis of Robotic Systems (ECE 854) taught by Dr. Adam W. Hoover at The Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. (November 2012).

[A02] **Ramos-Garcia, R.I.**, *"An introduction to hidden Markov models (HMM)"*. Guest lecturer in Analysis of Robotic Systems (ECE 854) taught by Dr. Adam W. Hoover at The Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. (November 2012).

[A01] **Ramos-Garcia, R.I.**, *"Developing GUIs with MATLAB"*. Guest lecturer in Digital Processing of Speech Signals (ECE 846) taught by Dr. John N. Gowdy at The Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC. (September 2011).

Workshops and Certified Courses

[B08] **Certified course: Digital Signal Processing (with distinction)**. Instructor(s): Dr. Paolo Prandoni (École Polytechnique Fédérale de Lausanne) and Dr. Martin Vetterli. (École Polytechnique Fédérale de Lausanne). Online: www.coursera.org. (February - April 2013)

[B07] **Certified course: Machine Learning**. Instructor(s): Dr. Andrew Ng (Stanford University). Online: www.coursera.org. (October - December 2011)

[B06] **Certified course: Artificial Intelligence**. Instructor(s): Dr. Sebastian Thrun(Stanford University) and Dr. Peter Norving (Stanford University). Online: www.coursera.org. (October - December 2011)

[B05] **Certified course: CISCO Systems** (CNNA1 Networks LAN Basics, CNNA2 Routers and Routing Basics, CNN3 Switching Basics and Intermediate). Instituto Tecnologico Superior de Cajeme, Sonora, Mexico. (August 2005 - December 2006)

[B04] **Workshop(s): Virtual Reality**. 21st International Symposium in Electronics Communications: RF Tech. Instituto Tecnologico de Monterrey, Nuevo Leon, Mexico. (March 2006)

[B03] **Workshop(s): PLC Programming, Cell-Phone Maintenance**. ITESCA 1st Electronic Engineering Week. Instituto Tecnologico Superior de Cajeme, Sonora, Mexico. (May 2005)

[B02] **Workshop(s): VHDL**. 4th Symposium in Computer and Electronic Engineering. Instituto Tecnologico Superior de Cajeme, Sonora, Mexico. (November 2004)

[B01] **Workshop(s): Matlab Programming**. 3rd Symposium in Computer and Electronic Engineering. Instituto Tecnologico Superior de Cajeme, Sonora, Mexico. (November 2003)

Awards and Honors

1. **SACNAS Travel Scholarship Award**. Attended 2013 SACANAS National Conference to present a poster at the Graduate Student Poster Session. (October 2013)
2. **Graduate Research Assistanship**. Developed image and computer vision software to automatic detect defects on cars for BMW Manufacturing Co., SC, USA. (May 2013)
3. **SACNAS Travel Scholarship Award**. Attended 2012 SACANAS National Conference to present a poster at the Graduate Student Poster Session. (October 2012)
4. **Recipient of a CONACYT Fellowship**. This fellowship (4 years funding) was provided by the Consejo Nacional de Ciencia y Tecnologia (CONACYT, www.conacyt.gob.mx) to pursue a Ph.D. in Electrical Engineering at Clemson University, Clemson, SC, US. (August 2008)
5. **Recipient of ITESCA Undergaduate Academic Excellence Award**. Recognition for outstanding academic work of the Electronic Engineering class of 2002-2006 at the Instituto Tecnologico Superior de Cajeme (ITESCA), Obregon, Sonora, Mexico. (December 2007)
6. **1st Place at the ITESCA Entrepreneur and Business Fair Competition**. Title: "PC-Link". A business plan was designed to sell and market a PlayStation controller adapter for PC games. The project was also presented at the 11th Entrepreneur National Event held at the Instituto Tecnologico de Acapulco (Guerrero, Mexico) in June 2005.(November 2004)
7. **ITESCA Mathematics Rally Award**. Instituto Tecnologico Superior de Cajeme, Obregon, Sonora, Mexico. (November 2002)
8. **2nd Place (state phase), III State Competition of Basic Sciences**. Instituto Tecnologico de Hermosillo, Hermosilo, Sonora, Mexico. (November 2001)
9. **Recipient of the CBTis No. 64 Academic Excellence Award of the Class of 1998-2001**. Centro de Bachillerato Tecnologico Industrial y de Servicios No. 64, Navojoa, Sonora, Mexico. (July 2001)
10. **Recipient of a Computer Science Field Award of the Class of 1998-2001**. Centro de Bachillerato Tecnologico Industrial y de Servicios No. 64. (July 2001)
11. **2nd Place (state phase), Software Development Prototypes Competition**. Title: "Electric Circuits Software". 7th Presentation of Technological Prototypes, Didactic, and Software Development, Hermosillo, Sonora, Mexico. (June 2001)
12. **1st Place (regional phase), Software Development Prototypes Competition**. Title: "Electric Circuits Software". This software was coded using Visual Basic with the objective to create a learning tool of theory and simulation of electric circuits for high school students. Centro de Bachillerato Tecnologico Industrial y de Servicios No. 64, Navojoa, Sonora, Mexico. (May 2001)
13. **1st Place (regional phase), III State Competition of Basic Sciences**. Instituto Tecnologico de Huatabampo, Sonora, Mexico. (May 2001)

Leadership and Public Service

- **Tutor for Bridges for A Brighter Future Program** (www.bridgestoabrighterfuture.org). Through Clemson SACNAS Chapter of Clemson University, I provided tutoring in mathematics for high school students at Furman University, Greenville, SC. (August 2012 - May 2014)
- **President of the Electronic Engineering Student Society**. I coordinated and led meetings to establish a work plan and academic events for the electronic engineering students at the Instituto Tecnologico Superior de Cajeme (ITESCA), Obregon, Sonora, Mexico. (May 2005 - May 2006)

Skills

Software
Windows, Linux, C, C++, Win 32 API, Java, Python, Visual Basic, Assembly, OpenCV (image processing/computer vision), Blepo (image processing/computer vision), MATLAB, Octave, R, Hidden Markov Model Toolkit (speech processing and temporal data analysis tool), LaTex (document preparation system), Microsoft Office.

Languages

- Spanish (native language)
- English (native or bilingual proficiency)

Membership

Student member

Institute of Electrical and Electronics Engineers (IEEE)	2010, 2013-2014
IEEE Signal Processing Society (IEEE SPS)	2010, 2013-2014
IEEE Systems, Man, and Cybernetics Society (IEEE SMCS)	2013-2014
Society for Advancement of Chicanos and Native Americans in Science (SACNAS)	2012-2013

Bibliography

[1] Stmicroelectronics lis344alh: Mems inertial sensor. [Online]. Available: http://www.st.com/stonline/products/literature/ds/14337.pdf

[2] Stmicroelectronics lpr410al: Mems motion sensor. [Online]. Available: http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/CD00254123.pdf

[3] Stmicroelectronics lpy410al: Mems motion sensor. [Online]. Available: http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/CD00254136.pdf

[4] Harvard School of Public Health. [Online]. Available: http://www.hsph.harvard.edu/obesity-prevention-source/obesity-trends/

[5] National Center for Health Statistics. National Health and Nutrition Examination Survey. Questionnaires, datasets, and related documentation. [Online]. Available: http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm

[6] U.S. Department of Health and Human Services. Final review, Healthy People 2010: Nutrition and overweight. [Online]. Available: http://www.cdc.gov/nchs/data/hpdata2010/hp2010_final_review_focus_area_19.pdf

[7] World Health Organization. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs311/en/

[8] K. J. Acheson, I. T. Campbell, O. G. Edholm, D. S. Miller, and M. J. Stock, "The measurement of food and energy intake in man-an evaluation of some techniques." *The American Journal of Clinical Nutrition*, vol. 33, no. 5, pp. 1147–54, 1980.

[9] S. Adibi, "Link technologies and blackberry mobile health (mhealth) solutions: A review," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 586–597, 2012.

[10] D. W. Aha and R. L. Bankert, "A comparative evaluation of sequential feature selection algorithms," in *Learning from Data: Artificial Intelligence and Statistics V*, ser. Lecture Notes in Statistics. Springer-Verlag, 1996, ch. 4, pp. 199–206.

[11] M. J. Ahamed, R. Ben-Mrad, and P. Sullivan, "A drop-on-demand-based electrostatically actuated microdispenser," *Journal of Microelectromechanical Systems*, vol. PP, no. 99, pp. 1–9, 2012.

[12] M. Ahmadi and G. Jullien, "A wireless-implantable microsystem for continuous blood glucose monitoring," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 3, no. 3, pp. 169–180, 2009.

[13] A. Akl, C. Feng, and S. Valaee, "A novel accelerometer-based gesture recognition system," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6197–6205, December 2011.

[14] R. Almaghrabi, G. Villalobos, P. Pouladzadeh, and S. Shirmohammadi, "A novel method for measuring nutrition intake based on food image," in *2012 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2012, pp. 366–370.

[15] R. Alvarez, "The promise of e-health - a canadian perspective," *eHealth International*, vol. 1, no. 1, p. 4, 2002.

[16] O. Amft, "A wearable earpad sensor for chewing monitoring," in *2010 IEEE Sensors*, November 2010, pp. 222–227.

[17] O. Amft, D. Bannach, G. Pirkl, M. Kreil, and P. Lukowicz, "Towards wearable sensing-based assessment of fluid intake," in *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, April 2010, pp. 298–303.

[18] O. Amft, *Ambient, on-body, and implantable monitoring technologies to assess dietary behaviour*, 2011th ed., ser. International Handbook of Behavior, Food and Nutrition. Springer, 2011, no. 38, p. 3507–3526.

[19] O. Amft, H. Junker, and G. Tröster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *ISWC 2005: IEEE Proceedings of the Ninth International Symposium on Wearable Computers.*, IEEE Press. IEEE Press, October 2005, p. 160–163.

[20] O. Amft, M. Stger, and G. Trster, "Analysis of chewing sounds for dietary monitoring," in *In UbiComp 2005*. Springer, 2005, pp. 56–72.

[21] O. Amft and G. Tröster, "Recognition of dietary activity events using on-body sensors," *Artificial Intelligence in Medicine*, vol. 42, no. 2, pp. 121–136, 2008.

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

[23] U. Anliker, J. Ward, P. Lukowicz, G. Troster, F. Dolveck, M. Baer, F. Keita, E. Schenker, F. Catarsi, L. Coluccini, A. Belardinelli, D. Shklarski, M. Alon, E. Hirt, R. Schmid, and M. Vuskovic, "Amon: a wearable multiparameter medical monitoring and alert system," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 415–427, 2004.

[24] M. Armenise, *Advances in Gyroscope Technologies*. Springer, 2011. [Online]. Available: http://books.google.com/books?id=lJUiyigJRBgC

[25] H. Badawi, M. Eid, and A. El Saddik, "Diet advisory system for children using biofeedback sensor," in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, May 2012, pp. 1–4.

[26] F. Bellisle, B. Guy-Grand, and J. L. Magnen, "Chewing and swallowing as indices of the stimulation to eat during meals in humans: effects revealed by the edogram method and video recordings," *Neuroscience and Biobehavioral Reviews*, vol. 24, no. 2, pp. 223–228, 2000.

[27] S. Berman and H. Stern, "Sensors for gesture recognition systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 3, pp. 277–290, May 2012.

[28] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[29] M. Booth, N. C. for Overweight, Obesity, M. Ragg, N. S. W. D. of Health, and N. Health, *NSW Schools Physical Activity and Nutrition Survey (SPANS) 2004: Summary Report*. NSW Department of Health, 2006.

[30] S. Bowman, M. Lino, S. Gerrior, and P. Basiotis, "The healthy eating index," *Center for Nutrition Policy and Promotion (U.S.)*, pp. 1–28, 1995.

[31] C. R. P. Brian Wansink, "Eating behavior and obesity at chinese buffets," *Obesity*, no. 8, p. 19571960, 2008.

[32] F. Bruhn, F. Carsey, J. Kohler, M. Mowlem, C. German, and A. Behar, "Mems enablement and analysis of the miniature autonomous submersible explorer," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 1, pp. 165–178, January 2005.

[33] T. Bui and L. Nguyen, "Recognizing postures in vietnamese sign language with mems accelerometers," *IEEE Sensors Journal*, vol. 7, no. 5, pp. 707–712, May 2007.

[34] L. E. Burke, J. Wang, and M. A. Sevick, "Self-monitoring in weight loss: A systematic review of the literature," *Journal of the American Dietetic Association*, vol. 111, no. 1, pp. 92–102, 2011.

[35] B. Caballero, "The global epidemic of obesity: An overview," *Epidemiologic Reviews*, vol. 29, no. 1, pp. 1–5, 2007.

[36] S. Cadavid, M. Abdel-Mottaleb, and A. Helal, "Exploiting visual quasi-periodicity for real-time chewing event detection using active appearance models and support vector machines," *Personal and Ubiquitous Computing*, pp. 1–11, June 2011.

[37] C. J. Caspersen, K. E. Powell, and G. M. Christenson, "Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research." *Public health reports (Washington, D.C. : 1974)*, vol. 100, no. 2, pp. 126–131, 1985.

[38] C. M. Champagne, G. A. Bray, J. P. Delany, A. A. Kurtz, J. B. R. Monteiro, E. Tucker, and J. Volaufova, "Energy intake and energy expenditure: a controlled study comparing dietitians and non-dietitians," *Journal of the American Dietetic Association*, vol. 120, no. 10, pp. 1428–1432, 2002.

[39] H.-C. Chen, W. Jia, Z. Li, Y.-N. Sun, and M. Sun, "3d/2d model-to-image registration for quantitative dietary assessment," in *2012 38th Annual Northeast Bioengineering Conference (NEBEC)*, March 2012, pp. 95–96.

[40] W.-C. Chen and R.-Y. Lyu, "A hmm-based fundamental motion synthesis approach for gesture recognition on a nintendo triaxial accelerometer," in *2011 5th International Conference on Signal Processing and Communication Systems (ICSPCS)*, December 2011, pp. 1–5.

[41] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, "Hand gesture recognition research based on surface emg sensors and 2d-accelerometers," in *2007 11th IEEE International Symposium on Wearable Computers*, October 2007, pp. 11–14.

[42] T. Chomutare, "A context-sensitive framework for mobile terminals for assisting Type 2 diabetes patients," Master Thesis, Department of Computer Science, University of Tormso, 2008.

[43] N. Cleven, J. Muntjes, H. Fassbender, U. Urban, M. Gortz, H. Vogt, M. Grafe, T. Gottsche, T. Penzkofer, T. Schmitz-Rode, and W. Mokwa, "A novel fully implantable wireless sensor system for monitoring hypertension patients," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 11, pp. 3124–3130, 2012.

[44] P. Cohen, D. McGee, S. Oviatt, L. Wu, J. Clow, R. King, S. Julier, and L. Rosenblum, "Multimodal interaction for 2d and 3d environments [virtual reality]," *Computer Graphics and Applications, IEEE*, vol. 19, no. 4, pp. 10–13, July/August 1999.

[45] J. L. Colquitt, J. Picot, E. Loveman, and A. J. Clegg, *Surgery for obesity*, 2nd ed. John Wiley & Sons, Ltd, 2009.

[46] J. Conway, M. Irwin, and B. Ainsworth, "Estimating energy expenditure from the minnesota leisure physical activity and tecumesh occupational activity questionnaries-a doubly labeled water validation," *Journal of Clinical Epidemiology*, vol. 55, no. 4, pp. 293–299, 2002.

[47] K. A. Coon, J. Goldberg, B. L. Rogers, and K. L. Tucker, "Relationships between use of television during meals and children's food consumption patterns," *Pediatrics*, vol. 107, no. 1, p. e7, 2001.

[48] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[49] A. M. Coulston and C. J. Boushey, *Nutrition in the Prevention and Treatment of Disease*. Academic Press, 2008.

[50] N. Day, N. McKeown, M. Wong, A. Welch, and S. Bingham, "Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium," *International Journal of Epidemiology*, vol. 30, no. 2, pp. 309–317, 2001.

[51] D.-M. Denk, H. Swoboda, and E. Steiner, "Physiology of the larynx," *Der Radiologe*, vol. 38, no. 2, pp. 63–70, 1998.

[52] K. Dermitzakis, A. Arieta, and R. Pfeifer, "Gesture recognition in upper-limb prosthetics: A viability study using dynamic time warping and gyroscopes," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, September 2011, pp. 4530–4533.

[53] J. Dixon, "The effect of obesity on health outcomes," *Molecular and Cellular Endocrinology*, vol. 316, no. 2, pp. 104–108, 2010.

[54] B. Dong and S. Biswas, "Swallow monitoring through apnea detection in breathing signal," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 6341–6344.

[55] Y. Dong, "Tracking Wrist motion to Detect and Measure the Eating Intake of Free-Living Humans," PhD dissertation, Electrical and Computer Engineering Department, Clemson University, SC., 2012.

[56] Y. Dong, A. Hoover, and E. Muth, "A device for detecting and counting bites of food taken by a person during eating," in *IEEE Int'l Conf on Bioinformatics and Biomedicine*, 2009, pp. 265–268.

[57] Y. Dong, A. Hoover, E. Muth, and J. Scisco, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Applied Psy-chophysiology and Biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[58] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "Detecting eating using a wrist mounted device during normal daily activities," in *Proceedings of the 9th International Conference on Embedded Systems and Applications*, 2011.

[59] Y. Dong, J. Scisco, M. Wilson, E. Muth, and A. Hoover, "Detecting periods of eating during free-living by tracking wrist motion," *IEEE Journal of Biomedical and Health Informatics*, in press.

[60] M. Drennan, "An Assessment of Linear Wrist Motion during the Tracking of a Bite of Food," Master Thesis, Eclectrical and Computer Engineering Department, Clemson University, 2010.

[61] B. Edelman, D. Engell, P. Bronstein, and E. Hirsch, "Environmental effects on the intake of overweight and normal-weight men," *Appetite*, vol. 7, no. 1, pp. 71–83, 1986.

[62] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 20–26, January 2008.

[63] A. Ershow, A. Ortega, J. Baldwin, and J. Hill, "Engineering approaches to energy balance and obesity: Opportunities for novel collaborations and research: Report of a joint National Science Foundation and National Institutes of Health workshop," *Journal of Diabetes Science and Technology*, vol. 1, no. 1, pp. 96–105, 2007.

[64] V. G. Fabiano-Alves, E. E. Moreira-da Rocha, M. C. Gonzalez, R. B. Vieira-da Fonseca, M. H. do Nascimento-Silva, and C. A. Chiesa, "Assessement of resting energy expenditure of obese patients: Comparison of indirect calorimetry with formulae," *Clinical Nutrition*, vol. 28, no. 3, pp. 299–304, 2009.

[65] C. Fester, N. J.I., and E. Levitt, "The control of eating," *Journal of Mathetics*, vol. 1, pp. 87–109, 1962.

[66] W. Fink, C. X. You, and M. A. Tarbell, "Microcomputer-based artificial vision support system for real-time image processing for camera-driven visual prostheses," *Journal of Biomedical Optics*, vol. 15, no. 1, p. 016013, 2010.

[67] E. Finkelstein, O. Khavjou, H. Thompson, J. Trogdon, L. Pan, B. Sherry, and W. Dietz, "Obesity and severe obesity forecasts through 2030," *American Journal of Preventive Medicine*, vol. 42, no. 6, pp. 563–570, 2012.

[68] E. A. Finkelstein, I. C. Fiebelkorn, and G. Wang, "National medical spending attributable to overweight and obesity: how much, and who's paying?" *Health Affairs Web Exclusives*, pp. 219–26, 2003.

[69] 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," *J. American Medical Association*, vol. 307, no. 5, pp. 491–497, 2012.

[70] J. M. Fontana and E. S. Sazonov, "A robust classification scheme for detection of food intake through non-invasive monitoring of chewing," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, September 2012, pp. 4891–4894.

[71] R. C. Foster, L. M. Lanningham-Foster, C. Manohar, S. K. McCrady, L. J. Nysse, K. R. Kaufman, D. J. Padgett, and J. A. Levine, "Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure," *Preventive Medicine*, vol. 41, no. 3-4, pp. 778–783, 2005.

[72] P. Galambos, J. Lantz, M. Baker, J. McClain, G. Bogart, and R. Simonson, "Active mems valves for flow control in a high-pressure micro-gas-analyzer," *Journal of Microelectromechanical Systems*, vol. 20, no. 5, pp. 1150–1162, October 2011.

[73] C. Gao, F. Kong, and J. Tan, "Healthaware: Tackling obesity with health aware smart phone systems," in *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO*, December 2009, pp. 1549–1554.

[74] C. Garcia Ulen, M. M. Huizinga, B. Beech, and T. A. Elasy, "Weight regain prevention," *Clinical Diabetes*, vol. 26, no. 3, pp. 100–113, 2008.

[75] D. G. Gaul, E. Craighead, and M. J. Mahoney, "Relationship between eating rates and obesity," *Journal of Consulting and Clinical Psychology*, vol. 43, pp. 123–125, 1975.

[76] J. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," *Speech and Audio Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 291–298, 1994.

[77] Y. Gianchandani and K. Najafi, "Batch fabrication and assembly of micromotor-driven mechanisms with multi-level linkages," in *IEEE International Conference on Micro Electro Mechanical Systems*, February 1992, pp. 141–146.

[78] M. Gillman, S. Rifas-Shiman, A. Frazier, H. Rockett, C. Camargo, A. Field, C. Berkey, and G. Colditz, "Family dinner and diet quality among older children and adolescents," *Archives of Family Medicine*, vol. 9, no. 3, pp. 235–240, 2000.

[79] M. I. Goran and M. S. Treuth, "Energy expenditure, physical activity, and obesity in children," *Pediatric Clinics of North America*, vol. 48, no. 4, pp. 931–953, 2001.

[80] M. M. Huizinga, "Weight-loss pharmacotherapy: A brief review," *Clinical Diabetes*, vol. 25, no. 4, pp. 135–140, 2007.

[81] I. Ioakimidis, M. Zandian, L. Eriksson-Marklund, C. Bergh, A. Grigoriadis, and P. Södersten, "Description of chewing and food intake over the course of a meal," *Physiology and Behavior*, vol. 104, no. 5, pp. 761–769, 2011.

[82] J. Jakicic, C. Winters, W. Lang, and R. Wing, "Effects of intermittent exercise and use of home exercise equipment on adherence, weight loss, and fitness in overweight women: A randomized trial," *JAMA*, vol. 282, no. 16, pp. 1554–1560, 1999.

[83] S. S. Jonnalagadda, D. C. Mitchell, H. Smiciklas-Wright, K. B. Meaker, N. V. Heel, W. Karmally, A. G. Ershow, and P. M. Kris-Etherton, "Accuracy of energy intake data estimated by a multiplepass, 24-hour dietary recall technique," *Journal of the American Dietetic Association*, vol. 100, no. 3, pp. 303–311, 2000.

[84] J. W. Judy, "Microelectromechanical systems (mems): fabrication, design and applications," *Smart Materials and Structures*, vol. 10, no. 6, p. 1115, 2001.

[85] 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.

[86] A. Kant, G. Block, R. Zielger, and M. Nestle, "Food group intake patterns and associated nutrient profiles of th US propulation," *Journal of American Diet Association*, vol. 91, pp. 1532–1537, 1991.

[87] D. L. Kaplan, "Eating style of obese and nonobese males," *Psychosomatic Medicine*.

[88] S. Kausar and M. Javed, "A survey on sign language recognition," in *Frontiers of Information Technology (FIT), 2011*, December 2011, pp. 95–98.

[89] N. L. Keim, C. A. Blanton, and M. J. Kretsch, "America's obesity epidemic: Measuring physical activity to promote an active lifestyle," *Journal of the American Dietetic Association*, vol. 104, no. 9, pp. 1398–1409, 2004.

[90] J. M. Kerver, E. J. Yang, S. Obayashi, L. Bianchi, and W. O. Song, "Meal and Snack Patterns Are Associated with Dietary Intake of Energy and Nutrients in US Adults," *Journal of the American Dietetic Association"*, vol. 106, no. 1, pp. 46–53, 2006.

[91] M. Khan, S. Ahamed, M. Rahman, and J.-J. Yang, "Gesthaar: An accelerometer-based gesture recognition method and its application in nui driven pervasive healthcare," in *2012 IEEE International Conference on Emerging Signal Processing Applications (ESPA)*, January 2012, pp. 163–166.

[92] B.-H. Kim, J.-B. Kim, and J.-H. Kim, "A highly manufacturable large area array mems probe card using electroplating and flipchip bonding," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 4, pp. 1079–1085, April 2009.

[93] T.-S. Kim, S.-S. Lee, Y. Yee, J.-U. Bu, C.-G. Park, and M.-H. Ha, "Large tilt angle electrostatic force actuated micro-mirror," *IEEE Photonics Technology Letters*, vol. 14, no. 11, pp. 1569–1571, November 2002.

[94] H. Kinnunen, M. Tanskanen, H. Kyrlinen, and K. Westerterp, "Wrist-worn accelerometers in assessment of energy expenditure during intensive training," *Physiological Measurement*, vol. 33, no. 11, p. 1841, 2012.

[95] F. Kong and J. Tan, "Dietcam: Regular shape food recognition with a camera phone," in *2011 International Conference on Body Sensor Networks (BSN)*, May 2011, pp. 127–132.

[96] V. Kosmidou and L. Hadjileontiadis, "Sign language recognition using intrinsic-mode sample entropy on semg and accelerometer data," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2879–2890, December 2009.

[97] S. Kumar, W. Nilsen, M. Pavel, and M. Srivastava, "Mobile health: Revolutionizing healthcare through transdisciplinary research," *Computer*, vol. 46, no. 1, pp. 28–35, 2013.

[98] N. Leenders, W. M. Sherman, H. N. Nagaraja, and C. L. Kein, "Evaluation of methods to assess physical activity in free-living conditions," *Medicine and Science in Sports and Exercise*, vol. 33, pp. 1233–1240, 2001.

[99] L. Li and D. Uttamchandani, "Design and evaluation of a mems optical chopper for fibre optic applications," *IEE Proceedings Science, Measurement and Technology*, vol. 151, no. 2, pp. 77–84, March 2004.

[100] Y. Li, X. Chen, X. Zhang, K. Wang, and Z. Wang, "A sign-component-based framework for chinese sign language recognition using accelerometer and semg data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2695–2704, October 2012.

[101] C.-W. Lin, Y.-T. Yang, J.-S. Wang, and Y.-C. Yang, "A wearable sensor module with a neural-network-based activity classification algorithm for daily energy expenditure estimation," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 5, pp. 991–998, September 2012.

[102] J. Liu, E. Johns, L. Atallah, C. Pettitt, B. Lo, G. Frost, and G.-Z. Yang, "An intelligent food-intake monitoring system using wearable sensors," in *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks (BSN*, May 2012, pp. 154–160.

[103] M. Livingstone, A. Prentice, J. Strain, W. Coward, A. Black, M. Barker, P. Mckenna, and R. Whitehead, "Accuracy of weighed dietary records in studies of diet and health." *BMJ: British Medical Journal*, vol. 300, no. 6726, p. 708, 1990.

[104] I. Lopes, B. Silva, J. Rodrigues, J. Lloret, and M. Proenca, "A mobile health monitoring solution for weight control," in *2011 International Conference on Wireless Communications and Signal Processing (WCSP)*, November 2011, pp. 1–5.

[105] P. Lopez-Meyer, R. Makeyev, S. Schuckers, E. L. Melanson, M. R. Neuman, and E. Sazonov, "Detection of food intake from swallowing sequences by supervised and unsupervised methods," *Annals of Biomedical Enigneering*, vol. 38, no. 8, pp. 2766–2774, 2010.

[106] Y. Ma, E. R. Bertone, E. J. Stanek, G. W. Reed, J. R. Hebert, N. L. Cohen, P. A. Merriam, and I. S. Ockene, "Association between eating patterns and obesity in a free-living us adult population," *American Journal of Epidemiology*, vol. 158, no. 1, pp. 85–92, 2003.

[107] D. J. Macfarlane, "Automated metabolic gas analysis systems: A review," *Sports Medicine*, vol. 31, no. 12, pp. 841–861, 2001.

[108] S. Mahabir, D. J. Baer, C. Giffen, B. A. Clevidence, W. S. Campbell, P. R. Taylor, and T. J. Hartman, "Comparison of energy expenditure estimates from 4 physical activity questionnaires with doubly labeled water estimates in postmenopausal women," *The American Journal of Clinical Nutrition*, vol. 84, no. 1, pp. 230–236, 2006.

[109] N. Maluf and K. Williams, *Introduction to microelectromechanical systems engineering*. Artech house publishers, 2004.

[110] J.-F. Mari, J.-P. Haton, and A. Kriouile, "Automatic word recognition based on second-order hidden markov models," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 22–25, Jan 1997.

[111] C. Martin, S. Kaya, and B. Gunturk, "Quantification of food intake using food image analysis," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, September 2009, pp. 6869–6872.

[112] B. McCabe-Sellers, "Advancing the art and science of dietary assessment through technology," *Journal of the American Dietetic Association*, vol. 110, no. 1, pp. 52–54, 2010.

[113] M. Meguid, A. Laviano, and F. Rossi-Fanelli, "Food intake equals meal size times mean number," *Appetite*, vol. 31, no. 3, p. 404, 1998.

[114] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, may 2007.

[115] L. S. Muhlheim, D. B. Allison, S. Heshka, and S. B. Heymsfield, "Do unsuccessful dieters intentionally underreport food intake?" *International Journal of Eating Disorders*, vol. 24, no. 3, pp. 259–266, 1998.

[116] K. Murao, T. Terada, A. Yano, and R. Matsukura, "Evaluating gesture recognition by multiple-sensor-containing mobile devices," in *2011 15th Annual International Symposium on Wearable Computers (ISWC)*, June 2011, pp. 55–58.

[117] A. Must, J. Spadano, E. Coakley, A. Field, G. Colditz, and W. Dietz, "The disease burden associated with overweight and obesity," *JAMA: The Journal of the American Medical Association*, vol. 282, no. 16, pp. 1523–1529, 1999.

[118] T. A. Nicklas, T. Baranowski, K. W. Cullen, and G. Berenson, "Eating patterns, dietary quality and obesity," *Journal of the American College of Nutrition*, vol. 20, no. 6, pp. 599–608, 2001.

[119] G. Niezen and G. Hancke, "Evaluating and optimising accelerometer-based gesture recognition techniques for mobile devices," in *AFRICON, 2009.*, September 2009, pp. 1–6.

[120] J. Nishimura and T. Kuroda, "Eating habits monitoring using wireless wearable in-ear microphone," in *3rd International Symposium on Wireless Pervasive Computing (ISWPC)*, May 2008, pp. 130–132.

[121] T. L. O'Connell, "An overview of obesity and weight loss surgery," *Clinical Diabetes*, vol. 22, no. 3, pp. 115–120, 2004.

[122] C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal, *Prevalence of obesity in the United States, 2009-2010.* US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2012.

[123] J. Oh, S.-J. Cho, W.-C. Bang, W. Chang, E. Choi, J. Yang, J. Cho, and D. Y. Kim, "Inertial sensor based recognition of 3-d character gestures with an ensemble classifiers," in *Ninth International Workshop on Frontiers in Handwriting Recognition*, October 2004, pp. 112–117.

[124] S. Ong and S. Ranganath, "Automatic sign language analysis: a survey and the future beyond lexical meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873–891, June 2005.

[125] A. Pantelopoulos and N. Bourbakis, "A survey on wearable biosensor systems for health monitoring," in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2008, pp. 4887–4890.

[126] J. Pärkkä, L. Cluitmans, and M. Ermes, "Personalization algorithm for real-time activity recognition using pda, wireless motion bands, and binary decision tree," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1211–1215, September 2010.

[127] S. Päßler and W.-J. Fischer, "Acoustical method for objective food intake monitoring using a wearable sensor system," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, May 2011, pp. 266–269.

[128] S. Päßler, W.-J. Fischer, and I. Kraljevski, "Adaptation of models for food intake sound recognition using maximum a posteriori estimation algorithm," in *2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, May 2012, pp. 148–153.

[129] S. Päßler, M. Wolff, and W.-J. Fischer, "Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food," *Physiological Measurement*, vol. 33, no. 6, p. 1073, 2012.

[130] V. Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–695, July 1997.

[131] G. Plasqui and K. Westerterp, "Physical activity assessment with accelerometers: an evaluation against doubly labeled water." *Obesity (Silver Spring)*, vol. 15, no. 10, pp. 2371–9, 2007.

[132] G. Plasqui, A. M. Joosen, A. D. Kester, A. H. Goris, and K. R. Westerterp, "Measuring free-living energy expenditure and physical activity with triaxial accelerometry," *Obesity Research*, vol. 13, no. 8, pp. 1363–1369, 2005.

[133] M.-Z. Poh, K. Kim, A. Goessling, N. Swenson, and R. Picard, "Cardiovascular monitoring using earphones and a mobile device," *IEEE Pervasive Computing*, vol. 11, no. 4, pp. 18–26, 2012.

[134] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation of food intake using a mobile device," in *2009 Workshop on Applications of Computer Vision (WACV)*, December 2009, pp. 1–8.

[135] J. Putnam and J. Allshouse, *Food Consumption, Prices, and Expenditures, 1970-97*, ser. Statistical bulletin. U.S. Department of Agriculture, ERS, 1999.

[136] L. Rabiner and B. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[137] L. Rabiner, S. Levinson, A. Rosenberg, and J. Wilpon, "Speaker-independent recognition of isolated words using clustering techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, no. 4, pp. 336–349, 1979.

[138] M. Rahmana, M. Pickering, D. Kerr, C. Boushey, and E. Delp, "A new texture feature for improved food recognition accuracy in a mobile phone based dietary assessment system," in *2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, July 2012, pp. 418–423.

[139] D. Reynolds and R. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.

[140] A. Salarian, H. Russmann, C. Wider, P. Burkhard, F. Vingerhoets, and K. Aminian, "Quantification of tremor and bradykinesia in parkinson's disease using a novel ambulatory monitoring system," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 313–322, February 2007.

[141] N. Saranummi, "In the spotlight: Health information systems," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 15–18, 2010.

[142] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, E. Melanson, M. Neuman, and J. Hill, "Toward objective monitoring of ingestive behavior in free-living population," *Obesity*, vol. 17, no. 10, pp. 1971–1975, 2009.

[143] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiological Measurement*, vol. 29, no. 5, p. 525, 2008.

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

[145] E. Sazonov and J. Fontana, "A sensor system for automatic detection of food intake through non-invasive monitoring of chewing," *Sensors Journal, IEEE*, vol. 12, no. 5, pp. 1340–1348, May 2012.

[146] E. Sazonov, G. Fulk, J. Hill, Y. Schutz, and R. Browning, "Monitoring of posture allocations and activities by a shoe-based wearable sensor," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, pp. 983–990, April 2011.

[147] E. Sazonov, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. Melanson, and M. Neuman, "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 626–633, March 2010.

[148] N. Sazonova, R. Browing, and E. S. Sasonov, "Prediction of bodyweight and energy expenditure using point pressure and foot acceleration measurements," *The Open Biomedical Journal*, vol. 5, pp. 110–115, December 2011.

[149] D. A. Schoeller, "Limitations in the assessment of dietary energy intake by self-report," *Metabolism*, vol. 44, Supplement 2, no. 0, pp. 18–22, 1995.

[150] D. A. Schoeller *et al.*, "Measurement of energy expenditure in free-living humans by using doubly labeled water," *The Journal of Nutrition*, vol. 118, no. 11, pp. 1278–1289, 1988.

[151] J. Scisco, "Sources of Variance in Bite Count," PhD dissertation, Pshycology Department, Clemson University, 2012.

[152] J. Scisco, E. Muth, Y. Dong, A. Hoover, P. O'Neil, and S. Fishel-Brown, "Usability and acceptability of the bite counter device," in *Annual Meeting of the Human Factors and Ergonomics Society*, 2011.

[153] J. Shang, M. Duong, E. Pepin, X. Zhang, K. Sandara-Rajan, A. Mamishev, and A. Kristal, "A mobile structured light system for food volume estimation," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, November 2011, pp. 100–101.

[154] M. Shuzo, S. Komori, T. Takashima, G. Lopez, S. Tatsut, S. Yanagimoto, S. Warisawa, and J.-J. D. I. Yamada, "Wearable eating habit sensing system using internal body sound," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 4, no. 1, pp. 158–166, 2010.

[155] A. M. Siega-Riz, B. M. Popkin, and T. Carson, "Trends in breakfast consumption for children in the united states from 1965-1991." *The American Journal of Clinical Nutrition*, vol. 67, no. 4, pp. 748S–756S, 1998.

[156] K. A. Siek, K. H. Connelly, Y. Rogers, P. Rohwer, D. Lambert, and J. L. Welch, "When do we eat? an evaluation of food items input into an electronic food monitoring application," in *Pervasive Health Conference and Workshops, 2006*, December 2006, pp. 1–10.

[157] M. Singh, A. Basu, and M. Mandal, "Human activity recognition based on silhouette directionality," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1280–1292, September 2008.

[158] A. J. Smeets and M. S. Westerterp-Plantenga, "Acute effects on metabolism and appetite profile of one meal difference in the lower range of meal frequency," *British Journal of Nutrition*, vol. 99, pp. 1316–1321, 5 2008.

[159] J. Speakman, *Doubly Labelled Water: Theory and Practice*. Springer, 1997.

[160] T. Spiegel, "Rate of intake, bites, and chews-the interpretation of lean-obese differences," *Neuroscience and Biobehavioral Reviews*, vol. 24, no. 2, pp. 229–237, 2000.

[161] G. B. Spurr, A. M. Prentice, P. R. Murgatroyd, G. R. Goldberg, J. C. Reina, and N. T. Christman, "Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry." vol. 48, no. 3, pp. 552–9, 1988.

[162] E. Stellar and E. E. Shrager, "Chews and swallows and the microstructure of eating," *The American Journal of Clinical Nutrition*.

[163] R. Stuart and B. Davis, *Slim Chance in a Fat World: Behavioral Control of Obesity.* Research Press, 1974.

[164] M. Sun, Q. Liu, K. Schmidt, J. Yang, N. Yao, J. D. Fernstrom, M. H. Fernstrom, J. P. DeLany, and R. J. Sclabassi, "Determination of food portion size by image processing," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 2008, pp. 871–874.

[165] M. Sung, C. Marci, and A. Pentland, "Wearable feedback systems for rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, no. 1, p. 17, 2005.

[166] A. Swartz, L. Squires, and S. Strath, "Energy expenditure of interruptions to sedentary behavior," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 8, no. 1, p. 69, 2011.

[167] F. Takeda, K. Kumada, and M. Takara, "Dish extraction method with neural network for food intake measuring system on medical use," in *2003 IEEE International Symposium on Computational Intelligence for Measurement Systems and Applications (CIMSA)*, July 2003, pp. 56–59.

[168] F. Thompson and A. Subar, *Dietary Assessment Methodology*, 2nd ed. Academic Press/Elsevier, 2008.

[169] F. E. Thompson, A. F. Subar, C. M. Loria, J. L. Reedy, and T. Baranowski, "Need for technological innovation in dietary assessment," *Journal of the American Dietetic Association*, vol. 110, no. 1, pp. 48–51, 2010.

[170] R. M. A. Triches and E. R. J. Giugliani, "Obesity, eating habits and nutriotional knowledge among school children," *Revista de Saude Publica*, vol. 39, pp. 541–547, 08 2005.

[171] C. C. Tsai, G. Lee, F. Raab, G. J. Norman, T. Sohn, W. G. Griswold, and K. Patrick, "Usability and feasibility of pmeb: A mobile phone application for monitoring real time caloric balance," in *Pervasive Health Conference and Workshops*, 2006, pp. 1–10.

[172] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 147–1488, November 2008.

[173] R. Tuschl, P. Platte, R. Laessle, W. Stichler, and K. Pirke, "Energy expenditure and everyday eating behavior in healthy young women," *The American Journal of Clinical Nutrition*, vol. 52, no. 1, pp. 81–6, 1990.

[174] D. Vasquez and J. Judy, "Optically-interrogated zero-power mems magnetometer," *Journal of Microelectromechanical Systems*, vol. 16, no. 2, pp. 336–343, April 2007.

[175] H. Vathsangam, A. Emken, E. Schroeder, D. Spruijt-Metz, and G. Sukhatme, "Determining energy expenditure from treadmill walking using hip-worn inertial sensors: An experimental study," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2804–2815, October 2011.

[176] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi, "An image procesing approach for calorie intake measurement," in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA),*, May 2012, pp. 1–5.

[177] W. Walker and D. Bhatia, "Towards automated ingestion detection: Swallow sounds," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, September 2011, pp. 7075–7078.

[178] J.-S. Wang and F.-C. Chuang, "An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 7, pp. 2998–3007, July 2012.

[179] G. Welk, *Physical Activity Assessments for Health-Related Research*, 1st ed. Human Kinetics, August 2002.

[180] K. R. Westerterp, "Physical activity assessment with accelerometers," *International journal of obesity and related metabolic disorders*, vol. 23 Suppl 3, April 1999.

[181] J. C. Witschi, "Short-term dietary recall and recording methods," *Nutritional epidemiology*, vol. 4, pp. 52–68, 1990.

[182] E. Wohlers, J. Sirard, C. Barden, and J. Moon, "Smart phones are useful for food intake and physical activity surveys," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, September 2009, pp. 5183–5186.

[183] W. Wolfe, C. CC, H. Frongillo, Jr, and JD, "Overweight school children in New York state: prevalence and characteristics," *American Journal of Publioc Health*, vol. 84, pp. 807–813, 1994.

[184] W. Wu and J. Yang, "Fast food recognition from videos of eating for calorie estimation," in *2009 IEEE International Conference on Multimedia and Expo (ICME)*, July 2009, pp. 1210–1213.

[185] R. Xu, S. Zhou, and W. Li, "Mems accelerometer based nonspecific-user hand gesture recognition," *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1166–1173, May 2012.

[186] N. Yao, R. Sclabassi, iang Liu, and M. Sun, "A video-based algorithm for food intake estimation in the study of obesity," in *2007 IEEE 33rd Annual Northeast Bioengineering Conference (NEBC)*, March 2007, pp. 298–299.

[187] J. Yin, Q. Yang, and J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, August 2008.

[188] L. Yu, F. Tay, D. Guo, L. Xu, M. Nyan, F. Chong, K. Yap, and B. Xu, "A mems-based bioelectrode for ecg measurement," in *2008 IEEE Sensors*, October 2008, pp. 1068–1071.

[189] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and emg sensors," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, November 2011.

[190] F. Zhu, M. Bosch, I. Woo, S. Kim, C. Boushey, D. Ebert, and E. Delp, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 4, pp. 756–766, August 2010.