Activity Recognition for Trans-Tibial Prosthesis Users

Morgan Redfield

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Electrical Engineering

University of Washington

2013

Committee:

Shwetak Patel, Chair

Jeffrey Bilmes, Chair

Brian Hafner

Joan Sanders

Program Authorized to Offer Degree:
Department of Electrical Engineering

University of Washington

Abstract

Activity Recognition for Trans-Tibial Prosthesis Users

Morgan Redfield

Co-Chairs of the Supervisory Committee:
Shwetak Patel
Department of Electrical Engineering

Jeffrey Bilmes
Department of Electrical Engineering

Prosthetists and medical researchers often find it difficult to assess the results of specific prosthesis prescriptions or therapies, which causes difficulties in medical decisions regarding an amputee. Currently, subject self-reporting is often used to determine how well a particular prosthesis is performing. Subjects may also be asked to perform certain actions in the clinic to assess the effectiveness of a given prosthesis. These methods are limited because self-reporting may be unreliable and experiments in a lab or doctor's office do not adequately approximate behavior and activities in a subject's daily life.

To determine how subjects use their prostheses outside of the clinic, it has become common to use accelerometers to enable long-term monitoring of a subject's activities. Accelerometers can offer information about how often the prosthesis is worn, how active the subject is over time, and what body positions subjects take. Currently available systems tend to only measure activity level or to require subjects to wear an accelerometer (or multiple accelerometers) at inconvenient positions on their body. The use of a single accelerometer mounted on a prosthesis may avoid the high compliance demands of similar methods while still allowing for accurate classification.

We propose and test three methods of classifying the activities and body postures of prosthesis users using only data from a single prosthesis mounted accelerometer. The classification methods tested were a rule-based system, a K-Nearest Neighbor system, and a

Hidden Markov Model. Personalized and non-personalized classifiers were tested for the K-Nearest Neighbor and Hidden Markov Model systems. All three classification methods were tested on data collected from eleven subjects with trans-tibial amputation. Data was collected while the subjects performed an hour long semi-structured activity protocol. The activity protocol included donning and doffing the prosthesis, sitting in various positions, standing, walking, and using the stairs. The predicted activity from each classification method was compared to the ground truth record and the overall accuracy for each method was calculated.

All three methods were able to classify activities with approximately 90% accuracy. Personalization improved the accuracy of both K-Nearest Neighbor and Hidden Markov Model classifiers. All classifiers were able to accurately differentiate periods of motion, but periods where the prosthesis was doffed, the subjects were sitting, or the subjects were standing were sometimes confused. These methods may be applicable for clinical use, and may improve the quality of care available to prosthesis users.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

GLOSSARY

BDT: Binary Decision Tree.

DFT: Discrete Fourier Transform.

G: Earth gravity at sea level. Equivalent to 9.81 m/s^2.

HMM: Hidden Markov Model.

IRB: Institutional Review Board.

KNN: K-Nearest Neighbor.

MFCL: Medicare Functional Classification Level. The MFCL is a measure of the level of activity that a person is capable of performing. It can take integer values from 1 to 4.

OFFLINE PROCESSING: Processing data after the full dataset has been gathered. This allows processing to make use of data from the "future".

ONLINE PROCESSING: Processing data in real-time.

RESIDUUM: That portion of the limb that remains after an amputation. Also called the residual limb.

SD: Standard Deviation.

NOTICE OF WORK ALREADY PUBLISHED

Portions of this work were submitted to the Journal of Rehabilitation Research and Development. A paper regarding heuristic and rule-based activity recognition will be published under the title "Classifying Prosthetic Use via Accelerometry in Persons with Trans-Tibial Amputations," by Morgan T. Redfield, John C. Cagle, Brian J. Hafner, and Joan E. Sanders.

ACKNOWLEDGMENTS

I'd like to thank the advisors who have made this work possible: Shwetak Patel, Joan Sanders, Brian Hafner, and Jeff Bilmes. Their support and advice have been instrumental in my development as an engineer and a researcher.

The Sanders' Prosthetics Lab staff and students have made my time as a graduate student fun and memorable. They continually shared their experience and knowledge. Thanks go to John Cagle for his organizational help and especially for his devotion to our engineering meetings, Christian Redd for his help with experiment design, David Swartzendruber for his help with electronics, Dave Gardner for all the bugs he found in my code, and Colin Dietrich for his wry wit. Special thanks go to Kate Allyn and Dan Harrison, who offered support and assistance in scheduling subjects and organizing experiments.

Eric Larson was incredibly helpful in this work. I gained a lot from his deep knowledge of signal processing and machine learning, patience with all of my questions, and friendliness. His impact on my education is perhaps greater than that of any other that I met in graduate school.

Thanks also to my family. Their constant encouragement and support were always available when I needed it most.

And finally, thanks to Susan Taylor for everything.

DEDICATION

For my family.

Chapter 1

INTRODUCTION

There are currently over one and a half million people in America who have lost a lower limb due to cancer, trauma, or dysvascular disease [68]. Dysvascular disease and related disorders cause the majority of lower limb amputations. Amputations may also be due to trauma from military conflicts and wartime injuries, motor-vehicle accidents, or natural disasters [19, 32]. Amputation rates are rising every year due to an increase in diabetes and improvements in treatment for traumatic injuries. The number of people living with limb loss is projected to double in the next 50 years [68].

The quality of life for people with lower limb amputations is heavily dependent on their level of mobility and the level of pain they feel in their residual limb. A well-fitting prosthesis is crucial to increasing mobility and reducing residual limb pain. It has been shown that those amputees who use a prosthesis have a higher quality of life than those that do not [59]. Certain components of the prosthesis, such as the liner, the type of foot chosen, and the pylon, can change the usability of the prosthesis [40, 63]. Choice of prosthesis components depends strongly on the lifestyle and goals of the prosthesis user [13]. It is therefore important to quantify how a person uses their prosthesis on a day to day basis.

The fit of a prosthesis has a direct impact on the amount of activity that a person with an amputation feels comfortable performing [40]. However, the fit of a prosthesis is not constant over time. A prosthesis that fits someone well in some circumstances may not fit them well when they are active and the volume of their residual limb has changed [67]. As the volume of a person's residual limb changes, unintended locations on their residuum may experience loading. This can cause discomfort, pain, and skin irritation or breakdown [32]. Problems with socket fit often lead prosthesis users to stop using their prosthesis for a period of time, which can severely affect the well-being of the person [41]. Prosthetics researchers may be interested in correlating the activities and bodily positions that prosthesis users perform with the resulting limb health. This requires an accurate method of monitoring prosthesis users as they go about their daily life.

The increasing shift in prosthetics to evidence-based practice has also driven a need to quantify prosthesis use. As third-party payers push for evidence that the interventions made are effective, the need has grown to accurately gauge what a person does with their prosthesis over time and how it changes [40]. Accurate knowledge of prosthetic use in free-living conditions would enhance prosthetic prescriptions, fitting processes, and measurement of outcomes [16].

Differentiation of body postures as well as gait types may be clinically important as sitting and standing can affect changes in residual limb volume and alter the fit of a prosthesis [55, 54]. Accurate knowledge of how much a prosthetic user sits or stands could thus be useful in determining changes in socket fit throughout the day. The characterization of prosthesis use could be partially achieved by quantifying prosthetic wear (e.g., donning and doffing) and users' engagement in locomotor activities (e.g., walking and stair climbing) and fundamental body postures (e.g., standing or sitting).

Despite its importance, prosthetists, physicians, and prosthetics researchers are challenged to describe how persons with limb loss use their prostheses outside the clinic or laboratory [12]. Performance tests such as the timed up and go test [48] or the six-minute walk [17] can be used to measure mobility of a prosthetic user in a clinic or laboratory [56, 36], but information on what prosthesis users do in their daily lives can be difficult to acquire. Characterizing ways that prostheses are used is complicated by the range of situations and environments users encounter.

Of the currently available activity monitoring solutions, many focus primarily on gait. Those that include posture often have low battery life or require high subject compliance. A system that requires prosthesis users to accurately attach a wearable sensor in the morning and detach it in the evening is more likely to experience compromised or missing data. A solution is needed that requires little to no active subject involvement. Accordingly, it is desirable to mount all sensors on the prosthesis.

1.1 Thesis Goals and Contributions

This thesis describes the development and evaluation of a set of signal processing algorithms to classify the data from an inexpensive, off-the-shelf accelerometer. Use of a single sensor mounted to a prosthesis would eliminate the need for the user to attach and remove the sensor and improve wear compliance. The hypothesis under investigation is that a system using only a prosthesis mounted sensor is capable of classifying periods of prosthesis use with high enough accuracy to be used in clinical applications. Periods of prosthesis use are categorized as prosthesis-off, movement, standing, or sitting. This adds postural identification based on data from a single accelerometer to the large body of research on gait identification from similarly located accelerometers. These activities are those that are of primary interest to prosthetists and prosthetics researchers.

The effects of online processing on accuracy are further investigated. Online processing, or generating results in real-time, is more limited than offline processing because it cannot make use of future data to identify what activity is currently being performed. While offline processing is acceptable for such applications as identifying the impact of different prosthesis components, online processing is necessary to inform the responses of a reactive prosthesis.

Chapter 2
BACKGROUND

Activity recognition has been under investigation for decades. As such, there exist many methods of performing activity recognition. Here, research related to activity detection in the community of prosthesis users is discussed. Relevant methods of activity classification are covered. Finally, the use of accelerometers in activity recognition is detailed.

2.1 Activity Recognition for Prosthesis Users

Previous methods for measuring prosthetic use outside of a gait laboratory or clinic include self-report surveys and personal activity monitoring devices (e.g., pedometers and step activity monitors). Self-report surveys have been used to quantify frequency and duration of prosthetic use [47]. However, self-report of activity among persons with limb loss has been noted to be unreliable when compared to a step-activity monitor [61]. Pedometers and step-counters have been used to objectively measure step-activity of persons wearing prostheses over extended periods of time [8, 23, 51, 61, 16, 11, 46]. While these sensors accurately measure gait activities, they are unable to provide information about body postures that may also be part of a person's habitual activity [21].

Prosthetists are increasingly turning to automated monitoring systems to record and classify the actions and activities performed by a person with an amputation. Much of this research is focused primarily on ambulation related parameters. Automated monitoring techniques have been used on persons with lower limb amputations to quantify step counts [22], estimate ambulation time [60], and describe gait patterns [58]. Algorithms have also been developed to identify locomotion and posture of individuals with an amputation from sensor data obtained over short time periods (i.e., up to several hours) [6, 9, 51, 44].

While prior studies have demonstrated the potential for activity and posture classification based on data from body-mounted sensors, there remain challenges to clinical use such as need for multiple sensors, subject donning requirements, low storage capacities, and short battery lives. Currently available sensors are also often restricted to short-term applications and/or require adherence to specific user protocols. Accordingly, more user-friendly and clinically-relevant solutions are needed to overcome these challenges.

2.2 Activity Recognition Methods

Automated activity monitoring is an important tool used in a variety of medical fields. It has been suggested as a technique that can be used to monitor recovery after surgery, to detect falls, to characterize types of gait, and to monitor activity levels for exercise. It has also been used with active medical devices, such as pacemakers, to inform the functioning of the device in real-time [28]. The majority of the methods investigated for activity monitoring make use of a machine learning algorithm that interprets data gathered from one or more sensors mounted on a persons body. Rule based systems have also proven popular [30].

Rule based methods are discussed because of their popularity and the ease with which they can be understood and implemented. K-Nearest Neighbor methods are discussed for their simplicity and general accuracy. Hidden Markov Models are discussed because their ability to account for structure in the series of activities performed may improve classification accuracy significantly.

2.2.1 Rule based methods

Rule based classification methods make use of ad hoc heuristics and tests to determine the activity being performed using sensor data. These rules may be learned by some algorithm or constructed by hand by researchers [30]. Classification is performed by testing a dataset against the rules, which are often organized in a tree structure [39]. In general, there is one or more rule for each type of activity that is to be classified. There may also be a default ”unknown” class to account for the case when all individual tests are inconclusive.

These methods are appealing to researchers because they are easy to understand and interpret. It may be clear that a duration of data is classified as walking, for example, because that duration had strong periodic structure. This may be one of the biggest reasons

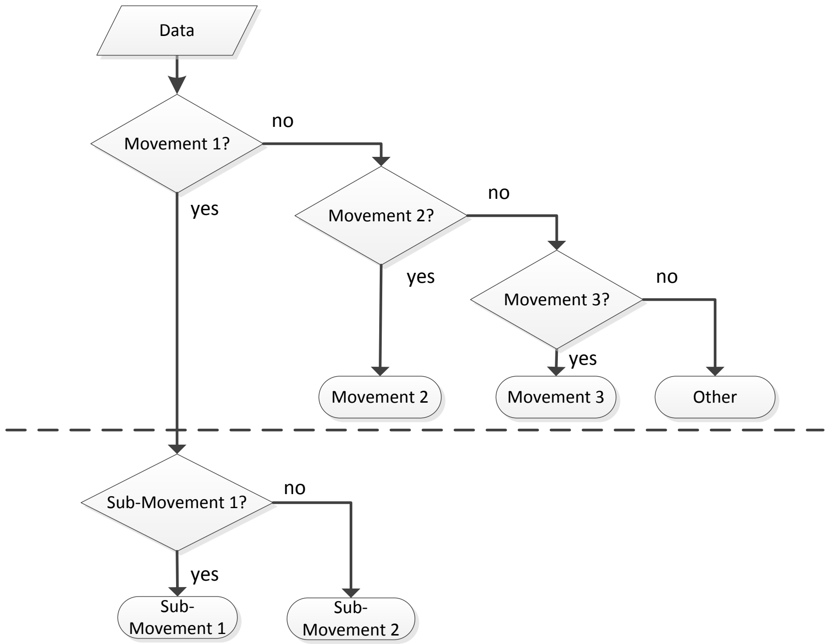

Figure 2.1: The general structure of a rule-based classification system. Each class is determined by a separate heuristic.

for the popularity of rule-based methods, since it can often be difficult to determine why a particular classification was made with machine learning methods.

One of the primary difficulties of rule based methods is determining the rules in the first place. This is often a task performed manually by researchers, which can be a labor intensive process. To do this, researchers must interpret the data themselves and determine what different heuristics may be used to classify data. Manually determined classification rules may be more appealing to researchers than automatically determined classification methods because the manually determined algorithms may generalize to different people more easily [30]. This may naturally avoid over-fitting the classification algorithms, which may occur when a classification algorithm performs well on training data but not on new data (a common problem with machine learning methods).

Previously developed rule-based classification systems [30, 39, 27] have analyzed accelerometer data to determine the activities and postures of able-bodied and amputee subjects. These systems used a sliding window to create sequential segments of data that could be classified as belonging to a single activity or posture of interest. The classification system in [39, 27] uses a hierarchical set of tests (Figure 2.1). The high level activity is identified first, and sub-activities are then determined based on the high level activity. This form of rule based classifier treats sequential activities as independent, although this is not required of rule based systems.

2.2.2 Machine Learning Methods

Supervised machine learning methods can be used to automatically determine the class of a piece of data based on sample data. In general, this is done by training an algorithm on manually classified data in such a way that the algorithm can accurately recognize new data as coming from one of the labeled sets [37]. To ensure that the trained classifier can accurately classify new data, it must be tested on data that it was not trained on. This can be done by separating the training data into two sections, training on one section and then testing on the other [50, 37]. In order to simplify this process, the ad hoc generation of features used in rule based methods is often formalized by generating feature vectors from the data and training the algorithm on those feature vectors. One or more transforms are often used on input data to reduce its dimensionality before doing the comparisons. This can dramatically reduce computation time and improve classification accuracy [50, 30, 37].

Here, two machine learning methods commonly applied to classification of data are considered.

K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is a method of estimating the class of some input vector by comparing it to pre-classified data vectors. The data in each vector can be any number of values that describe an event. In this algorithm, no computation is done on the training data. Instead, each feature vector in the training data is treated as a point in parameter-space. Ideally, the training data is such that it separates naturally into distinct clusters in that parameter-space. Classification of new data is performed by identifying

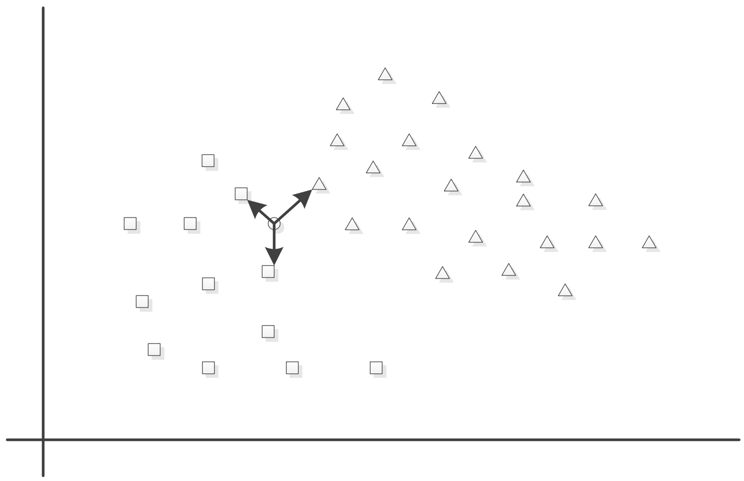

Figure 2.2: Identification of a single point using the KNN algorithm. In this case, the unknown point, represented by a circle, is compared to three nearest neighbors. Because the majority of the nearest neighbors belong to the square class, it is also classified as a square.

which cluster in parameters space best represents the new input vector [37].

The identification of the most representative cluster is accomplished by comparing a new data vector to each of the training vectors. The distances between the input vector and all pre-classified vectors are computed using the Euclidean distance formula

$d_i = \sqrt{\sum_{i=1}^{N}(S_i - X)^2}$

where d_i is the distance between points, S_i is the i-th pre-classified vector, and X is the input vector. The closest K of these pre-classified points are used to construct a class estimate. The class estimate can be determined by simple majority voting or weighted majority voting. In simple majority case, an input vector is determined to belong to the class that most of the near points belong to (Figure 2.2). In the weighted voting case, the importance of any of the neighbors is weighted by the distance, ensuring that nearer samples contribute more to the final determination of the class.

In using the KNN algorithm to classify activities from a time series, the time series must be broken up into windows (much like the rule based methods above). These windows are then treated as being independent, as the KNN algorithm cannot make use of correlations between subsequent windows. This may be valid if the features used for classification separate the classes enough in parameter space.

The KNN algorithm has previously been used to classify data from wearable sensors as being from a certain type of gait. One study made use of the KNN algorithm to differentiate between healthy and impaired gait [5]. That study used a head-mounted accelerometer to monitor subjects with simulated and real abdominal surgeries. It was found that a KNN approach applied to features generated after performing independent component analysis on the data was able to approach 100% classification accuracy. Another study has shown that it is possible to generate features which well-separate accelerometer data collected from the trunks of subjects as they walk on level ground, walk up stairs, and walk down stairs [57].

Hidden Markov Models

Hidden Markov Models, or HMMs, are dynamic graphical models in which each discrete time unit is modeled as an *observed* feature variable that depends upon a *hidden* state variable (Figure 2.3). The sequence of hidden variables forms a Markov Chain, while the observed variables are random variables that depend only on the state of the associated hidden variable [49, 37]. HMMs can be characterized by the transition matrix of the associated Markov Chain and the statistics determining the observed variables.

Both the observable and hidden variables may be known, in which case the parameters of the HMM can be learned via expectation maximization [42, 37]. Once the parameters of the model are known, a number of algorithms exist that can determine the most likely sequence of hidden states to explain a sequence of observed variables [37].

The observed feature vectors in an HMM can be treated as multi-dimensional Gaussian random variable. The state of the hidden variable determines the statistics of the observed distribution (i.e., the mean vector and covariance matrix of the observed distribution).

Due to the intrinsic Markov chain of an HMM, the correlational structure of a time series can be taken into account when performing classification. Subsequent windows of data are not seen as independent, as in the case of the KNN algorithm. For data in which one type

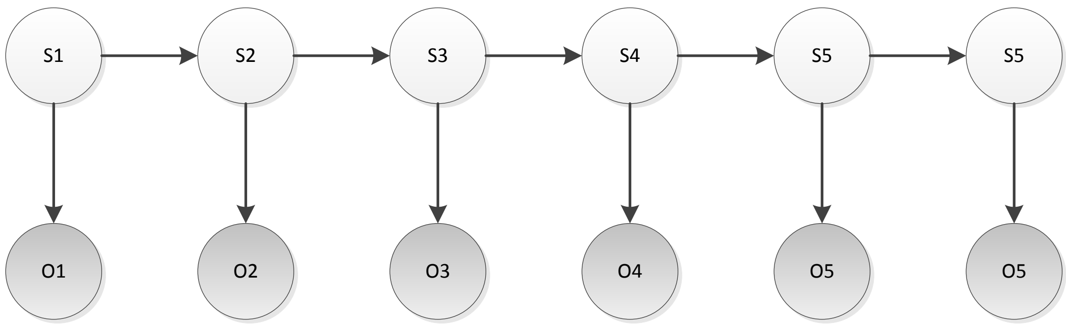

Figure 2.3: The graph describing a Hidden Markov Model. In this case, the S variables represent the state, or the activity being performed at a given time. The O variables represent the associated accelerometry observations.

of class follows another consistently (e.g., when sitting is consistently followed by a stand transition) this can improve accuracy significantly.

Classification of data using an HMM is often performed by first constructing observation feature vectors for windows of data. The Viterbi algorithm is then used to find the most likely sequence of states that could have generated those feature vectors [37].The Viterbi algorithm makes use of a two pass recursion where the entire dataset is taken into account. This allows for very accurate classifications, but it can only be done post hoc. Online classification, in which only current and past data are taken into account, can be performed using a similar algorithm [53].

In activity recognition applications, clinicians and researchers are often interested only in the activities of a subject over a given period of time. In this case, the offline Viterbi algorithm is satisfactory. If classifications are desired in real-time, for example if they are used as inputs to an active prosthesis, the online classification algorithm must be used. The decrease in classification accuracy of the online algorithm is application dependent, and must be determined experimentally.

People's daily activities can be modeled as an HMM in which the hidden variables are the activities performed by the subjects and the observed variables are feature vectors generated from sensor data. HMMs have been used to determine activity from data from multiple body worn sensors [34, 35]. That system used a number of static classifiers to produce a first order activity estimate, and the output of those static classifiers was used as feature vectors for an HMM to provide temporal smoothing and increased accuracy.

The process of using a static classifier (such as a KNN) to provide input to an HMM can improve classification accuracy. Static classifiers may incorrectly classify very short durations (e.g., one or two windows) of activity. Using an HMM on the output of the static classifiers acts like a low pass filter, removing the high frequency changes and very short duration classifications. This "smoothing" action can bring the durations of the classified activities closer to the true durations of activities performed, thus increasing classification accuracy.

2.3 Use of Accelerometers in Activity Recognition

Quantitative monitoring of human movement and posture has been investigated through the use of numerous different sensors. These include sophisticated kinetic measurement systems [21], force sensitive resistors (FSRs) [52], footswitches, and accelerometers [22]. Many of these methods, such as FSRs and kinetic measurements systems, have high power requirements. Others, such as footswitches, are only capable of measuring specific types of activities.

While a variety of sensors have been used for activity recognition in both able-bodied and trans-tibial populations, the most commonly used sensor may be the accelerometer. Accelerometers have been used to quantify activity levels [9, 62], measure daily postures and movement [7, 39], and to analyze gait [58, 51, 29].

Accelerometers are devices that are capable of measuring linear acceleration along an axis. Recent advances in Micro Electro-Mechanical systems (MEMS) have enabled the miniaturization of accelerometers [66]. Off-the-shelf products are now available that contain three orthogonal MEMS accelerometers in a package smaller than the size of a dime [66]. These accelerometers are also able to operate on very low power budgets, making them suitable for long term activity monitoring [22].

Because they measure both gravitational acceleration and acceleration due to motion, 3-axis accelerometers can be used to monitor stationary postures and dynamic movements. When the measured acceleration is constant and of magnitude equal to 1g, it can be assumed that the accelerometer is stationary with respect to the earth. The orientation of the accelerometer can then be determined from the direction of the acceleration. This can be used to determine, e.g., the posture that a person has assumed [22].

When an accelerometer undergoes motion, the resulting acceleration signals can be used to infer information about the type of motion. At the most basic level, simple thresholds on statistics such as mean, standard deviation, or skewness have been used to determine when instrumented humans are moving and what type of motion they are engaging in. Frequency domain methods, such as the Discrete Fourier Transform or the Wavelet transform, have also been used to determine quality or type of motion [22, 4, 18].

Several design criteria must be addressed when using accelerometers in activity recognition applications [22, 29]. These are

1. the location of the accelerometer
2. the orientation of the accelerometer
3. the dynamic range of the accelerometer
4. the attachment of the accelerometer to the body segment

The position and orientation of the accelerometer must be chosen such that the postures of interest can be differentiated using stationary accelerometer signals. The accelerometer must also experience acceleration due to the motions of interest. The dynamic range of the accelerometer must also exceed the accelerations to which it may subjected. The range of acceleration a sensor on a limb may see during walking is approximately ± 5G, while the range of acceleration induced by running is approximately ± 10G [29]. Finally, care must be taken in attaching the accelerometer. Ideally, it will be firmly affixed to a limb segment or body part to minimize noise and spurious accelerations. Connection of the sensor in a location that is near a bone and not over a soft tissue can also minimize spurious accelerations.

Several commercial products are currently available that perform activity monitoring using body-mounted accelerometers. Some of these are meant primarily for consumer exercise monitoring (e.g., Nike+ Fuel Band [43], FitBit [20], Jawbone UP [26], and applications available for modern smartphones [3]). Many of these consumer monitors track only activity, and do not track posture. They are also often unable to provide the raw data for analysis, and instead provide only summary statistics. Other activity monitors are specifically meant for research and clinical applications. These include the ActiLife ActiGraph GT3X+ and the OrthoCare StepWatch. Both the GT3X+ and the StepWatch have been used in a number of clinical studies [23, 51, 61, 22, 11, 46].

These accelerometer based monitoring systems are primarily designed with the goal of monitoring calorie expenditure and duration of activities. While some of the systems are capable of monitoring body postures, they must be properly mounted to the trunk of the body to do so. This limits their use to individuals who are likely to correctly attach the accelerometer. If the accelerometer is not properly attached, the data obtained from the sensor may produce incorrect classifications. Because both posture and activity have an impact of the volume of the residuum in people with lower-limb amputations, a new monitoring method is needed.

Chapter 3

APPROACH

Several systems were designed and tested to assess their ability to accurately classify data from a single prosthesis-mounted accelerometer as doffed, sitting, standing, or moving. An in-lab study was performed on a number of people with trans-tibial amputations. Subjects who participated in the study were asked to engage in a certain set of activities while the accelerometer was attached to their prosthesis. The data gathered during this protocol were used to train and validate activity classification systems.

Three different classification algorithms were investigated to assess the differences among them. A rule-based system was chosen because it is most intuitive to prosthetists and clinicians. As such, it may be more likely to be accepted for use than methods that are less intuitive. Two machine learning methods were investigated. The KNN algorithm was chosen for its simplicity and ease of implementation. Should it prove most accurate, it may be improved upon by using support vector machines or similar methods. A Hidden Markov Model was also assessed to investigate the impact of correlation between sequential activities. KNN and related methods assume all activities are independent of activities at other times, while an HMM model can take into account that some activities often occur in a specific order.

The overall classification accuracy of each method was calculated, and sensitivity analyses were performed on selectable model parameters. The rule-based and KNN systems treat activities as being independent, and thus operate equivalently for both offline and online processing. The HMM system was tested in both offline and online processing modes to determine the usability of a HMM for different applications.

3.1 Sensors

The proposed algorithms will all operate on data from any accelerometer that meets the sensitivity and range requirements. ActiLife ActiGraph GT3X+ accelerometers were chosen for an in-lab study due to their broad acceptance by the medical field. The GT3X+ device is commonly used in clinical activity monitoring tasks, and would therefore make adoption of this system by prosthetists straightforward. The GT3X+ is also capable of operating for long periods of time due to the high battery and storage capacities. This allows a direct implementation of the activity classification method for long term monitoring should the classification methods prove accurate on the in-lab study data. Furthermore, the GT3X+ has a small form-factor and is lightweight. This will ensure that the device does not impact the gait of the prosthesis user [15, 25]. It also means that the device is easy to cover with pants if the prosthesis user does not want to display the fact that they wear a prosthesis or other medical devices.

The GT3X+ (Figure 3.1 has a ±6G (gravitational acceleration) dynamic range, 0.00293G resolution, 100Hz maximum sampling rate, up to 31 days of battery life, and up to 40 days of data storage. A dynamic range of ±6G has previously been found to be acceptable for quantifying movement patterns during walking, but not running [33, 29]. The ActiGraph accelerometer is packaged in a 4.6x3.3x1.5cm water-resistant enclosure and weighs 19 grams. A 40Hz sample rate was used in the experiments because the frequency content of gait is primarily below 20Hz [2, 1]. At this sampling rate, the GT3X+ accelerometer is capable of recording data for up to 30 days.

The GT3X+ accelerometer was applied to each subject's own prosthesis to measure limb segment accelerations during the in-lab study.

3.2 Subject selection

Persons with trans-tibial amputations were recruited to test the developed classification algorithm in a semi-controlled activity protocol. All subjects were recruited from local prosthetic clinics, peer support groups, and hospitals. Subjects were considered if they were between the ages of 18 and 65 and had either unilateral or bilateral trans-tibial amputa-

Figure 3.1: The GT3X+ accelerometer.

tion. To ensure that the subjects would be able to complete the activity protocol, included subjects were required to have an ambulatory ability of at least Medicare Functional Classification Level (MFCL) 2 (limited community ambulator) [10]. Before taking part in the study, a research prosthetist evaluated each subject to ensure that they had the ability to complete approximately an hour of walking with frequent rests. Subjects were also required to be at least two years post-amputation and to have a healthy residual limb with intact skin.

Volunteers were excluded if they had a body weight over 120 kg. They were also excluded if their residual limb showed evidence of recent or current skin breakdown.

3.3 Experimental Protocol

Subjects provided informed consent prior to participation. All procedures were approved by a University of Washington Institutional Review Board (IRB) before study procedures were initiated.

One accelerometer was attached to the subject's prosthesis, proximal to the foot. Positioning the accelerometer at this location, instead of one more proximal, ensured the sensor was subjected to high accelerations (i.e., received a strong signal) during leg motions. The sensor was oriented with the positive x-axis along the limb axis and the positive z-axis in the medial-lateral direction (Figure 3.2). The accelerometer's sample rate was set to 40Hz. A 40Hz sampling rate was deemed acceptable as most of the energy in gait is concentrated be-

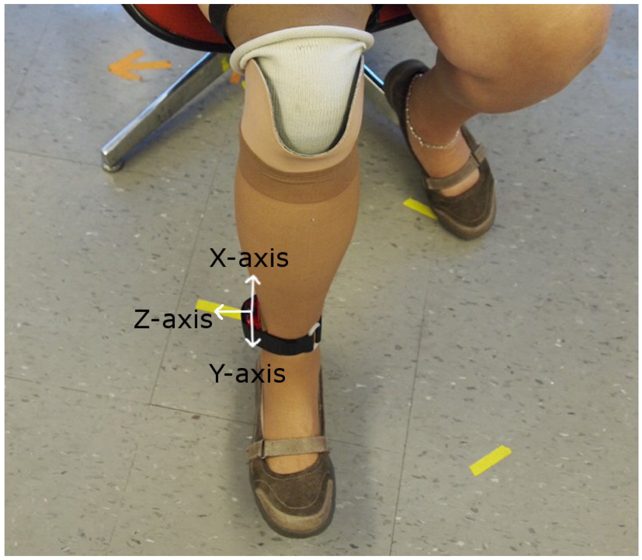

Figure 3.2: Accelerometer attachment and orientation. The accelerometer was connected securely to the subjects pylon with the z-axis facing in the medial-lateral direction and the x-axis facing in the long direction. The accelerometer was attached using sticky tape, and further secured using elastic.

low the resulting Nyquist frequency of 20Hz [2, 1]. Using a sample rate of 40Hz, rather than a higher rate typically used in gait laboratories, maximized the duration of data collection and still allowed for the identification of relevant gait events.

An experiment was performed to assess the accuracy of algorithms designed to classify use of prostheses as 1) movement (e.g., walking, using stairs, or transitioning from one posture to another), 2) standing (i.e., upright standing posture with minimal movement), 3) sitting (i.e., seated posture with minimal movement), or 4) doffed (i.e., prosthesis not being worn) based on data from the pylon-mounted accelerometer. These activities were chosen because they are likely to be of interest to prosthetists and prosthetics researchers.

Knowledge of when a prosthesis is doffed will allow prosthetists and researchers to calculate total wear time, which is important in determining the subject's overall quality of life. Recent research [55, 54] has revealed that sitting and standing may have different effects on the residual limb, and it is therefore of interest to differentiate these activities. Movement is of interest because it can be used to determine how active a person is, which is important in determining the quality of their prosthetic device. The decision was made not to focus on differentiating walking from running or using the stairs, and to treat all of these as movement. This was done because many other studies have focused on that classification problem, and those methods could be used here [22, 58, 51].

Subjects were asked to perform a pre-defined activity protocol that included walking over level ground (i.e., an indoor hallway); sitting on office chairs, sofas, or benches; standing; ascending and descending stairs; and doffing and donning the prosthesis. These activities were deemed to be most representative of the activities of clinical interest to prosthetists.

Subjects were asked to sit in each type of seat at least three times, stand and walk the hallway at least five times, use the stairs at least once, and doff/don their prosthesis once. Each time the subject performed an activity in the test sequence, they were asked to perform it for at least 60 seconds. Subjects were asked to perform the sequence in a set order, but were asked to engage in each activity, posture, or don/doffing action as they normally would (i.e., no instructions were given for walking speed, sitting posture, etc.). Subjects were visually monitored while they performed the test sequence. A researcher followed and timed each subject with a stopwatch to capture when subjects started and stopped each type of movement, posture, or don/doffing action. This record was used for ground-truth comparisons.

Chapter 4

IMPLEMENTATION

Here we discuss the detailed characteristics of acceleration signals obtained from a pylon-mounted accelerometer. The differences in acceleration data produced by different activities are emphasized. Next, the design and implementation of the three different classifiers is discussed. The rule based system is described, the features used for the machine learning system are described, and the machine learning systems themselves are discussed.

4.1 Characteristics of acceleration data

Acceleration signals from a pylon-mounted sensor show several differences based on the activity being performed by the prosthesis user.

When the prosthesis is off, the orientation of the prosthesis is arbitrary but constant. A person may leave their prosthesis standing upright, on its side, or leaning against some object. Therefore orientation cannot be used to determine the doffed status of a prosthesis. Because the prosthesis is stationary, the signal magnitude area (SMA) measured while a prosthesis is doffed will be zero (neglecting noise).

When a person is wearing their prosthesis and sitting or standing still, the orientation of the prosthesis gives strong information about what activity is being performed. In both positions, the prosthesis will most likely be upright. This being the case, the acceleration in the axial direction will be within a small range. The anterior-posterior accelerations also often differ between sitting and standing. In general, a person sitting will place their foot forward, whereas a person standing will place their foot directly underneath their center of gravity (Figure 4.1). This allows the differentiation of sitting and standing in many circumstances.

While people place their feet forward while sitting and upright while standing in general, it is by no means guaranteed. People may place most of their weight on their unaffected

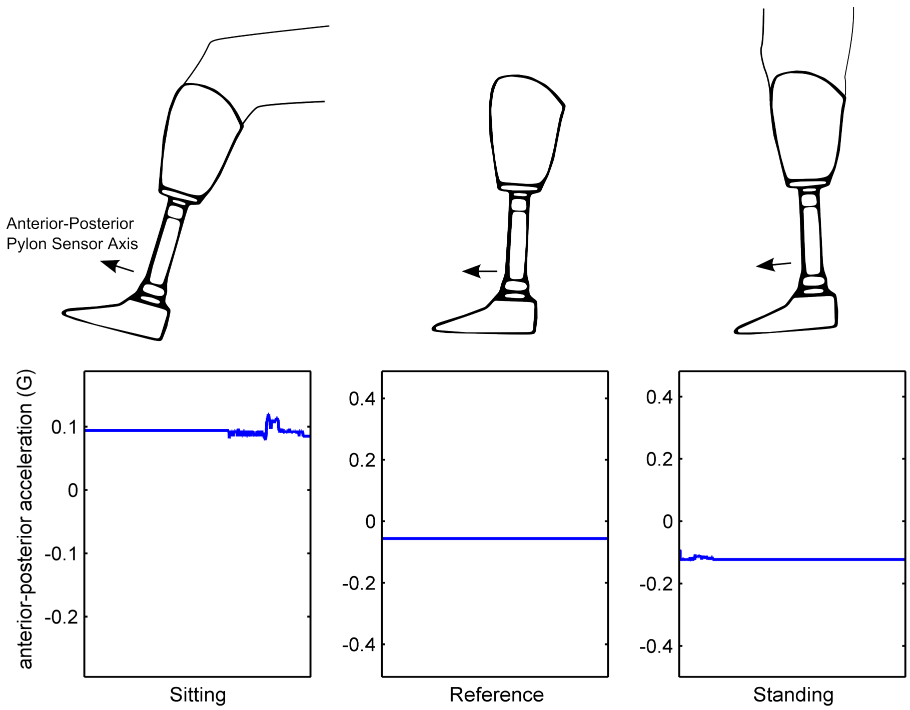

Figure 4.1: Pylon and thigh acceleration signals over a sixty second period when the subject was sitting, had doffed their prosthesis and placed the foot flat on floor (the reference position), or was standing.

limb while standing and hold their prosthesis in an unexpected orientation. If a person is sitting, they may bend their leg back behind their knee, especially if they are sitting in a higher than average chair. This variability complicates the identification of sitting and standing, as the orientation alone may not be enough to differentiate sitting from standing consistently.

Another difficulty that may negatively impact the ability of a classifier to recognize sitting and standing is that people do not remain still in these positions. People will shift back and forth, move their feet around, and generally adjust their posture over time to remain comfortable. This may introduce short periods of higher than normal SMA into a period that should be classified as all walking (or standing).

When a person is walking, the acceleration signals show an approximately periodic signal (Figure 4.2). Axial and medial-lateral accelerations both show large, slow changes

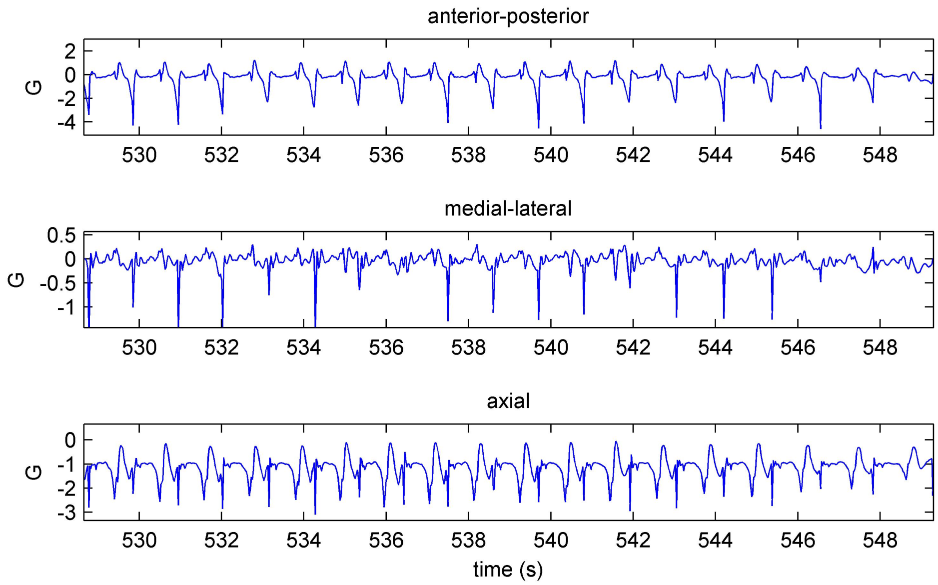

Figure 4.2: Acceleration signals during walking. Walking produces periodic acceleration signals in all three axes.

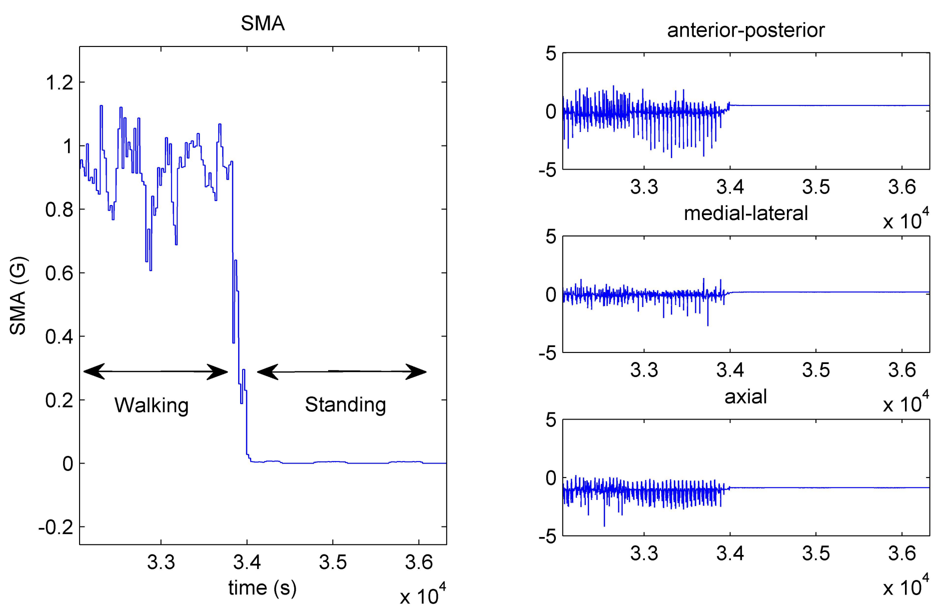

Figure 4.3: Signal Magnitude Area and associated accelerometer signals.

in acceleration while a foot is swinging, and sharp spikes of acceleration when a foot leaves the floor or lands. The axial acceleration shows changes of approximately 2 G during walking, while the anterior-posterior acceleration shows changes of 3 to 4 G. Medial-lateral acceleration is much less affected by walking, and shows spikes of approximately 1 to 2 G only when the foot hits the ground after a step.

The SMA of accelerometer data ranges from 0 during doffed up to between 0.6 and 2 during walking (Figure 4.3). During a transition from sitting to standing or vice-versa, the SMA ranges from 0.1 to 0.8. During standing or sitting, the SMA is often quite low. When subjects shift their weight or move their legs, the SMA may increase to approximately that of a transition in activities.

4.2 Data Analysis

To assess the suitability of the three different classification methods for activity recognition, we analyzed their performance on manually labeled datasets in which the ground truth of activity was known. The activities predicted by each classifier were compared to the true activity to determine the expected accuracy of the classifiers. Accuracy values were computed by calculating the percent of correctly classified windows of data. The true class of a window of data was set to be the class most prevalent in that window. All data analysis was performed using MathWorks Matlab 7.12.0 software and the Graphical Models Toolkit.

4.2.1 Rule based analysis

Raw acceleration data, such as those shown in Figure 4.2, obtained from the GT3X+ accelerometer were post-processed using a custom algorithm. No filtering was performed on the data. Data were buffered into short windows with an overlap of 50%. Window length was experimentally determined as described below and set to 50 samples (i.e., 1.125s).

A binary decision tree (BDT) algorithm [39] was designed to classify the windowed data. Data from all three axes of the accelerometer were used to determine if the subject was active, stationary (i.e., sitting or standing), or had doffed the prosthesis. Determination of posture was performed using only the anterior-posterior data of the accelerometer.

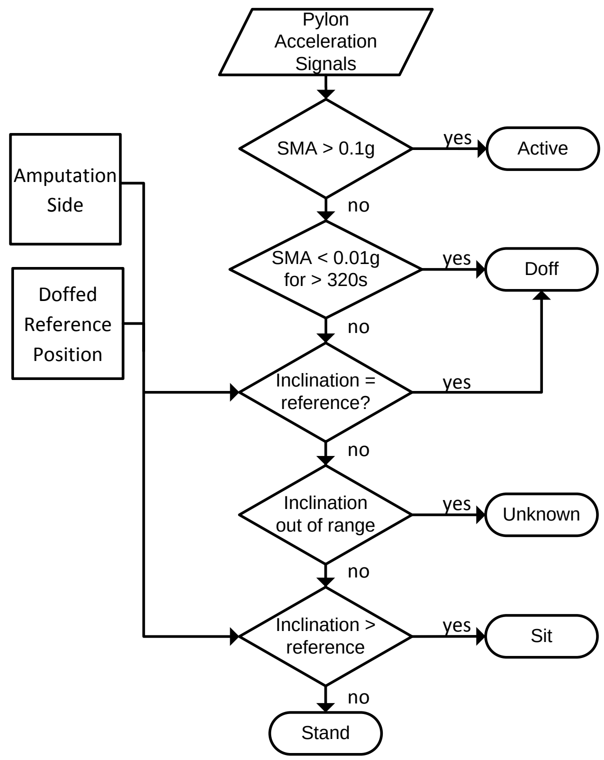

Figure 4.4: The binary decision tree algorithm used for activity and posture classification.

The BDT (Figure 4.4) used SMA to determine if the prosthesis was moving or stationary within each window. SMA was calculated by subtracting the mean of each accelerometer axis from each data point for that axis, integrating the absolute value of the result over a full window, and dividing by the window size. This method has previously been used to detect activity levels [38, 22, 14, 18]. SMA was evaluated using the following equation:

$$SMA = \frac{1}{n}\sum_{i=0}^{n} |X_i - \mu_X| + |Y_i - \mu_Y| + |Z_i - \mu_Z|$$

The developed algorithm required several subject-specific parameters for calibration. First, the accelerometer's location (i.e., left or right leg) was required to correctly orient the accelerometer's anterior-posterior axis; the axial direction did not change and the y-axis of the accelerometer rotated with leg change to remain pointing in the lateral direction. Second, the accelerometer's inclination while the prosthesis was doffed and standing upright with the foot on the floor was required. This doffed position served as a reference to differentiate sitting and standing postures. This strategy was effective because the anterior-posterior inclination angle (with respect to the vertical axis) was found to be greater than the doffed reference angle for sitting and less than it for standing (Figure 4.1).

Two activity thresholds were used to guide classifications. The lower and upper activity thresholds were experimentally determined via a sensitivity analysis using the laboratory-based experiment data as described below. These thresholds were set to 0.01g and 0.1g, respectively. When SMA was below the lower threshold, the subject was deemed either to be stationary or to have doffed their prosthesis. When SMA remained below the lower threshold for more than 320s, the prosthesis was considered doffed. The parameters to determine if the prosthesis was doffed were chosen based on the observation that subjects in stationary postures during the first experiment were not completely immobile for the full length of time they were in the posture, but further research will be needed to validate those parameter choices. Otherwise, the prosthesis was assumed to be donned and windows were classified as a stationary posture (i.e., standing or sitting). When SMA was between the lower and upper thresholds, the accelerometer data from that window were averaged to find the inclination [31]. Inclination was then compared to the subject's reference inclination to determine if the subject was sitting, standing, or doffed (Figure 4.4). If the prosthesis was oriented in a way that did not correspond to one of those postures, indicated by the inclination being outside of a range that could be obtained by a sitting or standing individual, the window was classified as unknown. Lastly, when SMA exceeded the upper threshold, the subject was considered to be engaged in movement.

A sensitivity analysis was performed on the three experimentally-determined parameters of the classification algorithm. Window size, upper activity threshold, and lower activity threshold were varied to determine the optimal values described above and to assess the sensitivity of the results to changes in those parameters.

4.2.2 Machine learning analysis

Data from the experiments were used to train the machine learning classifiers. Both machine learning algorithms under investigation in this study were tested using leave one out crossvalidation [37]. Two different training and testing regimens were used to determine the generalizability of these algorithms.

First, the algorithms were trained using data from all but one of the subjects. That subject's data was then used to test the resulting classifiers. The classifiers were used to produce predicted activities, which were compared to the true activities of that dataset. This was repeated such that data from all subjects was used for testing. The results of this testing approach revealed how generalizable the algorithm is to subjects that have not been seen before.

The second training and testing regimen was intended to determine the effect of adding personalized training for each subject. Each subject's experimental dataset was split into two halves, and the classification algorithms were trained on all but one of the resulting datasets. The resulting classifier was thus trained on data from other subjects and also data from the subject whose data was being classified. The final dataset was used to validate the accuracy. This was repeated until each of the datasets had been used to validate the models. Each subject thus had two sets of results: one for the first half of their dataset and one for the second. These accuracy values were averaged to produce the overall accuracy in classifying data from that subject using a personalized approach.

Selected Features

Prior to classification, the accelerometer data was processed to produce features that were representative of the activities being classified. If the accelerometer had been placed on the subject's left leg, the anterior-posterior data (y-axis of the accelerometer) was multiplied by -1 before generating features to account for the rotation of the accelerometer.

The acceleration data was then split into windows with 50% overlap. Choice of window length was determined through a sensitivity analysis for the KNN method, and the window size was set to the same as that for the rule-based method for the HMM method. A feature vector containing 6 features was then generated for each window of data. The features were chosen for their expected ability to differentiate the activities of interest.

The six features chosen were:

1. anterior-posterior axis orientation
2. medial-lateral axis orientation
3. axial axis orientation
4. SMA
5. peak frequency of the data
6. stationary duration

Orientation values were calculated for all three axes. These values made use of a pre-recorded orientation of the prosthesis in an upright and doffed position. This was done to account for any difference in angle for different prosthesis pylons. This also simplified the attachment of the accelerometer to the pylon, since small inaccuracies in attachment would be removed by shifting the acceleration values. To calculate the orientation value for a given axis, a subject-specific measurement of doffed orientation for that axis was subtracted from the mean value of the accelerometer data. The result was then normalized such that the maximum acceleration value, 6 G, would correspond to a feature value of 1.

Signal Magnitude Area, or SMA, was calculated for each window to provide a measure for the amount of movement the accelerometer was experiencing.

The frequency most representative of the data was also included in the feature vector. This frequency was calculated by taking the 128 point Discrete Fourier Transform of the data. The frequency that contained the largest magnitude was included in the feature vector after being normalized to the Nyquist frequency of 20Hz.

To provide information about whether the prosthesis was doffed, a feature was calculated that is similar to the SMA threshold duration of the rule based method above. If the SMA is below a set threshold for more than a certain number of windows, this feature was set to 1, otherwise it was set to 0. This feature cannot be calculated from only the data available in the current window, it requires information about past (but not future) windows. The SMA threshold for this feature was chosen to match the optimal lower threshold found for the rule based method. The number of windows the SMA must be below the SMA threshold for this feature to trigger was chosen to be 5, although more research is needed to validate this number.

This feature cannot be calculated from only the data available in the current window, it requires information about past (but not future) windows. The SMA threshold for this feature was chosen to match the optimal lower threshold found for the rule based method. The choice of how much to shift the arctangent function was made to ensure that short periods of time without movement had a large difference from long periods.

K-Nearest Neighbor

The K-Nearest Neighbor algorithm was used to classify feature vectors generated as described above. Simple majority voting was used to determine the predicted classification from the class of the nearest neighbors. A sensitivity analysis was performed to determine which value of K was optimal. A subsequent analysis was performed to determine what window length was optimal. The optimal values for K and for window length were then used to determine the classification accuracy of the KNN algorithm on a set of manually classified data.

Hidden Markov Model

A Hidden Markov Model was used to investigate the effect on classification accuracy of taking into account the temporal correlations in the activities performed by subjects. There is some preliminary evidence that the geometric distribution that HMMs use to represent activity durations may match the durations of stationary activities, though it may be a poor fit for the doffed and motion activities [45].

The Graphical Model Toolkit (GMTk) was used to train an HMM with the experimental data. The experimental data was buffered into overlapping windows, and feature vectors were generated for the windows as described above. The window length was chosen to be the same as that used for the KNN validation.

Observations in the HMM were represented as multi-dimensional Gaussian mixtures.

The means, variances, and number of components for each mixture were determined through expectation maximization. The transition matrix for the associated Markov chain was also determined through expectation maximization.

Both offline and online validation were performed for each dataset. Offline state prediction, using the Viterbi algorithm, determines the most likely state of each hidden variable based on all observations. Online state prediction, using a variant of the alpha recursion, determines the most likely state of each hidden variable based only prior and current observations.

These investigations into the accuracy of an HMM for the purposes of activity recognition should be seen as preliminary steps. There exist many types of dynamic graphical models (DGMs) other than HMMs that may also perform well. Furthermore, the use of HMMs (and other DGMs) to smooth the classification output from other classifiers may also be beneficial. Many avenues remain open to investigate the use of DGMs for activity recognition.

Chapter 5
RESULTS

The experimental protocol was run on eleven subjects with a trans-tibial amputation. The BDT, KNN, and HMM classification algorithms were run on the resulting datasets. All of the algorithms were capable of detecting active movement with greater than 95% accuracy. All methods classified sitting, standing, and doffed with lower accuracy. These stationary classes were confused with each other. The different methods classified data from subjects with differing levels of accuracy, and some subjects produced data that was accurately classified by one method but not by another.

5.1 Participant Demographics

Eleven subjects (nine male and two female) were recruited to participate in data collection protocol (Table 5.1). All subjects had a unilateral trans-tibial amputation. Subjects' mean age was 53 (SD=12), mean weight was 90.4 kg (SD=11.7 kg), mean height was 178.0 cm (SD=7.8 cm), and mean time since amputation was 19.6 years (SD=18.4 years). Subjects were classified as MFCL-2 (n=4), MFCL-3 (n=5), and MFCL-4 (n=2) by the study prosthetist based on interview and clinical evaluation.

5.2 Classification Accuracy

5.2.1 Rule based method

Window lengths from 20 to 200 samples were tested to determine the effect of window length on classification accuracy. The algorithm had a maximum classification accuracy at approximately 50 samples per window (Figure 5.1).

Accuracies for classifications derived using lower activity thresholds between 0.001 and 0.02 G and upper activity thresholds between 0.01 and 0.2 G were computed. It was found that if the activity thresholds were low, accuracy decreased substantially because stationary

Table 5.1: Subject demographics.

Subject	Limb	Sex	Age	Etiology	Years Since Amputation	Weight (kg)	height (cm)	MFCL
1	r	m	69	trauma	47	99.5	182	3
2	r	m	58	trauma	5	80.9	177	4
3	l	f	56	trauma	9	100.0	167	2
4	r	m	31	infection	2	88.7	170	3
5	l	m	49	tumor	12	108.5	177	2
6	l	m	49	trauma	22	100.0	182	4
7	l	m	36	trauma	2	98.2	190	3
8	r	f	65	trauma	58	78.4	167	2
9	l	m	53	trauma	6	109.5	183	2
10	l	m	51	trauma	5	85.3	165	3
11	l	m	27	trauma	7	110.0	178	3
Mean			49.4		15.9	96.3	177	
SD			11.1		18.9	10.4	8	

Table 5.2: Classification accuracy for all subjects using the rule-based and KNN methods.

Subject	Rule Based	KNN	Personalized KNN
1	99.6	95.8	96.0
2	96.8	94.0	94.4
3	90.0	83.0	83.9
4	99.1	90.8	91.0
5	96.2	84.5	89.3
6	97.5	94.9	95.1
7	99.3	94.4	95.2
8	94.0	91.7	92.3
9	81.2	90.6	90.8
10	84.6	92.1	92.4
11	78.7	85.6	96.1
Mean	92.5	90.7	91.5
SD	7.3	4.2	3.7

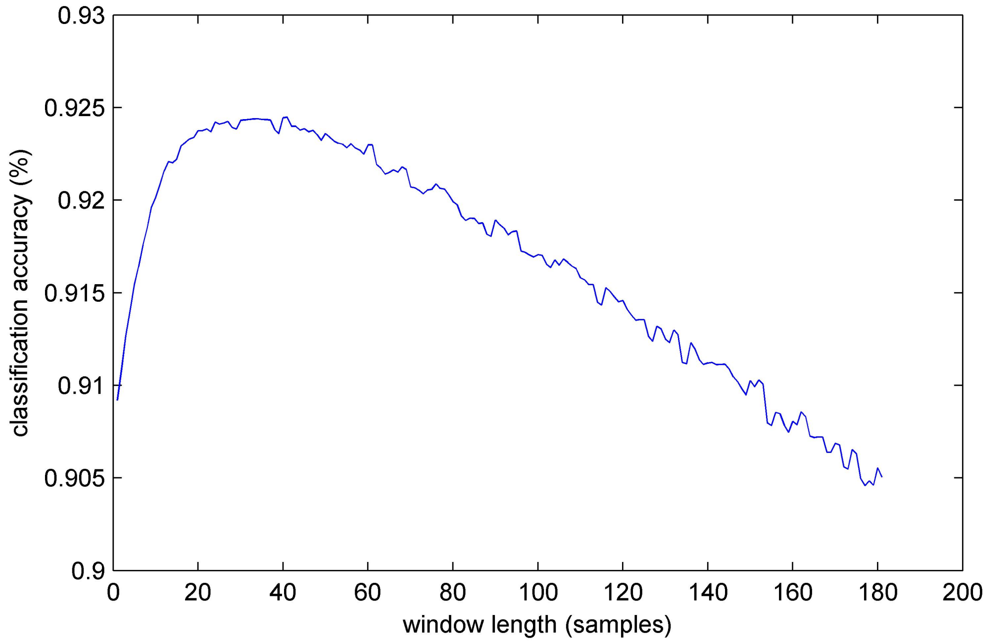

Figure 5.1: Sensitivity of the rule-based classification algorithm to the choice of window length.

postures (e.g., standing or sitting) were classified as active use (Figure 5.2). As thresholds increased from their optimal value, accuracy dropped off due to periods of activity being mis-classified as sitting, standing, or doffed. Maximum accuracy was achieved for the pylon data classification algorithm using a 0.01 G lower threshold and a 0.11 G upper threshold. Maximum accuracy was achieved for the pylon and thigh data classification algorithm using a 0.01 G lower threshold and a 0.11 G upper threshold. These optimal thresholds were used in all subsequent analyses.

Overall classification accuracy for each subject ranged from 78.7% to 99.6% (Table 5.2). Mean classification accuracy was 92.5% (SD=7.3%). The most commonly mis-classified body posture was sitting, which was typically mis-classified as standing. The confusion matrix (Figure 5.5) shows the percentage of actions that were classified correctly or mis-classified as other activities. The rule-based classification algorithm classified activities for subjects 1 through 8 with accuracy of 90.0% or above. Subjects 9, 10, and 11 produced

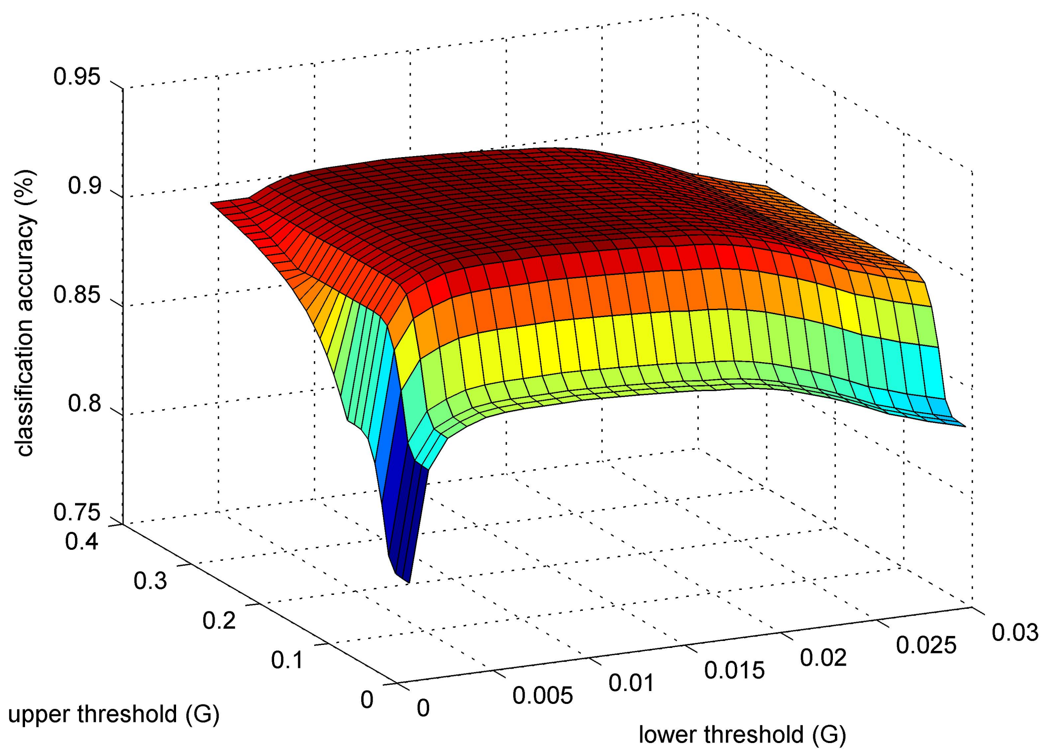

Figure 5.2: Sensitivity of the rule-based classification algorithm to the choice of SMA thresholds.

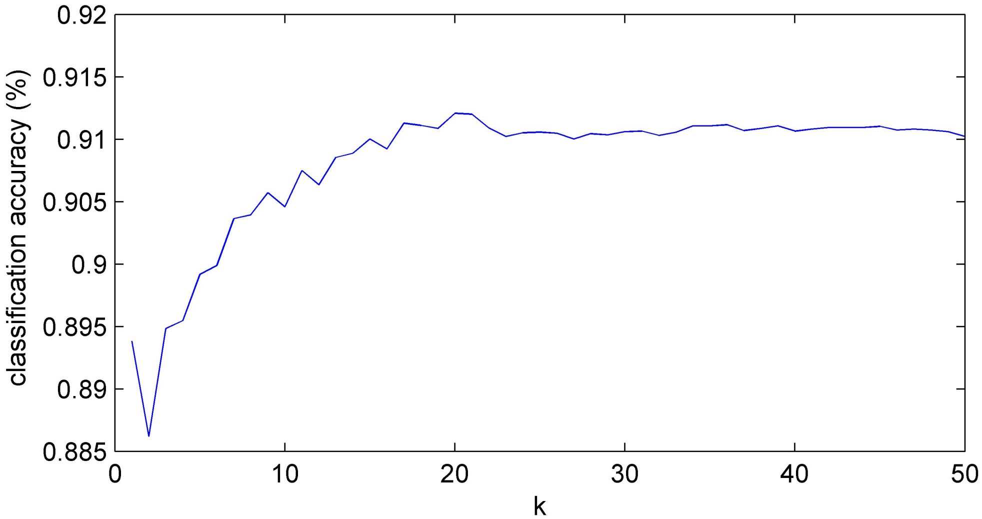

Figure 5.3: Sensitivity of the KNN algorithm to the choice of K, the number of nearest neighbors used for comparison.

data that was classified with lower accuracy, ranging from 78.7% to 84.6%.

5.2.2 Machine learning results

The machine learning methods had slightly lower classification accuracies than the rule based method, but they all had lower variance. Personalization improved classification accuracy slightly for the KNN algorithm and substantially for the HMM algorithm. The offline HMM method classified activities with slightly higher accuracy than the online HMM method.

K-Nearest Neighbor

Values of K from 1 to 50 were tested to determine the effect that the number of nearest neighbors considered had on classification accuracy. It was found that personalized accuracy increased with K from 1 to 20 and decreased for higher (Figure 5.3). The value K=20 was used for all subsequent analysis.

Window lengths from 30 to 150 samples were tested to determine the effect of window

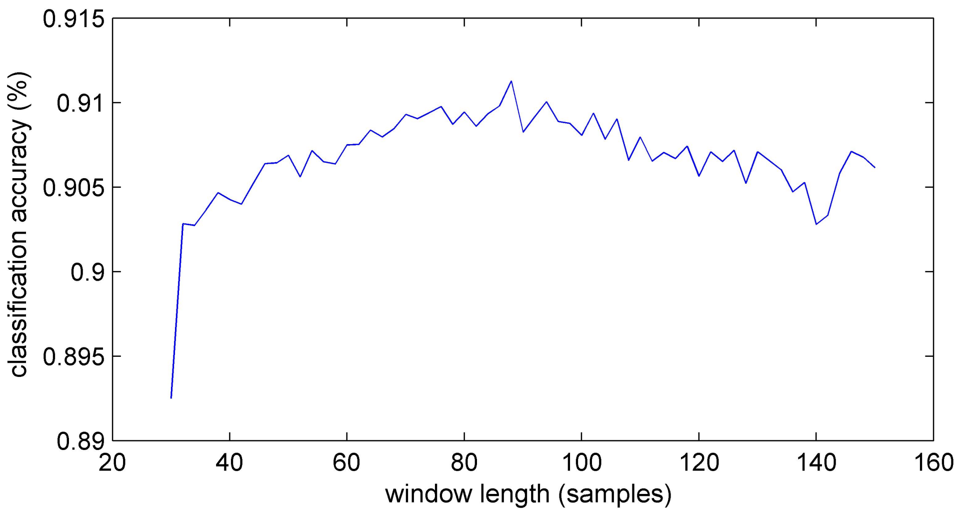

Figure 5.4: Sensitivity of the KNN algorithm to the choice of window size.

length on classification accuracy. The algorithm had a maximum personalized classification accuracy at 88 samples per window (Figure 5.4).

Overall classification accuracy for each subject using the general K-Nearest Neighbor algorithm ranged from 83.0% to 95.8% (Table 5.2). Mean classification accuracy was 90.7% (SD=4.2%). Classification accuracies for all but three subjects were higher than 90.0%. Subjects 3, 5, and 11 had lower classification accuracy. Accuracy for these subjects were 83.0, 84.5, and 85.6 % respectively. The most commonly mis-classified activity was sitting, which was typically confused with standing. The confusion matrix (Figure 5.5) shows the percentage of actions that were classified correctly or mis-classified as other activities.

For the personalized classifier, mean accuracy was 91.5% (SD=3.7%). This is 0.8% higher than the generalized classifier. Classification accuracy increased for all subjects. Subjects 3 and 5 still had classification accuracies below 90%, but subject 11 classification accuracy increased by over 10% to 96.1% correct. The confusion matrix (Figure 5.5) shows that the confusions for all activities decreased(accuracy increased). Sitting is still the most commonly confused activity, and it is still most often confused with standing.

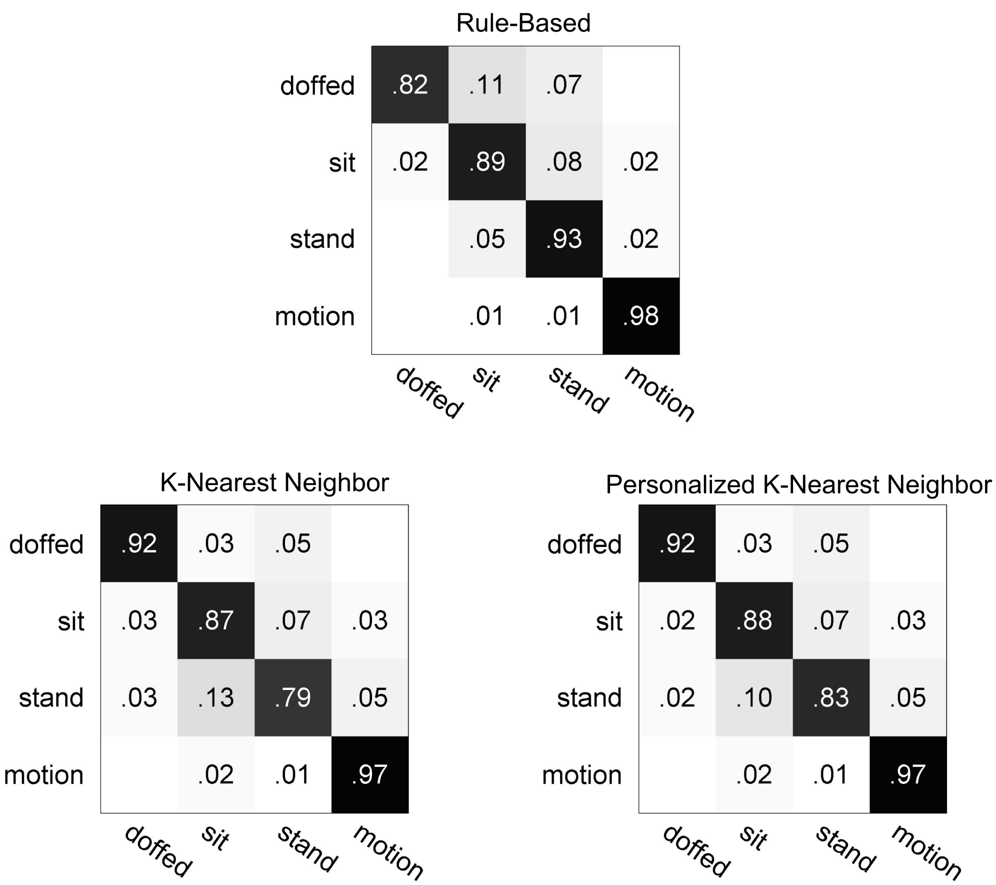

Figure 5.5: Confusion matrices for the rule-based and KNN classification methods.

Table 5.3: Classification accuracy for all subjects using the HMM methods.

Subject	HMM Offline	Personalized HMM Offline	HMM Online	Personalized HMM Online
1	96.8	97.3	95.0	95.1
2	93.7	96.3	95.1	95.3
3	90.3	83.4	79.9	82.1
4	89.3	93.2	92.3	93.1
5	78.3	89.9	73.8	89.5
6	94.7	93.7	93.8	91.1
7	80.0	96.5	85.7	95.7
8	88.2	90.6	89.2	90.0
9	86.6	92.9	84.8	90.0
10	93.5	93.5	90.2	91.3
11	79.9	85.5	84.7	83.6
Mean	88.3	92.1	87.7	90.6
SD	6.2	4.2	6.4	4.2

Hidden Markov Model

Overall classification accuracy for each subject using the generalized offline HMM algorithm ranged from 79.9% to 96.8% (Table 5.3). Mean classification accuracy was 88.3% (SD=6.2%). The most commonly mis-classified activity was sitting, which was typically confused with standing. The confusion matrix (Table 5.5) shows the percentage of actions that were classified correctly or mis-classified as other activities. Only five of the subjects produced data that could be classified with accuracies higher than 90.0%. The others, subjects 4, 5, 7, 8, 9, and 11, all produced data that could be classified with accuracy ranging from 78.3% to 89.3%.

Overall classification accuracy for each subject using the generalized online HMM algorithm ranged from 73.8% to 95.1% (Table 5.2). Mean classification accuracy was 87.7% (SD=6.4%). The most commonly mis-classified activity was sitting, which was typically

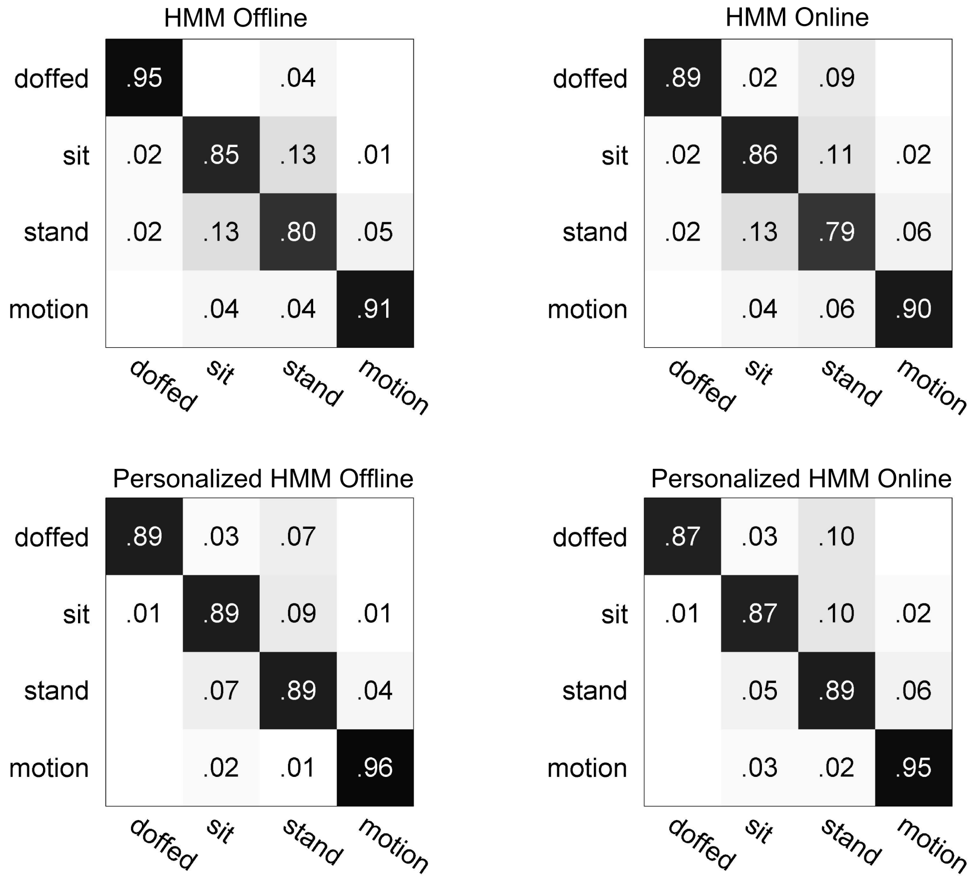

Figure 5.6: Confusion matrices for the HMM classification methods.

confused with standing. The confusion matrix (Table 5.6) shows the percentage of actions that were classified correctly or mis-classified as other activities.

The online HMM classification produced results that were lower on average than the offline HMM classification for most subjects, but subjects 2, 4, 7, 8, and 11 all had increased classification accuracy with the online classification. The average drop in classification accuracy due to using the online classification algorithm was 1.0% (SD=4.3%). The highest drop in accuracy was 2.3% for subject 9, and the lowest drop was 10.1% for subject 3. The highest gain in accuracy was 4.7% for subject 11.

The use of personalization for the Hidden Markov Model classifier improved classification accuracy for the majority of subjects in both offline and online classification. Personalized offline classification had a mean accuracy of 92.1% (SD=4.2%), while personalized online classification had a mean accuracy of 90.6% (SD=4.2%). The variance in accuracies decreased with the use of personalization.

Chapter 6

DISCUSSION

The choice of prosthesis components and alignment can have a large effect on the quality of life of the prosthesis' user. A well-fitting prosthesis is crucial to reducing pain and enabling people to take part in their desired daily activities. In choosing prosthesis components and alignment, prosthetists must take into account how often the person will wear their prosthesis, the durations of sitting and standing that the person will do, and the amount of activity that a person engages in throughout the day. The recent focus on evidence based practice has also increased the importance of monitoring prosthesis use both before and after a change to the prosthesis. Currently available technologies for monitoring prosthesis use have several limitations, including short battery lives, small data storage capacity, and an inability to monitor stationary postures. Those monitors that are capable of measuring posture and activity for long periods of time require strict subject adherence to the monitor wear guidelines. The methods described here make use of an off-the-shelf sensor and custom algorithms to monitor activity and posture when the sensor is mounted on the prosthesis. By mounting the sensor on the prosthesis, user involvement in the monitoring process is minimized and the chance of corrupted or missing data is reduced.

The results of this study show that there exist at least three algorithms that perform at approximately 90% accuracy when classifying activities as one of doffed, sitting, standing, or in motion. All of the classification methods were able to classify movement with greater than 95% accuracy, but the classification of stationary postures and prosthesis doffed were less accurate. This was expected, as an accelerometer placed at an ankle does not afford as much information about body posture as one placed at the trunk of a person's body. Nonetheless, a classification accuracy of approximately 90% may be acceptable for clinical and research use.

All of the classification methods sometimes misclassified sitting and standing as motion.

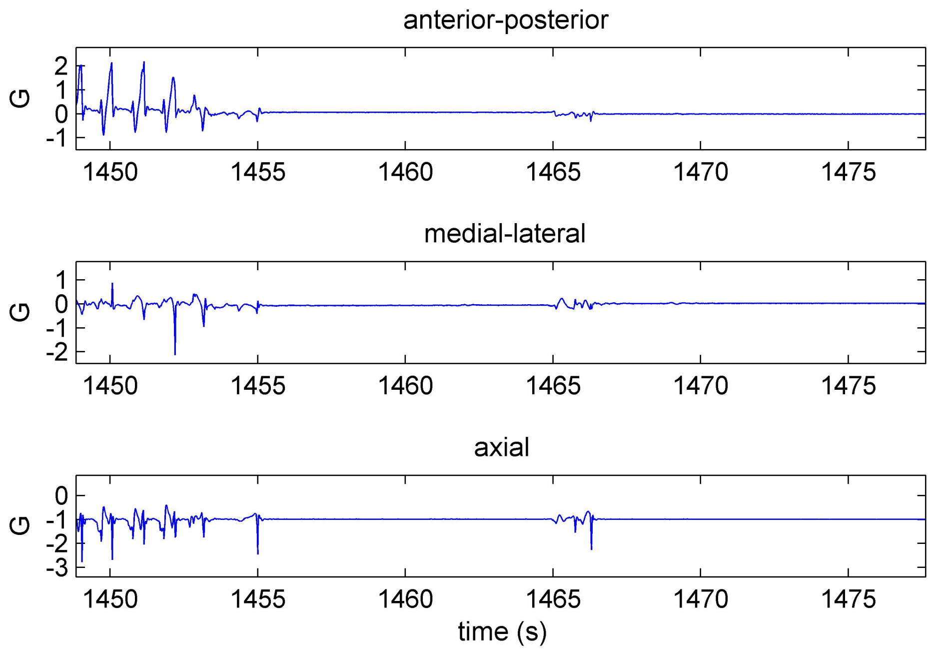

Figure 6.1: Plot of acceleration data from subject 5. The plot shows data where the subject is walking, stops, and sits down. After sitting down, periodic accelerations can be seen in all three axes at approximately 1465s. Subjects were not completely immobile while sitting, which led to the mis-classification of some periods of sitting as motion.

This is most likely because subjects did not sit or stand motionlessly. Subjects often shifted, shuffled, and readjusted their limbs while sitting. This led to large increases in SMA (Figure 6.1). Presumably this type of mis-classificaiton could be minimized by increasing the window size used to generate SMA. This would result in lower SMA readings for readjustment while sitting, but approximately the same SMA readings for periods of walking or using the stairs. However, as shown in Figure 5.1 and Figure 5.4, longer window lengths led to slightly lower classification accuracies. This indicates that there may be a tradeoff between time resolution of activities and accuracy of individual windows.

Another potential way to minimize misclassifications caused by somebody moving their leg while sitting is to include classification of transitions, such as transitioning from sitting to standing or vice versa. This could be used to rule out walking during a period of time between a stand-to-sit transition and a sit-to-stand transition. In order for this to be effective, the identification of sit-to-stand and stand-to-sit transitions must be very accurate. Otherwise it would introduce noise to the surrounding classifications. Match filtering may work well to identify these transitions [24].

6.1 Discussion of individual methods

The different classification methods all had mean classification accuracies within one standard deviation of one another. However, a closer examination of the methods is warranted because they seemed to do well or poorly classifying different activities or on data from different subjects.

6.1.1 Rule based method

The rule based method investigated here showed very high classification accuracy for some subjects, but also very low classification accuracy for some subjects. This subject to subject variability is most likely due to differences in the default postures that these subjects assume. Some subjects consistently leave their prosthesis pointed far forwards of their body while sitting, which enables accurate classification of the sitting posture. Other subjects, notably Subject 11, leave their prosthesis directly under their knee even when sitting (Figure 6.2). Due to the high variability in how individuals sit and stand, a one-size fits all approach to rule based activity classification may be sub-optimal. Potential solutions to this issue include personalized rules for different individuals or a more complex decision rule to identify sitting. Including personalization in a rule based classification method may make the method prohibitively complex, since new subjects would require researchers to manually edit the classification system.

The rule based method was also challenged to identify periods of time when the subjects had doffed their prosthesis. Identifying doffed periods is done in two ways using this method. One is to identify periods when the prosthesis is in the doffed reference position, which is prone to error when the subject is sitting or standing. The other is to identify long periods of no motion as measured by low Signal Magnitude Area. Identifying periods of no motion is more accurate than orientation, but also requires the choice of a suitable duration time

Figure 6.2: Subject 11 in his chosen sitting posture. He consistently leaves his prosthesis directly under his knee in all seats, but allows his knees to bend outwards. This lowers the classification accuracy of the rule based classifier for his sitting postures.

limit. If the time limit is chosen to be too short, then periods of standing or sitting may be identified as doffed. If the time limit is chosen to be too long, then short periods where the prosthesis is doffed may be missed. Short periods of doffing may be important, as some subjects doff their prosthesis periodically to change socks and this would be an important feature to note for prosthetists and researchers.

6.1.2 K-Nearest Neighbor

Because the K-Nearest Neighbor algorithm had high accuracy and low variance among subjects, KNN or a similar method may be the most useful for clinical applications. KNNs were analyzed in two different ways. In one analysis, only generalized training data from other subjects was used. In the second personalized training data was used that included some data from the subject being tested.

The *generalized K-Nearest Neighbor* algorithm had a mean accuracy of 97.0% for the in motion activity. The classification accuracy of the doffed activity was 92.0%, but the classification accuracy of the standing activity was only 79%. Standing was often mis-classified as sitting, and was sometimes mis-classified as doffed. These confusions may be important, and their implications should be taken into account if the method is used for any specific application. The standard deviation of classification accuracy for the K-Nearest Neighbor algorithm was only 4.2%, meaning that it was the most consistent method.

When using a *personalized K-Nearest Neighbor* algorithm, the classification accuracy of periods when the prosthesis was doffed or when it was in motion remain as accurate as the generalized KNN method. The classification accuracy of sitting increases slightly, but the classification accuracy of standing increases by 4%. The mean accuracy rose to 91.5% and the standard deviation dropped to 3.7%. Personalization improved the classification accuracy for all subjects.

The success of the KNN method indicates that the correlational structure of these activities is not very important in classifying data from prosthesis users. The KNN method was chosen because of its lack of correlational structure and simplicity. Because the KNN produced the best results and had the lowest variance among subjects, it may be of interest to investigate the use of more efficient algorithms. A support vector machine may provide equivalent accuracy for lower computational cost during classification [37].

6.1.3 Hidden Markov Model

The preliminary investigations made into classifying activities using an HMM are promising. The representation of observations by Gaussian mixtures allows for classification accuracy that is comparable to that obtained with a KNN or rule-based algorithm. Furthermore, the use of an HMM or similar graphical model on classifications made using other classifiers may improve accuracy by smoothing the output classifications and removing short periods of mis-classification.

Four different analyses were performed using the Hidden Markov Model classifier. These included the various permutations of offline or online and personalized or generalized. Generalized online HMM classification performed poorly compared to all other methods, while personalized offline classification performed well compared to all other methods. These results indicate that if classification is to be performed using an HMM, then personalization may need to be used to ensure good results.

The *generalized offline* algorithm had the highest classification accuracy for periods of time when the prosthesis was doffed, but it also had the lowest classification accuracy for periods of time when the prosthesis user was sitting. Sitting and standing were often confused. The low accuracy for sitting and standing was common among all subjects. Classification of periods of movement was approximately 91%, which was lower than the rule based and KNN methods. This is most likely because subject 5 walked slowly and with short steps, and thus had a low SMA. The low SMA for that subject resulted in many misclassifications.

The *generalized online* algorithm had the lowest classification accuracy of all methods tested. Using this method, several subjects had classification accuracies below 80%, and only five subjects had classification accuracy above 90%. Sitting and standing were often confused with each other, and periods when the prosthesis was doffed were confused with both sitting and standing.

The *personalized offline* algorithm performed well, with a mean classification accuracy of 92.1%. The classification of sitting and standing was more accurate than any other method. Personalization also allowed the movement of subject 5 to be taken into account, increasing accuracy for movement overall. Interestingly, the inclusion of personalization caused classification accuracy for periods when the prosthesis was doffed to drop from 95% for the generalized offline HMM to only 89% for the personalized offline HMM. This misclassifications of sitting and standing as doffed also decreased when using personalization, so it seems that the generalized method erred by classifying periods of use as doffed more.

The *personalized online* algorithm performed better than either generalized HMM algorithm, but otherwise performed poorly. Periods when the prosthesis was doffed were confused with sitting and standing, which were each confused with each other.

6.2 Suitability for online classification

While offline classification of periods of prosthesis use is of interest to clinicians and researchers so that they correlate activities to prosthetic interventions, online classification of prosthesis use has the capability to inform active prosthesis elements. Several prosthesis components (notably active knees and active ankles [64]) already attempt to identify the activity of the wearer so that they can adjust. Classification of activities as studied here could potentially be used with an active volume adjustment socket to improve the health of the residual limbs of prosthesis users.

The rule based and KNN classification methods are by nature online systems. They do not depend on future knowledge to make accurate classifications of activity. The HMM method, on the other hand, uses the Viterbi algorithm to determine the activity that produced each observation vector. The Viterbi algorithm does make use of future values, and could thus not be easily implemented on an active prosthesis. The forward algorithm used by online decoding for HMMs could be used for online classification, but results in lower classification accuracy. Our results show that the loss in accuracy is approximately 1.5% when using personalization (Table 5.2). This could potentially be improved upon by modifying the graphical model or using the model on the output of the KNN or rule-based classifiers to achieve temporal smoothing.

6.3 Personalization

Personalization increased mean activity classification accuracy and decreased variance across different subjects for both KNN and HMM methods. Many of the subjects in the study produced data that could be classified with accuracies of approximately 95% using a general model, but other subjects produced data that could be classified only at approximately 80%. The subjects who already had high classification levels gained approximately 1% of accuracy when personalization was used, but the other subjects had significantly improved accuracy. One subject even had an increase of approximately 10% when using a personalized KNN model instead of a general KNN model. This could be explained by two different phenomena.

1. The first is that ten subjects is not sufficient to train the algorithms to accurately recognize activities among many different people. This may the case if activities among different people are fairly similar. If this is the case, then a non-personalized classifier could be constructed that would work as well as a personalized classifier if more data from more subjects were collected.

2. The second possibility is that each person's methods of sitting, standing, and walking are different enough that personalization is required to guarantee classification accuracies higher than 90%. If this is the case, then all new subjects must go through a test protocol to collect data to personalize the algorithms before the method can be used for long term monitoring.

If the first conclusion is true and data from more subjects will increase the generalizability of the classification methods, then more data should be collected. If the second conclusion is true, then the collection of more data will lead to small improvements at best. More research is needed to determine which condition holds.

6.4 Potential Improvements

6.4.1 Rule-based method

Improving the rule-based classifier could be done by revising the heuristics it uses. Careful analysis must be made of the data and the heuristics improved based on patterns noticed. The rule-based system performed with very high accuracy for most subjects, but some subjects had lower classification accuracy. The decision heuristics could be modified to account for the differences seen among different subjects.

Sitting and standing, the most commonly mis-classified activities, are currently classified by determine whether the anterior-posterior acceleration is greater than or less than the acceleration in a reference position. Because only one axis is taken into account, subjects who sit in an unexpected way, like subject 11, had low classification accuracy for sit and stand activities. This could be improved by creating volumes in the state-space of prosthesis orientations which correspond to sit or to stand, instead of simply thresholding on one axis.

6.4.2 Machine Learning Methods

Machine learning methods of classification can generally be improved by training them on larger sets of data. By performing the protocol described here on more subjects, or developing a longer protocol for the same subjects, classification accuracy could be improved. Additional refinements to the features used and to the graphical model itself should be investigated to determine if even higher classification accuracy is possible. Other improvements are more specific to the type of classifier.

K-Nearest Neighbor

The K-Nearest Neighbor method was chosen because it is simple to implement and often performs well on classification problems. It has several drawbacks, including requiring large amounts of training data and performing slowly during classification. A more sophisticated method, such as a Support Vector Machine, may maintain the high accuracy of the KNN method while decreasing the time and computation power required for classification.

Hidden Markov Model

There are many refinements that could be made to the Hidden Markov Model classification method to improve the results. Two of these refinements that may result in the largest gains are duration modeling and the use of a bigram model for activities.

HMM duration modeling may improve the results from the hidden Markov model classifier [65]. Duration modeling consists of determining a probability distribution over the duration that subjects perform a given activity. This can be used when constructing the transition matrix of the Markov chain, which would later inform the classifier. Duration modeling may be applicable because many of the activities in which subjects engage will be of similar duration from day to day. However, duration modeling will require a different activity protocol that encourages subjects to spend as much time in each activity as they normally would while performing the experiment.

Preliminary analysis of the utility of duration modeling has been performed on long term acceleration data. Two subjects, subjects 6 and 10, were fitted with an GT3X+ as described in this study. They were asked to go about their daily lives for three weeks. When they returned, the data was downloaded and analyzed using the rule-based classifier. The durations of the activities were determined. While these durations should be interpreted with care due to the imprecise nature of their collection, general conclusions can be drawn from their distribution. In general, activities are very likely to be short, though there are some activities that can be quite long (Figure 6.3). The large number of activities with very long durations are all periods when the prosthesis was doffed.

This finding is consistent with the literature, in which it has been found that activities are primarily very short in duration. Orendurff, et. al., have shown that rest bout duration (duration of sitting or standing events) has a distribution that could be modeled using a geometric distribution. Walking duration has a distribution that could be modeled with a negative binomial distribution [45].

A standard HMM models durations as geometric distributions. Based on data from [45], the classification accuracy of the HMM could potentially be improved by including explicit duration modeling for the in-motion and doffed activities. Periods of time when the

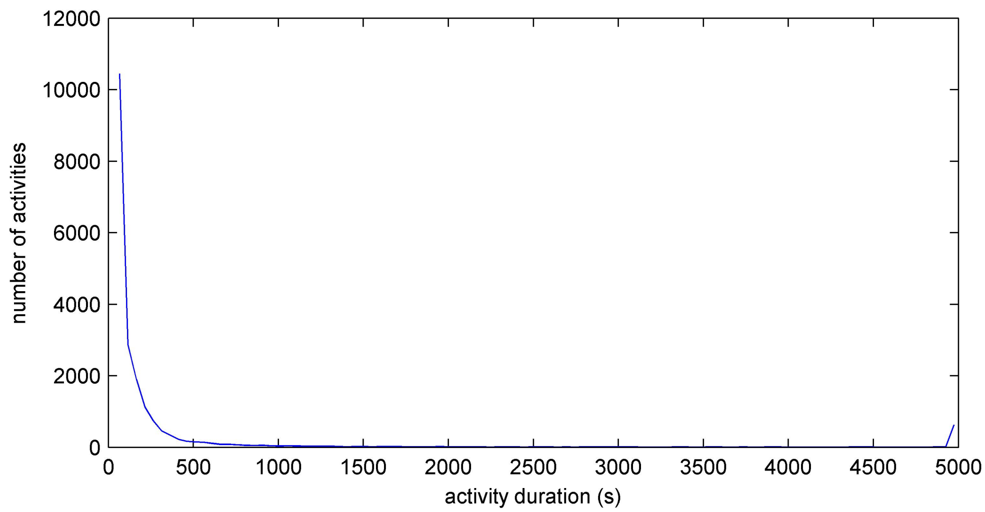

Figure 6.3: A histogram of the activity durations for three weeks of data each from subjects 6 and 10.

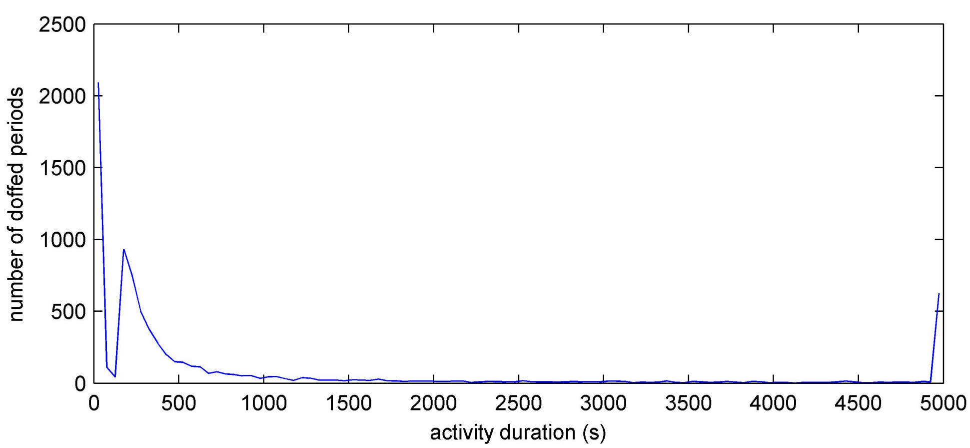

Figure 6.4: A histogram of the doffed durations for three weeks of data each from subjects 6 and 10.

prosthesis was doffed have a multi-modal distribution (Figure 6.4). Periods of time where the prosthesis was in motion could be modeled with a negative binomial distribution. The distributions of activities found in [45] were made with window sizes of ten seconds, which indicates that a longer window size than that used in the analysis here may be beneficial for HMM classification.

Another change that may improve the classification accuracy of the HMM is the use of a bigram model (Figure 6.5). Strictly speaking, this is a more complex model than the simple Hidden Markov Model described here. In a bigram model, the probability of transitioning to some state from the current state depends on both the current state and the most recent different state. This form of modeling may improve accuracy if people often perform actions in the same order, for example if they often walk somewhere, sit, and remove their prosthesis. Taking this structure into account may improve the classifier.

Furthermore, an HMM or related graphical model could be used to improve the other classifiers. Because the rule-based and KNN classifiers treat each window as independent, there is the possibility that a single window may be classified as one activity even if it within two long periods of a different activity. Dynamic graphical models could allow for temporal smoothing of the classifications, reducing the probability of such short duration errors [34, 35].

6.5 Suitability for other uses

The use of commodity hardware in this study allows the method to be easily implemented for prosthetists and prosthetics researchers. Because other accelerometers can also be used with the same algorithms, the choice of hardware can be made based on other factors, such as desired battery life, cost, or data storage capacity. Interest in other activities or application domains could also be addressed with the same hardware and algorithms.

Classification of other activities than those investigated here is possible. Differentiation of level walking, walking upstairs, walking downstairs, or walking on non-level ground are all activities that may be of interest to prosthetists and prosthetics researchers. All of the classification methods investigated here could be extended to classify other activities, or to classify sub-types of activity. To do this, more data would need to be collected and

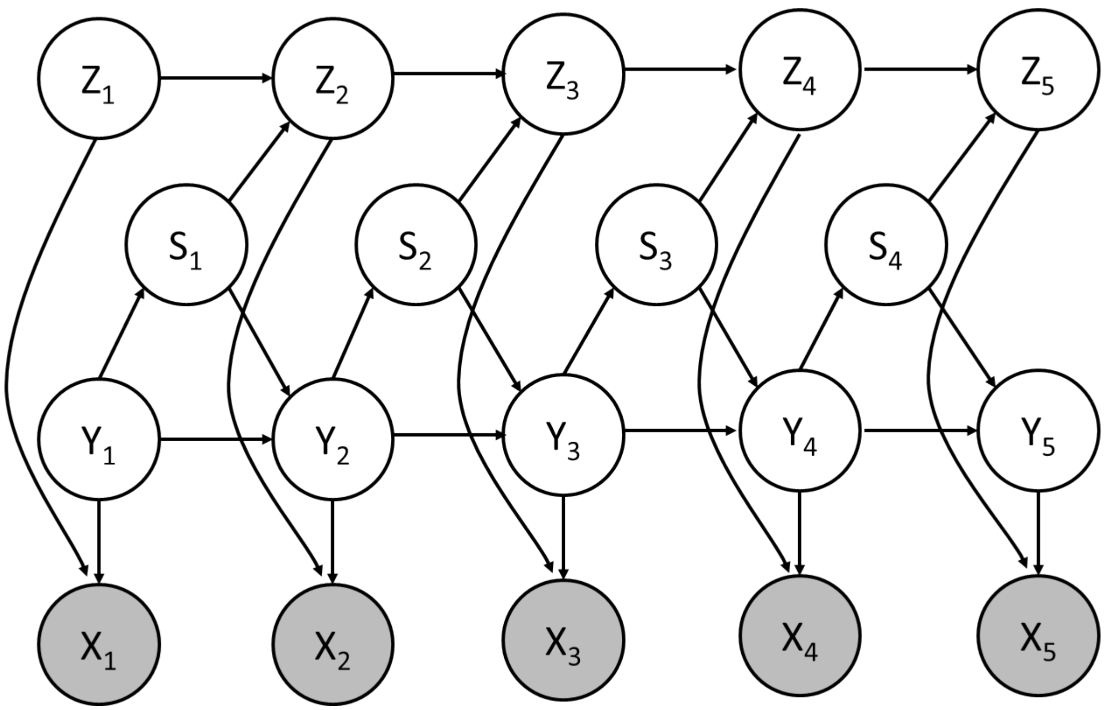

Figure 6.5: A bi-gram dynamic graphical model. The Z nodes represent the activity that was performed prior to the current activity. These Z nodes do not necessarily change at every time step; they change only when the activity state Y changes from one activity to another. This is mediated by the switching nodes S.

additional heuristics or features investigated.

While all three different algorithms tested here are capable of being extended to classify other activities, the limitations of the accelerometer used must be taken into account. The GT3X+ has a dynamic range of only ±6G, but ±10G is needed to accurately characterize running [29]. These methods may have difficulty differentiating running and walking based on data from this accelerometer. If classification of running is of interest, more research should be performed to validate the use of the current accelerometer or to test a different accelerometer.

While the differentiation of sitting and standing using data from a pylon-mounted accelerometer was difficult, a thigh mounted accelerometer may provide data that is easier to interpret. When someone is sitting, their upper leg is in a much different position than it is when they are standing. This is difficult to make use of in people with trans-tibial amputations, because attaching an accelerometer to the thigh leads to high compliance requirements. However, the method may be useful for people with trans-femoral amputations. A similar system to the one described here may have better accuracy for trans-femoral prosthesis users.

The method of instrumenting a prosthetic device and using the resulting sensor data to classify activity could also be extended to other assistive and orthotic devices. Potential applications include the instrumentation of canes, walkers, or wheelchairs. Activity classification for the use of these devices may allow rehabilitation specialists and other medical professionals to tailor their care to the actual behaviors of their patients.

Chapter 7

FUTURE WORK

This research is intended to serve as a starting point for activity classification methods that make use of only sensors mounted on prostheses. Current results are encouraging, and the approach should be further explored, validated, and then implemented for use in the field.

Additional research is needed before any of these classification strategies can be recommended for clinical use. The methods must be validated more thoroughly while prosthesis users move through their free-living environments. The behavior of subjects may change when they are in their home environment, and it is important to verify that these methods continue to perform adequately if that is the case.

One potential limitation in use of the developed posture and activity classification strategy is the unknown effect of riding in a motorized vehicle. Further research is required to isolate or account for external accelerations to which a user may be subjected while riding in vehicles. Additionally, other postures (e.g., lying down) or specific activities (e.g., stair climbing) of interest to researchers should be explored to ensure they can be appropriately classified with this system.

Another potential limitation is that subject postures and behaviors may change over time. If a personalized classifier is used, it is unknown if it will remain accurate over the long term. Subjects who lose weight or undergo rehabilitation therapies may sit or stand in ways that were not represented in their training data. The effects of weight loss or rehabilitation on this system should be investigated, and subjects who are experiencing lifestyle changes while they are being monitored should be monitored more closely.

The addition of other sensors may also be useful in improving classification accuracy. While other sensors may decrease the battery life or duration that data can be recorded for, the additional robustness or classification accuracy may be worthwhile for some applications. One sensor that seems particularly promising is the force sensitive resistor. When placed between the prosthesis' socket and the subject's residual limb, it can provide data on interface forces that may prove useful in accurate determination of posture.

Once an algorithm has been developed that accurately classifies activities of interest to clinicians and prosthetists, it must be tested in the daily lives of prosthesis users. The activity protocol used in this study is limited in that it consisted of subjects moving about primarily indoors, primarily for no purpose of their own. It may be the case that when subjects are in their free-living environment, they perform activities in a different manner. A longer term, free-living experiment would also provide data that could be used in duration modeling for the HMM classifier, as described above.

Clinical uses of this technology (e.g., rehabilitation training, componentry evaluation, etc.) should be explored to determine if the accuracy of which this system is capable is sufficient for such applications. The long battery life and large storage capacity of the ActiGraph GT3X+ accelerometer suggest that this sensor may be suitable for long-term data collection. Anticipated advances in battery and storage technologies will also likely extend the length of time such sensors can be used to monitor subjects. Such research will help to determine if prosthesis-integrated monitoring systems can enhance clinical care and improve quality of life for prosthesis users.

BIBLIOGRAPHY

[1] C. Angeloni, P.O. Riley, and D.E. Krebs. Frequency content of whole body gait kinematic data. *IEEE Transactions on Rehabilitation Engineering*, 2(1):40–46, March 1994.

[2] EK Antonsson and RW Mann. The frequency content of gait. *Journal of Biomechanics*, 18(1):39–47, 1985.

[3] AppsZoom. Activity Tracking Android Apps.

[4] Louis Atallah, O Aziz, GZ Yang, and Benny Lo. Detecting walking gait impairment with an ear-worn sensor. In *6th International Workshop on Wearable and Implantable Body Sensor Networks*, pages 175–180, Berkeley, California, 2009.

[5] Louis Atallah, Omer Aziz, Edward Gray, Benny Lo, and Guang-Zhong Yang. An ear-worn sensor for the detection of gait impairment after abdominal surgery. *Surgical innovation*, 20(1):86–94, February 2013.

[6] H B Bussmann, P J Reuvekamp, P H Veltink, W L Martens, and H J Stam. Validity and reliability of measurements obtained with an "activity monitor" in people with and without a transtibial amputation. *Physical Therapy*, 78(9):989–98, September 1998.

[7] J B J Bussmann, W L J Martens, J H M Tulen, F C Schasfoort, H J G van den Berg-Emons, and H J Stam. Measuring daily behavior using ambulatory accelerometry : The Activity Monitor. *Behavior Research Methods, Instruments, & Computers*, 33(3):349–356, 2001.

[8] Johannes B Bussmann, Eleonore a Grootscholten, and Henk J Stam. Daily physical activity and heart rate response in people with a unilateral transtibial amputation for vascular disease. *Archives of Physical Medicine and Rehabilitation*, 85(2):240–244, February 2004.

[9] Johannes B Bussmann, Hannelore J Schrauwen, and Henk J Stam. Daily physical activity and heart rate response in people with a unilateral traumatic transtibial amputation. *Archives of Physical Medicine and Rehabilitation*, 89(3):430–4, March 2008.

[10] LLC. CGS Administrators. Juisdiction C Supplier Manual. chapter 16, pages 4–5. Nashville, TN, 2011.

[11] Teri G. Rosenbaum Chou, Joseph B. Webster, Maryam Shahrebani, Toni L. Roberts, and Roy D. Bloebaum. Characterization of Step Count Accuracy of Actigraph Activity Monitor in Persons With Lower Limb Amputation. *Journal of Prosthetics and Orthotics*, 21(4):208–214, October 2009.

[12] Elizabeth Condie, Helen Scott, and Shaun Treweek. Lower limb prosthetic outcome measures: a review of the literature 1995 to 2005. *Journal of Prosthetics and Orthotics*, 18(6):13–45, 2006.

[13] A Cortes, E Viosca, J V Hoyos, J Prat, and J Sanchez-Lacuesta. Optimisation of the prescription for trans-tibial (TT) amputees. *Prosthetics and Orthotics International*, 21(3):168–174, 1997.

[14] Davide Curone, Gian Mario Bertolotti, Andrea Cristiani, Emanuele Lindo Secco, and Giovanni Magenes. A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity. *IEEE Transactions on Information Technology in Biomedicine : a Publication of the IEEE Engineering in Medicine and Biology Society*, 14(4):1098–105, July 2010.

[15] Yaghoub Dabiri, Siamak Najarian, and M Reza Eslami. Effects of Mass and Momentum of Inertia Alternation on Individual Muscle Forces During Swing Phase of Transtibial Amputee Gait. *Kobe Journal of Medical Science*, 56(3):92–97, 2010.

[16] Nancy L. Dudek. Ambulation monitoring of transtibial amputation subjects with patient activity monitor versus pedometer. *Journal of Rehabilitation Research and Development*, 45(4):577–586, December 2008.

[17] Paul L Enright. The six-minute walk test. *Respiratory care*, 48(8):783–5, August 2003.

[18] Davide Figo, Pedro C. Diniz, Diogo R. Ferreira, and João M. P. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662, March 2010.

[19] Hannah Fischer. U . S . Military Casualty Statistics : Operation New Dawn , Operation Iraqi Freedom , and Operation Enduring Freedom. Technical report, 2010.

[20] Inc. Fitbit. Fitbit.

[21] Laurent Frossard, Nathan Stevenson, John Sullivan, Maggie Uden, and Mark Pearcy. Categorization of Activities of Daily Living of Lower Limb Amputees During Short-Term Use of a Portable Kinetic Recording System: A Preliminary Study. *Journal of Prosthetics and Orthotics*, 23(1):2–11, January 2011.

[22] A Godfrey, R Conway, D Meagher, and G OLaighin. Direct measurement of human movement by accelerometry. *Medical Engineering & Physics*, 30(10):1364–86, December 2008.

[23] Brian J Hafner, Laura L Willingham, Noelle C Buell, Katheryn J Allyn, and Douglas G Smith. Evaluation of function, performance, and preference as transfemoral amputees transition from mechanical to microprocessor control of the prosthetic knee. *Archives of Physical Medicine and Rehabilitation*, 88(2):207–17, February 2007.

[24] Robert Hecht-nielsen. Nearest matched filter classification of spatiotemporal patterns. *Applied Optics*, 26(10):1892–1899, 1987.

[25] Siobhan C Hillery and Eric S Wallaceœ. Trans-tibial amputee gait adaptations as a result of prosthetic inertial manipulation. *Disability and Rehabilitation*, 22(8):383–386, 2000.

[26] Jawbone. Jawbone Up.

[27] Dean M Karantonis, Student Member, Michael R Narayanan, Merryn Mathie, Nigel H Lovell, Senior Member, and Branko G Celler. Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):156–167, 2006.

[28] Károly Kaszala and Kenneth a Ellenbogen. Device sensing: sensors and algorithms for pacemakers and implantable cardioverter defibrillators. *Circulation*, 122(13):1328–40, September 2010.

[29] Justin J Kavanagh and Hylton B Menz. Accelerometry: a technique for quantifying movement patterns during walking. *Gait & posture*, 28(1):1–15, July 2008.

[30] K Kiani, C J Snijders, and E S Gelsema. Computerized analysis of daily life motor activity for ambulatory monitoring. *Technology and health care : official journal of the European Society for Engineering and Medicine*, 5(4):307–18, October 1997.

[31] JR Kwapisz, GM Weiss, and SA Moore. Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations*, 12(2):74–82, 2011.

[32] Sheridan Laing, Peter Vs Lee, and James Ch Goh. Engineering a trans-tibial prosthetic socket for the lower limb amputee. *Annals of the Academy of Medicine, Singapore*, 40(5):252–9, May 2011.

[33] W R Ledoux and H J Hillstrom. Acceleration of the calcaneus at heel strike in neutrally aligned and pes planus feet. *Clinical Biomechanics (Bristol, Avon)*, 16(7):608–13, August 2001.

[34] Jonathan Lester and Tanzeem Choudhury. A hybrid discriminative/generative approach for modeling human activities. In Alessandro Saffiotti Leslie Pack Kaelbling, editor, *International Joint Conference on Articial Intelligence*, pages 766–772, Edinburgh, Scotland, UK, 2005.

[35] Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A Practical Approach to Recognizing Physical Activities. In *Pervasive Computing*, pages 1–16, 2006.

[36] Suh-Jen Lin and Nisha Hathi Bose. Six-minute walk test in persons with transtibial amputation. *Archives of physical medicine and rehabilitation*, 89(12):2354–9, December 2008.

[37] Stephen Marsland. *Machine Learning: An Algorithmic Perspective.* CRC Press, Boca Raton, 2009.

[38] M J Mathie, a C F Coster, N H Lovell, and B G Celler. Detection of daily physical activities using a triaxial accelerometer. *Medical & Biological Engineering & Computing*, 41(3):296–301, May 2003.

[39] M.J. Mathie and BG Celler. Classification of basic daily movements using a triaxial accelerometer. *Medical & Biological Engineering & Computing*, 42(5):679–687, 2004.

[40] Michael R Menard, Margaret E McBride, David J Sanderson, and D. Duncan Murray. Comparative Biomechanical Analysis of Energy-Storing Prosthetic Feet. *Archives of Physical Medicine and Rehabilitation*, 73(5):451–8, 1992.

[41] Henk E J Meulenbelt, Pieter U Dijkstra, Marcel F Jonkman, and Jan H B Geertzen. Skin problems in lower limb amputees: a systematic review. *Disability and Rehabilitation*, 28(10):603–8, May 2006.

[42] Todd K. Moon. The Expectation Maximization Algorithm. *Signal Processing Magazine, IEEE*, 13(6):47–60, 1996.

[43] Inc. Nike. Nike+ Fuelband.

[44] SN Ooi and NAA Osman. Ambulatory Function Monitor for Amputees. In *5th International Conference on Biomedical Engineering*, pages 732–734, Kuala Lumur, Malaysia, 2011.

[45] Michael S. Orendurff. How humans walk: Bout duration, steps per bout, and rest duration. *The Journal of Rehabilitation Research and Development*, 45(7):1077–1090, December 2008.

[46] Kim Parker, R Lee Kirby, James Adderson, and Kara Thompson. Ambulation of people with lower-limb amputations: relationship between capacity and performance measures. *Archives of Physical Medicine and Rehabilitation*, 91(4):543–9, April 2010.

[47] Liliana E Pezzin, Timothy R Dillingham, Ellen J MacKenzie, Patti Ephraim, and Paddy Rossbach. Use and satisfaction with prosthetic limb devices and related services. *Archives of Physical Medicine and Rehabilitation*, 85(5):723–729, May 2004.

[48] Diane Podsiadlo and Sandra Richardson. The timed" Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society*, 39(2):142–148, 1991.

[49] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[50] Jebu Jacob Rajan. *Time series classification*. Ph. d., University of Cambridge, 1994.

[51] N Ramstrand and KÅNilsson. Validation of a patient activity monitor to quantify ambulatory activity in an amputee population. *Prosthetics and Orthotics International*, 31(2):157–166, 2007.

[52] Christian B Redd and Stacy J Morris Bamberg. A Wireless Sensory Feedback Device for Real-Time Gait Feedback and Training. *IEEE/ASME Transactions on Mechatronics*, 17(3):425–433, 2012.

[53] Soren Kamaric Riis. *Hidden Markov Models and Neural Networks for Speech Recognition*. PhD thesis, Technical University of Denmark, 1998.

[54] Joan E. Sanders and Stefania Fatone. Residual limb volume change: Systematic review of measurement and management. *Journal of Rehabilitation Research and Development*, 48(8):949, 2011.

[55] Joan E Sanders, Daniel S Harrison, Katheryn J Allyn, and Timothy R Myers. Clinical utility of in-socket residual limb volume change measurement: case study results. *Prosthetics and Orthotics International*, 33(4):378–90, December 2009.

[56] T Schoppen, A Boonstra, J W Groothoff, J de Vries, L N Göeken, and W H Eisma. The Timed "up and go" test: reliability and validity in persons with unilateral lower limb amputation. *Archives of physical medicine and rehabilitation*, 80(7):825–8, July 1999.

[57] M. Sekine, T. Tamura, T. Fujimoto, and Y. Fukui. Classification of walking pattern using acceleration waveform in elderly people. *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.00CH37143)*, 2:1356–1359, 2000.

[58] Ruud W Selles, Margriet a G Formanoy, Johannes B J Bussmann, Peter J Janssens, and Henk J Stam. Automated estimation of initial and terminal contact timing using accelerometers; development and validation in transtibial amputees and controls. *IEEE Transactions on Neural Systems and Rehabilitation Engineering : A Publication of the IEEE Engineering in Medicine and Biology Society*, 13(1):81–8, March 2005.

[59] Richa Sinha, Wim J a van den Heuvel, and Perianayagam Arokiasamy. Factors affecting quality of life in lower limb amputees. *Prosthetics and Orthotics International*, 35(1):90–6, March 2011.

[60] HJ Stam. A device for long term ambulatory monitoring in trans-tibial amputees. *Prosthetics and Orthotics International*, 19:53–55, 1995.

[61] Jacqueline M Stepien, Sally Cavenett, Leigh Taylor, and Maria Crotty. Activity levels among lower-limb amputees: self-report versus step activity monitor. *Archives of Physical Medicine and Rehabilitation*, 88(7):896–900, July 2007.

[62] Rita J van den Berg-Emons, Johannes B Bussmann, and Henk J Stam. Accelerometry-based activity spectrum in persons with chronic physical conditions. *Archives of physical medicine and rehabilitation*, 91(12):1856–61, December 2010.

[63] Harmen van der Linde, Cheriel J Hofstad, Alexander C H Geurts, Klaas Postema, Jan H B Geertzen, and Jacques van Limbeek. A systematic literature review of the effect of different prosthetic components on human functioning with a lower-limb prosthesis. *Journal of rehabilitation research and development*, 41(4):555–70, July 2004.

[64] Huseyin Atakan Varol, Frank Sup, and Michael Goldfarb. Multiclass real-time intent recognition of a powered lower limb prosthesis. *IEEE transactions on bio-medical engineering*, 57(3):542–51, March 2010.

[65] S.V. Vaseghi. State duration modelling in hidden Markov models. *Signal Processing*, 41(1):31–41, January 1995.

[66] Patrick L Walter. The History of the Accelerometer. *Sound and Vibration*, 41(1):84–93, 2007.

[67] Santosh G. Zachariah, Rakesh Saxena, John R. Fergason, and Joan E. Sanders. Shape and volume change in the transtibial residuum over the short term: Preliminary investigation of six subjects. *The Journal of Rehabilitation Research and Development*, 41(5):683, 2004.

[68] Kathryn Ziegler-Graham, Ellen J MacKenzie, Patti L Ephraim, Thomas G Travison, and Ron Brookmeyer. Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Archives of physical medicine and rehabilitation*, 89(3):422–9, March 2008.