

56

Use of a Wearable Ambulatory Monitor in the Classification of Movement States in Parkinson's Disease

by
David A. Klapper
M.D.

Albert Einstein College of Medicine, 1994

Submitted to the Division of Health Science and Technology
In partial fulfillment of the
requirements for the degree of
Master of Science
at the
Massachusetts Institute of Technology
August 2003

[September 2003]

Signature redacted

Signature of Author

David A. Klapper, MD

Signature redacted

Certified by

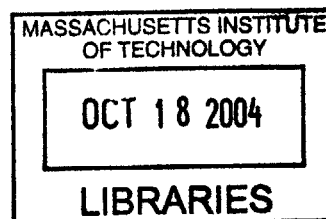
Lucila Ohno-Machado, MD, PhD

Signature redacted

Accepted by

Martha Gray, PhD

ARCHIVES



Abstract

For Parkinson's patients to function at their best, their medications need to be optimally adjusted to the diurnal variation of symptoms. For this to occur, it is important for the managing clinician to have an accurate picture of how the patient's bradykinesia / hypokinesia and dyskinesia fluctuate throughout the normal daily activities. This thesis proposes the use of wearable accelerometers coupled with machine learning and statistical techniques in order to classify the movement states of Parkinson's patients and to provide a timeline of how the patients fluctuate throughout the day. A pilot study was performed using 2 patients with the goal of assessing the ability to classify dyskinesia and bradykinesia / hypokinesia based on accelerometric data. The patients were observed and videotaped. Clinical observations of bradykinesia / hypokinesia and dyskinesia were noted every minute. Neural networks were able to classify better than classification trees with an average c-index (equivalent to the area under the ROC curve) of 0.905 for bradykinesia / hypokinesia and 0.926 for dyskinesia. A separate group of 5 patients were observed with the additional goal of building models that can classify the movement of a patient without requiring clinically annotated training data for the same patient. An enhanced protocol was used in the final study. Dichotomized linear regression was found to classify well with an average c-index of 0.8219 for body bradykinesia / hypokinesia and 0.8799 using as the gold-standard the patient's diary. Dyskinesia was classified at a c-index of 0.7522. Neural networks did not perform as well, possibly because of restrictions placed on adjusting parameters. The two most clinically important problems: predicting when the patient feels he/she is "off" or when he/she has "troublesome dyskinesia" were discriminated with c-indices of 0.96 and 1.0 respectively. The good result of the models despite the small number of patients is promising. Further studies with larger number of patients are therefore justified.

10/27/2014

Executive Summary

Background

Clinicians who care for Parkinson's patients must be able to manage and offset the hour-by-hour fluctuations in movement that these patients often experience. Symptoms of Parkinsonism such as bradykinesia, hypokinesia and akinesia, and medication-related side effects such as dyskinesia need to be reported to the clinician in a manner that accurately conveys the timing and severity of the symptoms. The clinician can then tightly adjust and titrate the timing and dosing of medication, allowing the patient to function at his or her best. Patient history and patient self-reporting diaries are currently used for this purpose, but they have problems related with patients' compliance, completeness and reliability. A monitor that could be worn by the patient while he or she is at home and could issue to the clinician a report of how the patient has been moving over the course of the day would be of great help.

Wearable devices have been studied for the measurement of movement in Parkinson's patients, but none have been designed in a manner that would be useful for the titration of medications. The data used to create the classification algorithms for these devices generally did not have the continual clinical annotation that would be needed to create a device that could produce a timeline of the patient's movement. Their classification algorithms were generally trained with data derived from structured tasks, and were therefore inappropriate for at-home ambulatory monitoring, which must be able to work in an unstructured environment. Their classification schemes generally address dyskinesia or tremor, but not bradykinesia / hypokinesia, which is clinically important. The few devices that attempt to detect bradykinesia / hypokinesia do not address them in a way that would be useful for adjusting medications. Furthermore, previous classification schemes generally used simplistic algorithms that could not address the complexity of this problem.

Previous Work

A pilot study was performed to demonstrate the feasibility of an accelerometer-based movement monitor for patients with Parkinson's disease. The device that was studied consisted of five 3-axis accelerometers attached to all 4 limbs, as well as to the hip. Two Parkinson's patients were observed by a neurologist for a cumulative total of 640 minutes. Each patient's state was recorded by a neurologist using a 5-point scale for bradykinesia / hypokinesia and a 5-point scale for dyskinesia. In addition, videotapes of the sessions were reviewed by the neurologist, allowing for the neurologist to resolve the state of bradykinesia / hypokinesia and dyskinesia down to a minute by minute basis. Features were derived from the raw accelerometric data based on the absolute of the derivative of magnitude of acceleration as well as position and magnitude correlation between sensors. The data were randomly divided into training and test sets and each of the two scales of the neurologist's annotations were dichotomized. Neural networks and classification trees were used to predict dykinetic vs. not-dykinetic states, as well as bradykinetic / hypokinetic versus not bradykinetic / hypokinetic states.

Neural networks were able to detect bradykinesia / hypokinesia on the test set with c-indices (area under the ROC curve) of 0.880 and 0.921 for patients number 1 and 2, respectively. Dyskinesia was detected with c-indices of 0.911 and 0.941. Classification trees detected bradykinesia / hypokinesia with accuracies (percentage of correct classifications) of 0.748 and 0.853 and dyskinesia with accuracies of 0.806 and 0.916. Other information that was found to be useful:

- Detecting bradykinesia / hypokinesia required data from more sensors (four) than did detecting dyskinesia (two).
- The hip sensor (sensing trunk motion) appeared to be the most important sensor.
- The left upper extremity was not important in the models

Main Study

These preliminary results strongly supported the idea that our device could be used to classify movement states in Parkinson's disease. Still, there were not yet data from enough patients in order to build a truly useful classifier. In the pilot study, the classification models were built from data obtained from the same patients that the classification models were tested on. In the real clinical "world", the device would have to classify the movement states of a patient without any previous data on that patient. This would require sampling a larger group of patients. The main study described here sought to demonstrate that a useful movement state classifier could be constructed without using training data from the patient in which it was going to be employed. For this purpose, 5 additional patients were recruited. An improved clinical scoring protocol was used, which included more accepted measures of scoring dyskinesia, bradykinesia and hypokinesia, as well as a diary used by the patient to report his/her symptoms. Additionally, the two pilot study patients were re-scored and included in some of the analysis. Linear regression and neural network regression models were built and tested. A true test set could not be used because of the small number of patients. This problem was remedied by using a "leave-1-out" method. All choices for variables and other parameters were fixed as guided by the pilot study before the analyses were performed.

Despite the relatively small sample of patients, the results were encouraging. The neural network models did not perform well, likely because I did not permit parameters of the neural networks to be adjusted in order to optimize performance (so as to prevent over-fitting). Linear regression, which does not have a need to adjust parameters, did perform well. The linear regression models had an average c-index of 0.8219 for predicting the level of body-bradykinesia / hypokinesia and an average c-index of 0.8799 for predicting the patient's state as manifest in his/her diary entry. Dyskinesia was not modeled as well, with an average c-index of 0.7552.

Remarkably, the main study models performed best at the tasks that were the most important. The general clinical consensus is that the patient's report (i.e. diary) is more important than what the clinician observes and that detecting "off" states and "troublesome dyskinesia" helps most with adjusting mediations. Fitting with that, the linear regression models were able to differentiate "off" states (from "on" states) with an average c-index of 0.96 and to differentiate "troublesome dyskinesia" (from all other states) with an average c-index of 1.0.

Conclusion

This project has shown that it is possible to classify movement states in Parkinsonian patients solely based on data obtained from a set of wearable accelerometers. The methods used in this project were able to classify the movement state of patients even if no clinical information were available on those patients. The methods performed reasonably well in discriminating different movement states from each other despite the fact that there were only a handful of patients used in constructing the classification models. It is entirely possible that the discriminative power of the models would grow if the models were based on a larger set of patients.

Table of Contents

1. Introduction

- 1.1 The Problem
- 1.2 The Goal

2. Background of Parkinson's Disease

- 2.1 Role of Medication
- 2.2 Terminology
- 2.3 The Prevalence and Incidence of the Main Symptoms
- 2.4 Rating Scores and Diaries
 - 2.4.1 Rating Scales
 - 2.4.1.1 Scales for Dyskinesia
 - 2.4.1.2 Scales for Tremor
 - 2.4.1.3 Scaling Hypokinesia / Bradykinesia
 - 2.4.1.4 Scales for Parkinsonism
- 2.5 Previous Work on Device-Based Monitoring
 - 2.5.1 Methods of Collecting Sensor Data
 - 2.5.1.1 Different Types of Sensors, Their Pros and Cons
 - 2.5.1.2 Sensor Placement
 - 2.5.1.2.1 Placement to Detect Dyskinesias
 - 2.5.2 Detecting Posture
 - 2.5.3 Detecting Tremor
 - 2.5.4 Data Assessment
 - 2.5.4.1 The Frequency Domain
 - 2.5.4.1.1 Characteristics of Frequency Spectrum
 - 2.5.4.2 Methods Used to Process the Data
 - 2.5.5 This Project's Pilot Study
 - 2.5.5.1 Methods and Materials
 - 2.5.5.1.1 The Device
 - 2.5.5.1.2 Scoring Scheme
 - 2.5.5.2 Results
 - 2.5.5.3 Analyses
 - 2.5.5.3.1 Dichotomization of Scores
 - 2.5.5.3.2 Pre-Processing
 - 2.5.5.3.3 Feature Extraction
 - 2.5.5.3.4 Time Window
 - 2.5.5.3.5 Description of Features
 - 2.5.5.3.6 Neural Networks
 - 2.5.5.3.7 Classification Trees
 - 2.5.5.4 Discussion of Pilot Data

3. Final Study

- 3.1 Major Changes to Data Collection for the Final Study
- 3.2 Major Changed to Data Processing for the Final Study
- 3.3 Methods and Materials
 - 3.3.1 Analyses
- 3.4 Results
- 3.5 Analyses
- 3.6 Discussion

4. Conclusion

5. References and Notes

1. Introduction

This project seeks to investigate the potential usefulness of a device to monitor Parkinson's patients. This monitoring system is developed to record the effect of Parkinson's medication on the patients' movement, enabling Parkinson's medications to be optimally adjusted by clinicians who utilize this information.

1.1 The problem

Parkinson's disease is a common disorder affecting at least 750,000 one million Americans⁴⁸. Parkinson's disease causes progressive difficulty in moving. This includes slowness of movement, decreased amount of movement and difficulty initiating movement. Oftentimes there is an associated tremor as well as balance and posture problems. These difficulties with movement can be quite debilitating for patients. They seriously affect Parkinson's patients' quality of life as well as their ability to perform necessary activities of daily living.

Parkinson's disease is, however, a treatable condition. Medications can improve the debilitating decrease of movement. Unfortunately, medications often cause serious side effects such as abnormal movements (e.g. chorea) and abnormal posturing (i.e. dystonia) that are in and of themselves debilitating. The effectiveness of the medication as well as the side effects of the medication is related to the concentration of the medication in the patient's brain. For instance, too low a concentration may not relieve the Parkinson's symptoms and too high a level may lead to the abnormal movements. As a patient's disease gets worse over time, medications become less effective and the effect of each dose lasts less. In addition, abnormal movements, that are side effects of medication, tend to increase. For these reasons, patients who have had Parkinson's disease for several years are often in a delicate balance between the benefit of medication and the side effects of medication. Even a slight change of dosing or timing of a patient's Parkinson's medication may have profound effects on how that patient is able to function.

In order to maintain this delicate balance of medication, the managing clinician needs to have accurate and reliable information about how the patient's movement changes throughout the day. Slow and decreased movements at a particular time of day may lead the clinician to increase medication at that time. Abnormal movements such as chorea and dystonia may require different types of medication adjustments depending on the timing and type of abnormal movement. Other

abnormalities such as freezing or rapid fluctuations (i.e. "on-off") require their own interventions. Unfortunately, a clinician who sees a Parkinson's patient in the office is only able to witness the patient at a single point in time and has no observational information about the patient's daily fluctuations at home. Clearly, the history given by the patient is very useful, but it is prone to recollection errors as well as many patients' difficulty in judging precisely what sort of abnormal or impaired movements they were having during the course of the day. Impaired cognitive function is common in Parkinson's patients and may make getting an accurate and reliable history even harder. Having a patient take a diary can be helpful. However, patient self-reporting diaries are difficult to comply with (often not completely filled out) and also suffer from the same problems that can cause inaccurate histories.

1.2 The goal

This project seeks to address the problem of collecting accurate and reliable information about how Parkinson's patients' movements fluctuate throughout their day. The specific goal of this project is demonstrate that a wearable device can properly classify movement states in Parkinson's disease. To accomplish this, I used a set of wearable sensors (accelerometers) that can measure patient's movements while they are performing their normal activities. The data collected by these sensors were analyzed by classification algorithms. A clinically useful classifier would require more data collection than might be possible over the course of a masters program. This is therefore just a "demonstration of concept." Ultimately, the output of a fully developed device will be a timeline indicating when the patient had decreased movements, when the patient had fluid movements and when the patient had particular types of abnormal movements. This timeline would be used by the managing clinician to adjust the medications of the patient.

2. Background on Parkinson's disease

Parkinson's disease affects at least 750,000 Americans with the highest prevalence in older age groups⁴⁸. Parkinsonism is defined as two of the following: tremor at rest, rigidity, slow and decreased movements, flexed posture, loss of proper postural stabilizing reflexes and episodes of sudden inability to move with at least one being tremor or slow movement³¹. To be accurate, not all Parkinsonism is Parkinson's disease per se. To have Parkinson's disease, a patient should have a response to medication and not have Parkinsonism due to a known cause or as part of

a more complicated syndrome. Still, Parkinson's disease is the most common cause of Parkinsonism.

2.1 Role of medication

Most medications for Parkinson's disease work by compensating for the progressive loss of dopamine secreting neurons that is the cause for this disease. These include levodopa a precursor of dopamine, dopamine agonists, medications that inhibit the breakdown of dopamine and anticholinergics. Anticholinergics are thought to work by rebalancing the dopamine/acetylcholine balance in the brain that was offset by the loss of dopamine neurons. They are not a first line medication, but are used mostly in treating tremor. Levodopa is the standard medication and typically the most effective. However, many abnormal movements have been attributed to levodopa therapy and there is some thought that it may be toxic to the dopamine secreting neurons and may therefore make the disease worse. For this reason, some patients are started on other Parkinson's medications first. Still, in the end, patients typically end up on levodopa, usually in addition to other medications.

In addition to medication, there are neurosurgical procedures that can be used to treat Parkinson's disease. These include selectively ablative procedures (e.g. pallidotomy, thaladotomy), electrical stimulator implantation and cell transplants. These procedures have their specific indications, but are generally performed in more advanced disease and typically do not eliminate the need for medication.

2.2 Terminology

When a patient has decreased movement, he/she is said to be in the "off" state. Normal or increased movement is said to be "on". Oftentimes patients gradually turn "off" as the effect of medication wears off ("wearing off phenomenon"); sometimes the switch from "on" to "off" may be more abrupt ("on-off phenomenon" sometimes called "yo-yoing") and may rapidly switch back and forth from "on" to "off". This is usually a sign of advanced disease. Collectively, wearing off and yo-yoing are termed "response fluctuations".

Decreased movement in the "off" state can be in the form of slow movement ("bradykinesia"), paucity of movement ("hypokinesia") or difficulty initiating movement ("akinesia"). "Freezing", a sudden inability to move, may occur in the "off" state or the "on" state, with

different clinical ramifications ("off" state freezing may improve with more medication, but "on" state freezing is more problematic).

"Dyskinesia" is a general term for abnormal movements (other than tremor). There are many subtypes of dyskinesia. The two most relevant ones for our purposes are chorea (abnormal arrhythmic jerky movements) and dystonia (abnormal posturing). These are typically felt to be side effects of medication.

2.3 The prevalence and incidence of the main symptoms I wish to measure (dyskinesia and response fluctuations)

Both dyskinesia and response fluctuation are quite frequent in Parkinson's patients. Schrag and Quinn³³, in a community based study found that of the 70% of Parkinson's disease patients treated with levodopa, 40% had response fluctuations and 28% had dyskinesias. According to the DATATOP study, after only 18 months of treatment with levodopa (relatively early in disease course) 51% developed wearing off, 5% the more severe "on-off" fluctuations and 26% developed dyskinesias¹². These problems become more severe with the duration of disease. Marsden et al.²⁴ found response fluctuations to have an incidence of about 10% per year, whereas the more recent CR FIRST study⁵ found only about 20% after 5 years. In any event, symptoms tend to progress with increasing duration of disease and therefore the trend to increasing longevity will likely increase the prevalence of dyskinesias and response fluctuations even beyond where they stand today.

2.4 Rating scores and diaries

Parkinson's research relies on the ability investigators to compare the Parkinsonism (or dyskinesia) of one patient with that of another (or to compare the same patient at two different time points). For this reason, there is much literature on ways to quantify (i.e. score) patients' degree of Parkinsonism (or dyskinesia). The methods that are most commonly used to quantify a Parkinson's patient's state are rating scales and self report diaries. Device-based monitoring, both in laboratory and ambulatory, has also been used. I will review the literature on device-based monitoring later when I discuss previous work related to this project. Here, I will review rating scales and patient diaries.

The discussion of rating score and diaries is relevant to this project for several reasons. Firstly, rating scales and diaries are methodologies that compete with the type of device discussed here. A researcher (or

clinician) may decide to either use one of these older techniques or to use this device depending on the merits of the different methodologies. More importantly, rating scales and diaries are also the standards (albeit not “gold standards”) against which this project’s device would need to be validated. Lastly, rating scales and diaries are necessary in order to create the classifier for the wearable Parkinson's device. They need to be used to quantify the state of Parkinsonism, dyskinesia etc. so that the classification model can be properly trained and tested.

The most important reason for reviewing rating scales and patient self-reporting diaries is their use in creating the classifier. To build the classifier, I needed to obtain accelerometric recordings of Parkinson's patients and corresponding clinical annotations as to their state of movement (e.g. degree of dyskinesia, bradykinesia / hypokinesia). This was used to train the classification algorithm. How precisely this annotation is best is the question. Rating scales and patient self-reporting diaries could be used for this annotation. A good understanding of the various methods of annotation is necessary before deciding which ones should be used in collecting data.

2.4.1 Rating scales

As opposed to patient self reporting diaries, rating scales are typically used by professionals who are observing the patient and not by the patient himself/herself. They are intended to provide an "objective" measurement in contrast to self-reporting diaries.

There are rating scales for the staging of Parkinson’s as well as for the momentary level of Parkinsonism or dyskinesia. Since the staging of Parkinson’s is only relevant to this project as a means to stratify patients, I will focus on the momentary scales.

2.4.1.1 Scales for dyskinesia

“Continuum” scales

The simplest rating scale is a “continuum” scale. These scales typically treat tremor and bradykinesia (which occur in the "off" state) as the polar opposite of dyskinesia (which tends to occur in the "on" state). Such a scale may give a negative integer for an “off” state, zero for “on” with no dyskinesia and a positive value for "on" with dyskinesia. This is appealing since the whole rating of the patient is encapsulated in a single number. In addition, the rating is simple to do and can be repeated as

often as one needs (since it is an instantaneous measure). There are however several drawbacks. Firstly, dyskinesia and "off" are not truly opposites. Some dyskinesia can occur in the "off" state. Furthermore, different parts of the body may be in different states at the same time. For these reasons, such rating scales not commonly used in clinical investigations. Still, it should be noted that patient self-reporting diaries (which are commonly used) typically do rate "off" – "on"/dyskinetic as a single dimension, similar to these continuum scales.

It would probably not be advisable to use a continuum scale as the basis of how the clinician/observer scores how the subjects in the project are doing. Since such scales are not validated and are not commonly used, a device that can only produce output in terms of a continuum scale would likely have difficulty being approved and being accepted by clinicians. This stands in contrast to diaries that are completed by the patient (not the clinician). These are typically done on a continuum scale, but are generally accepted.

AIMS scale²

The AIMS (Abnormal Involuntary Movements Scale) scale for assessing dyskinesia is often used in clinical studies. It was developed originally for assessment of tardive dyskinesia (not Parkinsonism). Therefore it has a strong emphasis on oral and facial dyskinesias which are common in tardive dyskinesia, but are not common in Parkinsonism. For this reason, Parkinson's investigators will often modify the AIMS by leaving out the oral/facial parts of the scale. The scale uses 0-4 ratings which are simple and can be repeated every 15 minutes or so. However, there are no clear descriptions (anchors) that would tell an observer what each number on the 0-4 scale signifies. AIMS includes assessment of the trunk the arms and the legs and includes both observer and patient ratings of severity. It has not been psychometrically tested for Parkinson's patients.

The modified AIMS scale would be a useful method to clinically annotate for this project. It is commonly used and accepted and using it would likely make the device easier to be approved and more likely to be accepted. The modified AIMS scale does have somewhat more detail than is currently used for clinical purposes (e.g. separate subscales for upper extremities, lower extremities and trunk). Still, this added information may potentially be useful to clinicians.

Obeso scale²

Obeso scale is a scale for the assessment of patients prior to surgery for Parkinson's disease. It ranks intensity of dyskinesia on a 0-5 scale based on how much it impairs the patient. It also ranks duration of dyskinesia (i.e. present for what percentage of waking hours). It is not designed for continual tracking of symptoms and therefore is not useful for this project.

Rush (Goetz) scale²

Goetz modified the Obeso scale to create the Rush Dyskinesia Scale (also known as the Goetz Scale). This scale requires the clinician to ask the patient to do certain structured tasks. This would make this scale not very useful for the purposes of this project because I am attempting to monitor patients at home in their natural environment where they would not be asked to perform such tasks. Therefore, the scale could not be used unmodified to train or validate an ambulatory monitor. To the scale's benefit, it does differentiate the type of dyskinesia (chorea versus dystonia versus other), but does not isolate the anatomic distribution of the dyskinesia (e.g. arms, legs, trunk).

UPDRS²⁷ Part IV

UPDRS (Unified Parkinson's Disease Rating Scale) dyskinesia sub-score is another scale for dyskinesia. UPDRS is a general and commonly used assessment scale for Parkinson's. Part IV deals specifically with "complications of therapy" including questions specifically about dyskinesia. These are historical questions posed to the patient about how bad their dyskinesias are, when they have them, and what percent of time they have them. This subscale is not suitable for continual tracking of symptoms and therefore not useful for this project.

Dyskinesia Subjective Rating Scale

The Lang and Fahn scale (The Dyskinesia Subjective Rating Scale)²⁹ is another scale used in measuring dyskinesia. It is not suitable as an instantaneous measure however because it is based on history taken from the patient about how the dyskinesia affects various activities.

2.4.1.2 Scales for tremor

Measurement of tremor is not a major goal of this project since it is not very relevant clinically. It is not often that tremor needs treatment independent of how "off" a patient is. Tremor measurement, however, is useful as an outcome measurement especially for neuro-surgical

procedures (especially thaladotomy, but also deep brain stimulation and pallidotomy). It can also be an outcome measurement for drug trials. Such outcome measures are concerned with the general condition of the patient over long periods of time. Instantaneous measurements of tremor in contrast do not have as much utility and have not been used much. Scales that have been used include parts of UPDRS (which rates 0-4 for different parts of body and resting versus action/postural), the Webster tremor scale (which ranks tremor by amplitude)⁴⁷ or the Fahn scale (0-4 based on severity, not on localization). For our project, a rating scale that specifies location of tremor, like the UPDRS part 3 would be the most useful. It also appears to be commonly used in research.

2.4.1.3 Scaling hypokinesia / bradykinesia

The UPDRS motor score (based on a subsection of UPDRS) has been used as a means to score bradykinesia / hypokinesia. However, van Hilten et al³⁹ found poor correlation between the UPDRS motor score and activity counts (suprathreshold accelerations) implying that hypokinesia is poorly represented by the UPDRS scoring scheme. A single question from the UPDRS examination entitled "body bradykinesia / hypokinesia" has also been used³⁹. It does not require an active examination (as does much of the rest of the UPDRS motor exam). Therefore, it would be suitable for the continual, unobtrusive scoring that is necessary for this project.

2.4.1.4 Scales for Parkinsonism

These more general assessments of Parkinsonism could be of value to this project simply as a means to demonstrate how severely the patients were affected and whether the experience with them generalizes to other patient populations. These include:

- The UPDRS scale. This is a general assessment of Parkinsonism, having sections covering not only dyskinesia, but also bradykinesia, akinesia and tremor among others. Many studies use subsections of the UPDRS when trying to assess a particular one of the symptoms of Parkinsonism, however this does raise questions of validity of the subparts. The main problem with the UPDRS is that some aspects of it are momentary and some are historical. The same patient can have different UPDRS scores at different times. This limits how useful it would be as a staging of Parkinson's disease.

- Hoehn and Yahr staging. This is a 1-5 staging of the severity of Parkinsonism based on level of disability. The expanded UPDRS score (with all 6 sections) actually includes the Hoehn and Yahr staging scale. The Hoehn and Yahr scale is easy to do and does not mix up momentary and historical aspects. The drawback is that it does a very crude staging.

2.5 Previous work on device-based monitoring

Devices used to aid in the assessment of Parkinsonism may be "in laboratory only" devices or ambulatory/wearable devices. By far devices that are used are predominately "in laboratory" devices that monitor the patient for brief periods of time. However, it is the ambulatory/wearable devices are most relevant to this project.

2.5.1 Methods of collecting sensor data

Since the goal of this project is to investigate a wearable monitor for detecting movement in Parkinson's disease, it is relevant to review the literature to find what types of sensors to use and where to place them on the body in order to produce the best results.

2.5.1.1 Different types of sensors, their pros and cons

Many different types of sensors have been used to obtain movement data from patients. These include electromyography^{37,32}, ultrasound, radar, laser displacement detectors^{28,26,15,30}, mechanical coupling devices²⁵, video-based systems, in addition to accelerometers and rotation sensors⁷. For the purposes of ambulatory/wearable monitoring, only accelerometric and gyroscopic modalities and perhaps electromyography appear to be feasible. Video monitoring also may be feasible³⁸ so long as the involved limbs can be maintained in line of sight (which may be an issue).

a) Displacement and velocity measurements

The standard method for measuring movement in an ambulatory (outside of clinic) setting is accelerometry. There were some questions raised as to whether measuring velocity or displacement directly may have some benefit over measuring acceleration²⁸. Velocity and displacement sensors could be more sensitive to lower frequencies whereas acceleration measurements would be more sensitive to higher frequencies. Other than mention of this as a theoretical consideration, there has been no

demonstration that measuring acceleration alone misses types of movements that would be relevant to our study. In any case, a system used to measure velocity and displacement might not be suitable to be used as an ambulatory device. Video monitoring, though, does measure displacement and has been used as a wearable ambulatory device³⁸.

b) Electromyography³⁶

Electromyography (EMG) measures the electrical activity of muscle contraction. Certain information can be obtained using EMG that could not be obtained using accelerometry. For example, normally, when one muscle contracts, the opposing muscle relaxes. In dystonia, however, opposing muscles can contract simultaneously. This could, in theory, aid in detecting dystonia. Furthermore, since particular muscles are isolated, it is easier to detect actual activity of the muscle as opposed to accelerometry where the acceleration detected may be due to movement of another body part, movement of the whole body or gravitational pull. The main drawback of EMG as opposed to accelerometry is that it is harder to apply and wear on an ambulatory basis. Muscles need to be found and electrodes placed, requiring some expertise. In addition, maintaining good electrical contact over the course of a day is not as simple as maintaining placement of an accelerometer (e.g. the skin actually needs to be abraded). These problems limit the use of EMG on a large scale. Additionally, EMG only records activity of the muscles it is applied to, limiting the information obtained. This may or may not impair the ability of an EMG based monitor to determine the status of a Parkinson's patient. If the problem with easily applying EMG can be overcome, it is possible that EMG may yield data useful in classifying the state of movement of Parkinson's patients (either alone or in conjunction with accelerometry).

c) Video monitoring

Video monitoring coupled with machine vision has been used in detecting a variety of different movements^{38,51}. Georgia Tech has a group³⁸ that is investigating the use of a wearable pendant that detects hand motions by using computer vision algorithms and they have discussed its potential use in detecting Parkinsonian or essential tremors. They, however, have not published work where this was actually done. In any case, the patient appeared to need to bring his/her hands up towards the pendant in order for the pendant to capture the hand movements. Clearly, this would not work for our goals, since we are attempting to capture data without active participation of the patient. The line of sight issues may be resolved if

several video capture sensors were used, though it is unclear what specific advantages such a system would have over accelerometers.

d) Rotation-sensitive devices

Burkhard et al⁷ described a rotation-sensitive device for measurement of dyskinesia. This device seems to detect some useful characteristics of movement, since empirically dyskinesias appear to be more rotational than voluntary movements. However, not much can be said about how useful such a device would be for our purposes. The device was only used for brief periods of time in the laboratory (not ambulatory) and there was no attempt to try and differentiate dyskinesia from normal voluntary activities.

e) Simplified acceleration-based devices

The device used for this project uses a full 3-axis accelerometer. Various simplified versions of accelerometers have been studied. For example, activity counts have been used extensively both in Parkinson's studies as well as in addressing other medical problems⁴⁵. An activity monitor is essentially an accelerometer, but most of the information is discarded. All that is measured is the number of times acceleration rose above a fixed threshold. It typically registers movements only along a single axis. Still, activity monitors have found some usefulness. For example, Van Hilten et al⁴³ used a wrist worn activity monitor diurnal activity in Parkinson's patients and appeared to find patterns consistent with medication cycles. Nonetheless, it could not differentiate between voluntary activity, tremor and dyskinesia. It is also uncertain what precisely this device was detecting. A high level of activity counts may have been due to the voluntary activity of a well-medicated Parkinson's patient or due to the abnormal dyskinetic movements more typical of an overmedicated patient. This is because this study, like most studies of ambulatory monitors in Parkinson's, did not record continual clinical observations to correspond with the device recordings. Single axis and 2-axis accelerometers have also been used. Dunnewald, Jacobi and van Hilten asserted¹⁰ that 3 axis accelerometers "provided no relevant additional information" over 2 axis accelerometers. However, this likely does not apply to our project for many reasons. For instance, much of the information in their data was discarded (e.g. they dichotomized using an arbitrary cutoff threshold). If all the data were used, 3-axis accelerometers may have an advantage. In addition, the data they present do not necessarily support their contention that there is no advantage. Three axes outperformed 2 axes 3 out of 6 times which is more than

expected by chance though not with statistical significance (Since there are three 2-axis pairs and only one 3-axis pair, the expected probability of the 3-axis accelerometer performing better simply as a matter of chance would be 0.25. The expected number of occurrences would therefore be $0.25 * 6 = 1.5$. The probability of 3 or more occurrences, using the binomial distribution would be 0.169, which would not be significant if the standard cut-off of 0.05 is used.

f) Other modalities

I have not found much literature relevant to this project using other modalities such as ultrasound or RADAR. Both use Doppler shift to detect movement, though ultrasound uses mechanical vibration and RADAR uses radio frequency electromagnetic waves. Generally these are used to count movements in much the same way as accelerometers have been used to count movements (i.e. by counting all events above a certain fixed threshold). It appears that ultrasound was primarily used for detecting facial dyskinesia, for example, a detector to measure oro-facial dyskinesia²⁶, an apparatus to measure facial movement¹⁵ and a technique to measure tardive dyskinesia³⁰. Facial dyskinesias are not common in Parkinson's rather they are seen frequently in tardive dyskinesia (a side effect of antipsychotic medication). Therefore, ultrasound as it was used is less relevant to this project. RADAR may have some utility. Of course, one must consider that a device with electromagnetic emission might have more hurdles to overcome before being approved for clinical use. Mechanical coupling devices need the subject to physically move one part of the device relative to another part of the device in order to generate a signal. Generally, these devices are cumbersome and are therefore not useful for ambulatory purposes.

2.5.1.2 Sensor placement

Since the primary goal of this project is to use accelerometer readings to detect dyskinesias as well as level of "on" or "off", it is important to consider where sensors will need to be placed to detect these two parameters.

2.5.1.2.1 Placement to detect dyskinesias

In order to consider where to place sensors to detect dyskinesia, it is important to consider where dyskinesias occur. Dyskinetic episodes are felt to commonly involve the legs first (and sometimes only the legs). In fact, Fahn¹² discusses how diphasic dyskinesias, which have a different

clinical implication than standard peak dose dyskinesias, tend to occur most often in the legs. Truncal dyskinesia also may occur in the absence of appendicular dyskinesia. Therefore it may be prudent to place sensors on legs as well as on the trunk. Other sensors may be needed to be able to differentiate dyskinesia from voluntary activity. Beyond this, it is difficult to guess where to place sensors to detect dyskinesia in the absence of experimental data. In the pilot study (discussed in the section “Previous work”), I found that detecting dyskinesia actually required fewer sensors than detecting “on” and “off”. A good model could be built using 2 sensors (right upper extremity and right hip for pilot study patient #1 and right upper extremity and right lower extremity for pilot study patient #2). The hip accelerometer appeared to be important in both the neural network and classification tree models despite the fact that the hip accelerometer malfunctioned on patient #2 and yielded no data for a large part of the recording. Since the hip readings were so important, it suggests that having a truncal sensor would be useful (e.g. hip or other part of trunk).

The work in the literature regarding sensor placement cannot be well generalized to this project. It is difficult to state that a certain sensor site is not useful for detecting dyskinesia or “on”-“off” if the study did not try to detect these states the way this project attempts. Still, there is some relevant literature.

Van Hilten et al⁴⁴ attempted to address the problem of where to place accelerometers. They found “no difference” between placing an accelerometer on the dominant wrist, the non-dominant wrist or the hip. However, the limitation of their processing brings into question the generalizability of this statement. There was no attempt to combine information from several sensors (e.g. correlation). Therefore, they could not address which sites should be used if more than one site is to be used. Additionally, much of the information in the data was processed out. They distilled all the rich accelerometric data down to a simple determination of whether a movement “occurred” or not. For this they used a fixed threshold of 0.1g using only the 0.25-3Hz band (g is the acceleration of gravity at sea level). They then counted the number of movements (a measure of bradykinesia) as well as the number of 15-second epochs not containing a movement (a measure of hypokinesia). The conclusion that they reached using the “stripped down” version of the data likely does not say much about what would be if all the information in the data was used.

Manson et al²³ used a single 3-axis accelerometer attached to the shoulder of the most affected side as an ambulatory dyskinesia monitor. Their rationale was that at that site, the sensor could detect both limb and truncal dyskinesia in a single sensor. They used the correlation between mean acceleration (in the 1-3Hz band) and clinical dyskinesia scores as a measure of how well the device could detect dyskinesia. They found that the device detected dyskinesia well when tasks such as eating, drinking were done but found poor correlation when the patient was walking. This was sensible since walking involves significant movement of the shoulder which would make differentiating voluntary activity from dyskinesia hard. In fact, when the patients did tasks that would be expected to use their shoulders more (e.g. "putting on coat") the device did less well in sorting out presence of dyskinesia from its absence.

It is unclear from that study whether using only a single shoulder site would be adequate for our purposes. Still, it seems sensible to detect both arm and trunk movement. The shoulder would seem to be a difficult site to place a sensor at. Detecting arm and trunk movement may require two separate sensors.

2.5.2 Detecting posture

Several authors have used wearable devices to detect a patient's posture. By posture, I am referring to lying versus sitting versus standing. There does not appear to be much work on wearable devices detecting more subtle aspects of posture, for example, if the patient has a stooped posture as Parkinsonian patients often do. Simple mercury switches have been used to determine posture. For example, Walker et al⁴⁵ used mercury switches on the thigh and chest to determine posture. Other articles have used accelerometers on the thigh and chest similarly. By using the "DC" (e.g. 0-1Hz) component of the accelerometer signal, the accelerometer can produce a result similar to the mercury switches^{9,18}. These simple posture detection algorithms seem to work quite well. Dunnewold et al⁹ found a sensitivity of 99.6% and a specificity of 99.8% while performing a predetermined sequence of activities meant to reflect normal daily activities. Occasionally, though rarely, this technique was fooled. For instance, stair climbing occasionally appeared as sitting (because of the raised thigh) and a transfer from lying to sitting appeared as standing (because the thigh sometimes was perpendicular to floor).

Posture detection algorithms can yield useful information for this device. Having information as to whether a patient is lying sitting or standing while specific accelerometric recordings are made may help build a better

classifier. Still, it is not likely to help that much because since the vast majority of the time is spent in the seated position. Posture monitoring may also be useful for clinicians independent of its use in building a classifier. It can yield information as to the patients' activity and hence, their need for medication.

2.5.3 Detecting tremor

Parkinsonian tremor may affect all four extremities as well as the trunk and head. However, clinically, distal arm tremor is the most stereotypical of Parkinson's disease (e.g. "pill rolling tremor"). Devices to measure tremor in Parkinson's disease have focused on the distal upper extremities. Hoff Wagemans and van Hilten¹⁹ found that in structured tasks, a 3-axis accelerometer attached to the most affected wrist appeared to detect tremor reasonably well. The digits have also been used to measure tremor. As an example, Beuter and Edwards⁴ measured tremor on the index finger. A major reason that they chose the index finger is that the wrist has a resonant frequency of 8-12 Hz, which may be confused with physiological tremor whereas the resonant frequency of the index finger is about 17Hz. Since this project does not concern itself with physiological tremor, this is not likely to be important here.

2.5.4 Data assessment

2.5.4.1 The frequency domain

Different types of movements in Parkinson's patients tend to have different frequency characteristics. A review of the literature finds the following approximate frequency ranges for different types of movement. Dyskinesia has been found to be predominately in the lower frequency range (approximately 0.25Hz-3.5Hz with some variability between articles)^{23,7,9,44} and Parkinson's rest tremor at a higher frequency (4-6Hz)^{23,7}. Other types of tremor tend to be in a higher range (essential tremor 7-12Hz and physiological tremor 8-12Hz). Different types of dyskinesia were found to have different frequencies. For example, Burkhard et al.⁷ found dystonia to be in the 0.25-1.25Hz range and chorea in the 1.5-3.25Hz. One article⁶ found voluntary activity to be in the below 3.3Hz range with the majority less than 1Hz (except walking, which was about 2Hz). Unfortunately, none of these frequency ranges are "hard and fast". There is overlap in frequency range between different types of motion.

2.5.4.1.1 Characteristics of frequency spectrum

Rather than a single frequency, an accelerometer actually picks up a spectrum of frequencies. A device might be able to use the predominant frequency in order to better classify what type of movement is occurring. (e.g. voluntary activity versus dyskinesia versus tremor). However, there is more information in the frequency spectrum than simply the peak frequency or mean frequency. The distribution of the frequencies may also help to better classify the type of movement. There are examples in the literature of different ways of extracting features from the frequency spectrum that could help in classifying different types of movement:

Beuter and Edwards⁴ investigated the use of several frequency domain characteristics to differentiate Parkinsonian tremor from (normal) physiological tremor. They found that they were better able to discriminate the two types of tremor if they took into account the dispersion of frequencies and not just the overall amplitude of acceleration for a particular band. Factors that they used that take into account dispersion included the following:

- 1) dispersion about median frequency (width of band containing 68% of the power spectrum),
- 2) a measure of how close the spectrum is to a single peak (harmonic index, normalized to the highest peak),
- 3) proportion in 4-6 Hz (typical of Parkinson's tremor) and 7-12Hz (typical of physiology tremor),
- 4) the un-normalized harmonic index (a measure of how close the spectrum is to a single peak).

In addition, the article used a center of mass of the frequency, a factor that in theory could have advantages over other methods of detecting the most important frequency (e.g. peak frequency or median frequency). All these factors appeared helpful in discriminating Parkinson's patients from controls for the structured tests they performed.

Scholz et al.³² and Spieker et al.³⁷ made use of the signal to noise ratio (SNR) to analyze the frequency distribution of EMG recordings in order to detect tremor. Scholtz et al. used a fixed SNR cutoff of 4 as a threshold to determine the predominant frequency of the tremor. The SNR was also used as a measure of intensity of the tremor. Spieker et al also used a fixed cutoff for SNR to detect tremor. A period of time was labeled as tremor time if a frequency in the 3.7 – 10Hz had a SNR of 4 or larger. They found that by counting the amount of "tremor time," this technique

discriminated between tremor patients and controls with a specificity of 94% and a sensitivity of 96%. This result was somewhat misleading because there were no direct determinations that the patients indeed had tremor at those times. Therefore, no statement could be made as to how well this system discriminated periods of tremor from periods of non-tremor.

Standard Fourier transform (power or frequency spectrum) analysis could be thrown off by spurious frequencies that occur transiently. Van Someren et al.³⁵ attempted to remedy this problem by checking for a consistent waveform that exists over several periods. They labeled "half periods" as parts of the recording between zero crossings and if those "half periods" were within a certain range of period (i.e., a period that could be tremor) and if there were at least 12 consecutive "half periods", then that would be labeled as tremor. This system seemed to work well on the population it was tested on (pre and post-thaladotomy Parkinson's patients and normal controls). Only 4% of control time was labeled as tremor and tremor was found to decrease post surgically. It is unclear if this somewhat ad hoc technique would perform better than comparing frequencies of successive (finite) Fourier transforms. It does suggest that perhaps the frequency spectrum of a particular time window should be considered relative to the frequencies obtained during adjacent time windows or be compared to average frequencies obtained for the whole recording.

2.5.4.2 Methods used to process the data

The most common technique that has been used to analyze data obtained from wearable sensors has been some form of correlation^{37,13,4,23,7,35,10,25,18,11,19,39}. This may involve either comparison of the features derived from the device readings with some clinical score that was observed. It may also involve comparison between different readings without comparison to a clinical score, such as might be done to test reliability and validity. Other statistical methods used include analysis of variance^{9,42,40,44}, the kappa statistic⁴⁵ and linear regression⁴¹. Hidden Markov models have been used to detect gesture sequences³⁸, but not for purposes of detecting free form pathological movements. Neural networks appear to be the only major "machine learning" technique that has been explored^{21,22,1}.

Of the related articles using neural networks, only the work of Keijsers et al.²¹ was related to Parkinson's disease. They used neural network models to predict dyskinesia based on accelerometry. The subjects

performed specific tasks which leads to questions of how representative this would be of real world activity. Still there were a variety of tasks and so it would be closer to what actually would happen in an unstructured environment. They used 38 features derived from the accelerometry recordings as input to the networks. To measure success, they checked a Spearman correlation between the observed value for dyskinesia and the value predicted by the network. In general, they found good correlation, though some values were quite low. It did confuse voluntary activity with dyskinesia as noted by the poor correlation of predicted to measured values of dyskinesia when the patient was walking. They used a “leave-1-out” technique of cross-validation (no true validation set) and, since they adjusted parameters based on the entire set of data (including the validation sample), it may be possible that their results were overly optimistic.

2.5.5 This project’s pilot study

In order to assess the feasibility of using such a device to classify on-off range (i.e., hypokinesia and bradykinesia) and dyskinesia in Parkinson’s patients, I conducted a pilot study. In the pilot study I tested the device on two Parkinson’s patients in an observed setting and trained and tested two different models of movement classification based on the collected data.

2.5.5.1 Methods and materials

Two Parkinson’s patients (1 male and 1 female) were recruited from the Parkinson’s clinic at University Hospital of Brown University (Pawtucket, RI). Both patients were determined by their referring neurologist to have motor fluctuations. Patient #1 (male) was on Sinemet CR 3 times a day. Patient #2 was on levodopa/carbidopa 5 times a day in addition to entacapone and ropinirole. The patients were observed in the Parkinson’s Day Program room during the course of the study. For the length of the study, they were observed by a neurologist (the author) and were videotaped for later review by the same neurologist. During the study they wore the device for detecting motion. After the observation, the data was downloaded for offline analysis.

2.5.5.1.1 The device

I used a series of five 3-axis accelerometers (as developed at the M.I.T. Media Laboratory). The range of the accelerometers was from -1g to +3g with a resolution of 1/64g. Samples were taken by the accelerometers at

approximately 40 Hz. The five accelerometers were attached to the patient using Velcro straps at the following locations: One on the dorsum of each arm just proximal to the wrist and one on each leg just proximal to the lateral aspect of the ankle. One additional accelerometer was contained in the main unit which was attached to the patient's belt.

2.5.5.1.2 Scoring scheme

The observing neurologist queried the patient as to his state and later

Table 1. Scoring scheme used for pilot study

Dyskinesia (chorea only)

- | | |
|---|---|
| 0 | none |
| 1 | mild (does not appear to impair patient at all) |
| 2 | moderate (appears to cause mild impairment of activity) |
| 3 | significant (appear to cause moderate impairment of activity) |
| 4 | severe (appears to cause severe impairment of activity) |

On-Off (a measure of bradykinesia and hypokinesia)

- | | |
|---|---------------------------|
| 0 | Significantly off |
| 1 | Mildly off |
| 2 | Ambiguous or intermediate |
| 3 | Mildly on |
| 4 | Definitely on |

reviewed the video recording to obtain a synthesis assessment of the state of the patient. A 0-4 scoring was used for "on-off" (i.e. level of bradykinesia and hypokinesia) and a 0-4 scoring for dyskinesia (more specifically chorea). Table 1 shows the scoring scheme that was used.

Notation was made once per minute for the duration of the time the device was recording. If the patient was temporarily off the video or temporarily not observed (e.g. trip to bathroom), the neurologist would extrapolate the intermediate time points based on data known about the surrounding observed time points.

2.5.5.2 Results

A total of 306 minutes of data were obtained for patient #1 and 323 minutes of data obtained for patient #2. The belt accelerometer of patient #2 malfunctioned midway through the study and therefore it was not used in analysis. In addition, the left lower extremity accelerometer for patient #2 was switched midway through the study to the right proximal upper extremity (for testing purposes), so that data too are discarded from this study. Each patient experienced approximately 2 ½ on-off cycles.

2.5.5.3 Analyses

2.5.5.3.1 Dichotomization of scores

Although a scale of 0-4 was used for dyskinesia and (separately) for on-off, these scorings were then dichotomized based on a subjective analysis

Table 2. How the on-off and dyskinesia scales were dichotomized in the pilot study

Patient #1

	Low range	High range
On off	<1.5	>=1.5
Dyskinesia	<2.0	>=2.0

Patient #2

	Low range	High range
On off	<3.5	>=3.5
Dyskinesia	<2.0	>=2.0

Note: The text refers to “high level” for on-off as “on”, low level on-off as “off”. “High level” dyskinesia is referred to as “dyskinetic” and “low level” dyskinesia as “not dyskinetic”

of what would be a clinically relevant cut off. This was determined based on the range of variation of the patient. Table 2 shows the cutoffs were used for dichotomizing (using the original 0-4 scale):

2.5.5.3.2 Pre-processing

Data were processed using Java, Matlab, as well as Netlab (for neural network functions). The data were divided into training and test sets for use in a neural network (see Table 3). Data for each patient were handled

Table 3. Size of test sets and training sets for the pilot study

Patient	Size of Test Set	Size of Training Set	Total
Patient#1	124 one minute samples	186 one minute samples	310
Patient#2	132 one minute samples	198 one minute samples	330
Total	256 one minute samples	384 one minute samples	640

separately. Data were divided into one-minute windows and each window was assigned to the training set or test set randomly in a roughly 60:40 ratio. Therefore the test and training sets did not consist of contiguous time periods.

2.5.5.3.3 Feature extraction

From the accelerometric data, features were derived and used as the basis for neural network and classification tree classification models. Table 4 lists features that were used.

2.5.5.3.4 Time window

As mentioned previously, accelerometric recordings were obtained at a

Table 4. Features that were extracted from the accelerometry data and then used as inputs into machine learning algorithms (pilot study).

Name	Description
R1	Hip "magnitude"
R2	RUE "magnitude"
R3	LUE "magnitude"
R4	RLE "magnitude"
R5	LLE "magnitude"
R6	RUE and LUE "positional correlation"
R7	RUE and LUE "magnitude correlation"
R8*	RUE and LUE "positional mutual information"
R9	Not used
R10	RUE and RLE "positional correlation"
R11	hip and RUE "magnitude correlation"
R12	hip and LUE "magnitude correlation"

RUE = right upper extremity

LUE = left upper extremity

RLE = right lower extremity

LLE = left lower extremity

For description of "magnitude", "positional correlation", "magnitude correlation" and "positional mutual information" please see text.

* This feature was found not help much in classification and because it was very computationally intensive, it was not used in building any of the models

rate of about 40 readings per second. To enter this data into a machine learning program, there are two possibilities: One way would be to use to data from each individual reading (representing 1/40 of a second) as the input for the machine learning algorithms and the label (e.g. "on-off" and dyskinesia state) for the output. Another approach would be to window the processing in a way that features derived from a whole period of time (e.g. 1 minute) would be used instead of data from only a single reading cycle. I have chosen the windowing approach for several reasons. Firstly, if only a single reading were used as the basis for the model, then the

classification power of the system would have been very weak. The accelerations at a particular point in time are not likely to be nearly as good a predictor of movement state as those of an entire period of time. This problem could be partially remedied by letting the machine learning program make a prediction based on only a single reading cycle, but then combine these predictions to create a prediction for a whole time window. In that way the prediction for the whole time window would be more powerful because it combines the power of the many individual predictions that were made for each reading cycle. However, how best to combine these predictions into one larger prediction is not clear. For instance, they could be averaged or multiplied depending on different assumptions. The windowing technique used here of making the predictions based on the whole window, simplifies this problem. Another reason why using the whole window may be better than using only a single reading cycle is that it enables correlations or mutual information between different accelerometers to be generated. While it is true that the machine learning algorithm may “learn” how different features vary together even if only one time point at a time is taken, that would not take into account the range of values in the time window immediately preceding and following that time point. Time window correlation and mutual information measures adjust for near term variability.

2.5.5.3.5 Description of features

The features that were used as inputs for the machine learning algorithms are listed in table 4. Below is a description of some of the terms that were used.

Gross acceleration

The gross (measured) acceleration (used for all later processing discussed below) is obtained by simple vector addition (i.e. gross acceleration = $\sqrt{(x^2+y^2+z^2)}$)

“Magnitude”

The term “magnitude” does not refer to the magnitude of acceleration measured, rather to the mean of the absolute value of the derivative of acceleration with respect to time (i.e. $|da/dt|$ where a is the magnitude of acceleration and t is time). The reason this measure was chosen is because it was felt to more relevant and less susceptible to drift. The derivative of measured acceleration was used because the accelerometer readings tend to drift some over the course of the study. The derivative negates this drift because derivatives would be little changed by slow drifts over the course of hours. The absolute value was used because,

clearly averaging the derivative over anything but the shortest period of time would yield zero. Accelerations and decelerations are both measures of movement and here I count them equally.

“Positional correlation”

This was intended as a measure of common orientation of the limbs involved. Certain activities may be expected to entail different limb orientations. If two limbs have their positional orientations correlated, then it might be expected that they are working together. The way this measure was calculated was as follows: Six factors (the X, Y, and Z axis accelerations of both sensor sites) were correlated with each other in all possible permutations (except that a factor was not correlated with itself). The mean of these 15 correlations was called the “positional correlation”.

“Magnitude correlation”

This is actually the correlation of the derivative of measured acceleration over the time window involved (1 minute).

“Positional mutual information”

As discussed in the section about “positional correlation”, certain repeated positions might signify certain activities (e.g. walking). However, it might be expected that some position of the two sensors are common in a particular activity, but may not be detectable by simple correlation. For this reason, positional mutual information was used. Positional mutual information was calculated in a manner similar to “positional correlation” (described previously), however, instead of correlations, mutual information was used as it in theory might be more appropriate than simple correlation. The process to calculate the mutual information of 2 sensors (for a particular time window) is described in Table 5.

Table 5. Method for calculating mutual information using accelerometry data from two separate accelerometers (pilot study)

1. Take the accelerometry data of the 2 sensors in question for the time window in question. Call those 2 strings of data vector X and vector Y .
2. Discretize the vectors X and Y as follows: replace the values for acceleration with a number indicating what quartile (or decile) that value belongs to relative to the other values for acceleration in that same vector (in this project both quartiles and deciles were tested). Call these new discretized vectors X_d and Y_d . (Quartiles and deciles are labeled starting from zero)
3. Create a vector W by combining X_d and Y_d as follows: $W(t)$, the element of the vector W that corresponds to a particular timepoint t , is set to $X_d(t) * (\# \text{ of possible values of } Y_d) + Y_d(t)$.
4. Mutual information is calculated using Shannon's entropy: $MI = \text{entropy}(X_d) + \text{entropy}(Y_d) - \text{entropy}(W)$

10-minute moving average of dyskinesia

Rather than minute-by-minute dyskinesia as a target output, a 10-minute moving average was used. This produced better results on the training set, presumably because dyskinesia varies a lot over the very short term and may be missed using smaller windows.

2.5.5.3.6 Neural networks

To implement the neural network part of the experiment, I used Netlab (an extension of Matlab). Coding was done in Matlab, R and Java. The implemented neural network used a single hidden layer of neurons. Hidden nodes used a tanh activation function and the (single) output neuron used a logistic function.

The feature space (and neural network parameters) was explored using 5-fold cross-validation on the training set. Features for the test set were chosen based on results of the cross-validation on the training set. Table 6 shows the features that were selected.

Table 6. Features used for neural network models in the pilot study

Patient #1 “on-off” (Model #1)

Inputs:

Hip absolute of derivative of magnitude for the window (R1)*
RUE absolute of derivative of magnitude for the window (R2)
RLE absolute of derivative of magnitude for the window (R4)
LLE absolute of derivative of magnitude for the window (R5)
RUE/RLE positional correlation for the window (R10)

Output:

The average on-off rating for the 1 minute window

Neural net:

6 hidden nodes and 100 iterations

Patient #2 “on-off” (model #2)

Inputs (same as patient #1):

Hip absolute of derivative of magnitude for the window (R1)+
RUE absolute of derivative of magnitude for the window (R2)
RLE absolute of derivative of magnitude for the window (R4)
LLE absolute of derivative of magnitude for the window (R5)
RUE/RLE positional correlation for the window (R10)

Output (same as patient #1):

The average on-off rating for the 1 minute window

Neural net (same as patient #1):

6 hidden nodes and 100 iterations

Patient #1 dyskinesia (model #3)

Inputs:

Hip absolute of derivative of magnitude for the window (R1)
RUE absolute of derivative of magnitude for the window (R2)
Hip and RUE magnitude correlation for window (R11)

Output:

A ten minute moving average of dyskinesia

Neural net:

Hidden nodes 6 iterations 100

Patient #2 dyskinesia (model #4)

Inputs:

RUE absolute of derivative of magnitude for the window (R2)
RLE absolute of derivative of magnitude for the window (R4)

Output:

A ten minute moving average of dyskinesia

Neural net:

Hidden nodes 4, iterations 200

* Note: names of features in parentheses refer to the features listed in Table 4
 + Hip accelerometer yielded missing data for section of recording. For this part, this feature was assigned a value of zero

In order to assess the calibration of the neural network classification model, the Hosmer-Lemeshow c-hat and h-hat goodness-of-fit statistics were obtained. For the Hosmer-Lemeshow c-hat, the samples were divided into quartiles as in table 7:

Table 7. Ranges used by Hosmer-Lemeshow c-hat (pilot study)

Range #	Expected value
1	<25 percentile
2	>=25 and < 50 percentile
3	>= 50 and < 75 percentile
4	>= 75 percentile

For the Hosmer-Lemeshow h-hat, the samples were divided into 4 ranges as in table 8:

Table 8. Ranges used by Hosmer-Lemeshow h-hat (pilot study)

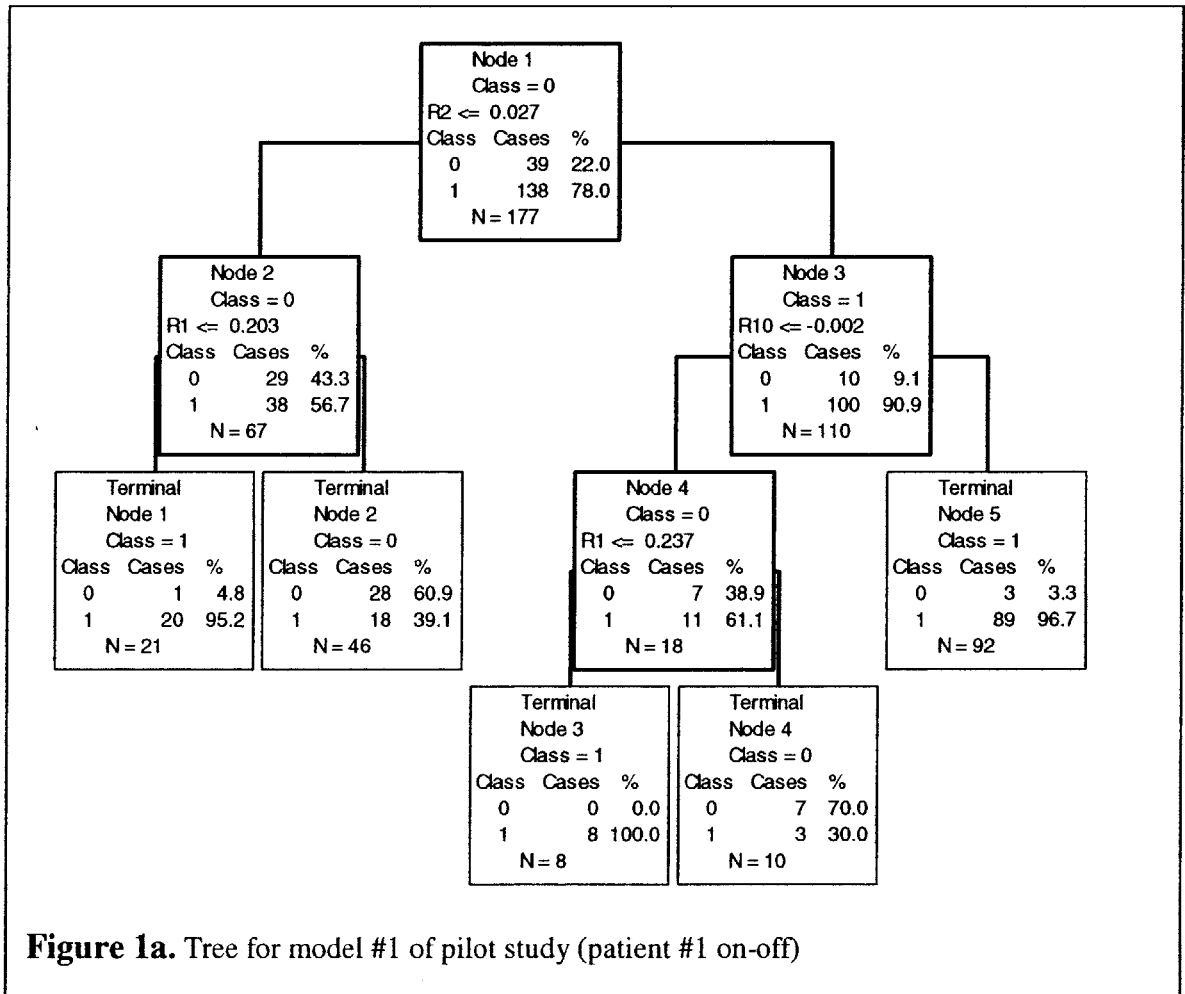
Range #	Expected value
1	<0.25
2	>=0.25 and <0.5
3	>=0.5 and <0.75
4	>0.75

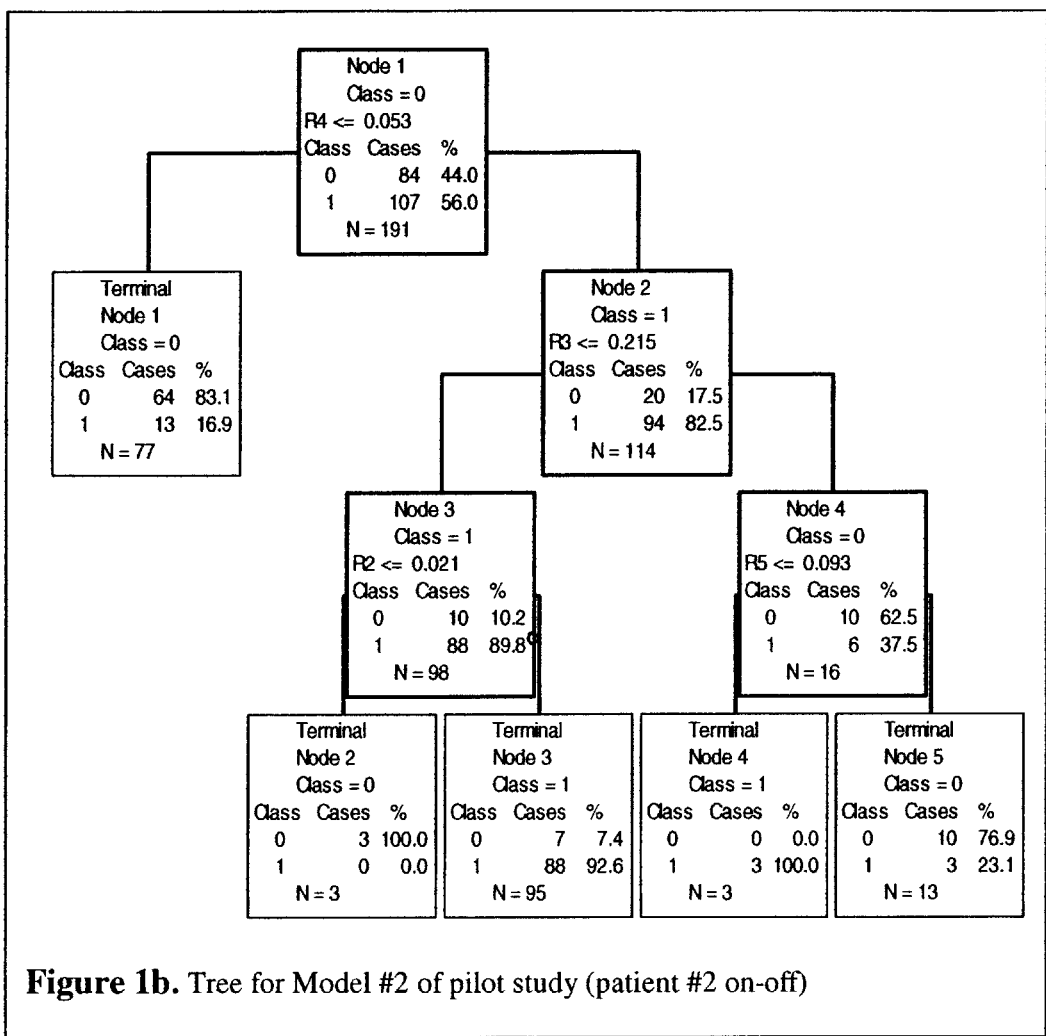
The results of the Hosmer-Lemeshow tests for each of the neural network models are shown in Table 9.

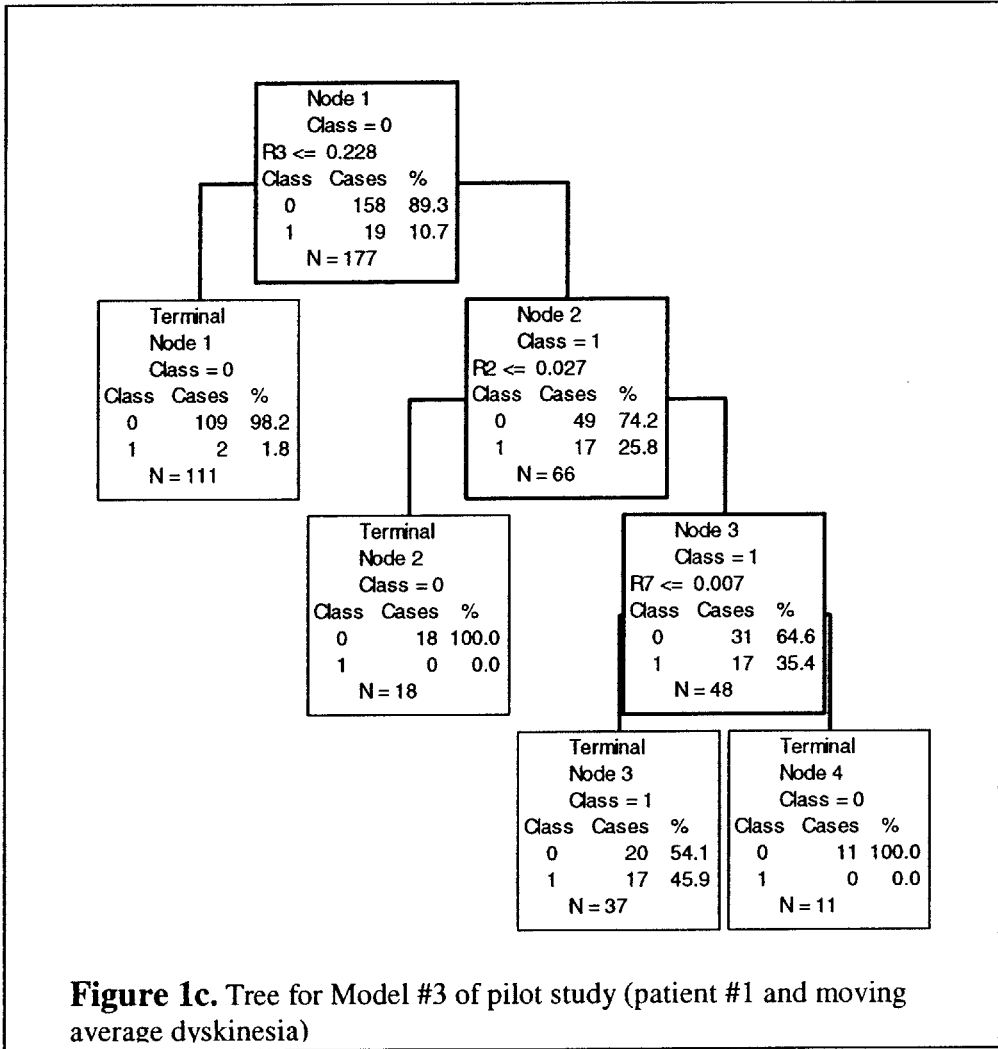
Table 9. Results (pilot study) of Hosmer-Lemeshow test for neural network models (using the test set only)		
Hosmer-Lemeshow c-hat		
Model	p-value	Degrees of freedom
Patient #1 on-off	0.8154	7
Patient #2 on-off	0.07559	7
Patient #1 dyskinesia	0.2438	7
Patient #2 dyskinesia	0.593	7
Hosmer-Lemeshow h-hat		
Model	p-value	Degrees of freedom
Patient #1 on-off	Not calculable	N/A
Patient #2 on-off	0.9864	3
Patient #1 dyskinesia	0.468	5
Patient #2 dyskinesia	0.7504	7

2.5.5.3.7 Classification trees

The classification tree section of this experiment was implemented using CART (Classification and Regression Trees) 4.0 from Salford Systems (San Diego, California). The same data (i.e. the same training and test sets and same dichotomization) that were used for the neural network part of the experiment were also used for the classification tree part of the experiment. All the features that were used during feature exploration for the neural networks were also used for constructing the classification tree models with the exception of positional mutual information, because it was computationally expensive and was not found useful in the neural network models. All defaults of CART were used for the building of the model (e.g. gini statistic used, 10-fold cross-validation for model building using training set). Figures 1a, 1b, 1c, 1d show the classification trees that were obtained. Tables 10a, 10b, 10c, 10d show the success of the trees in classifying the test set.







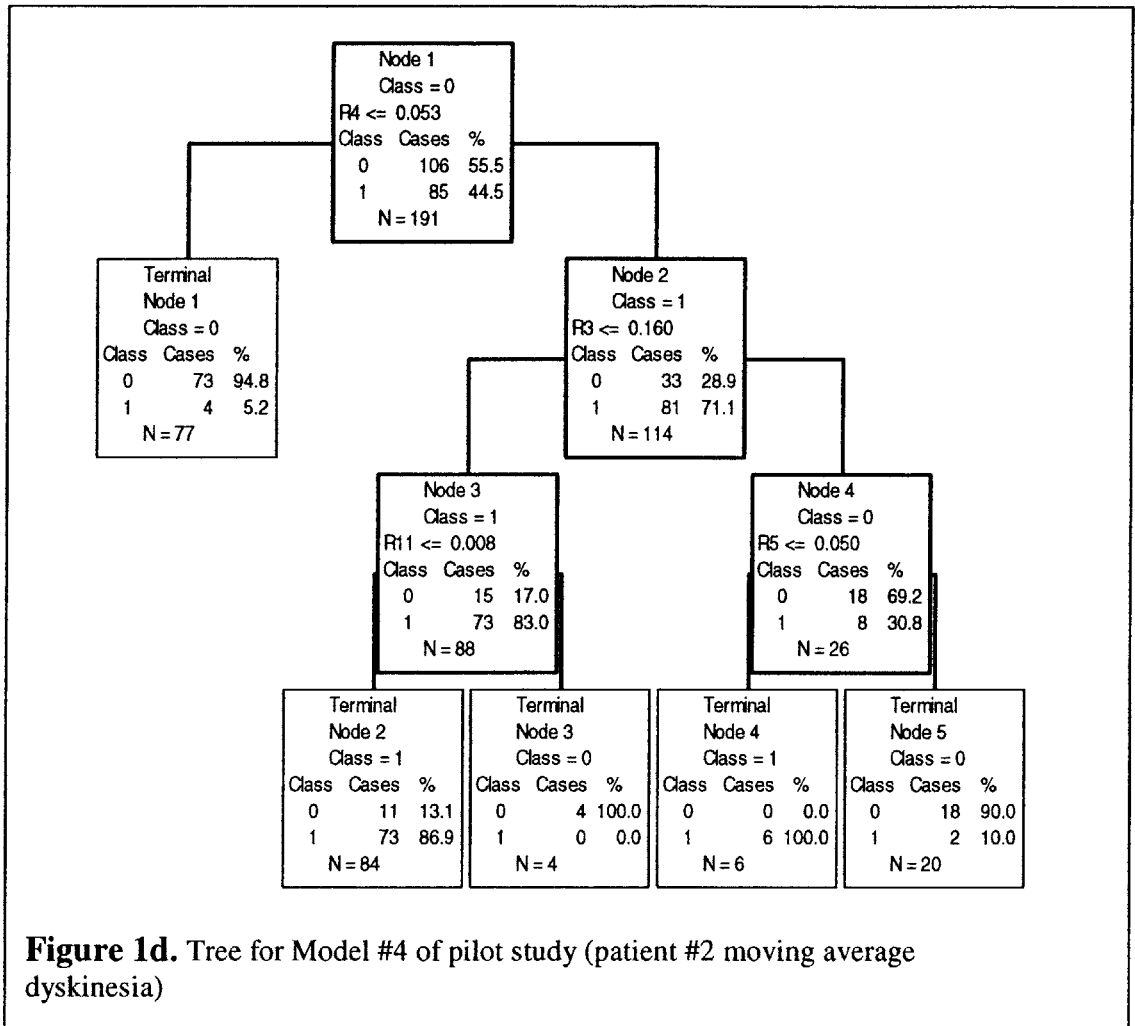


Table 10a. Confusion matrix and accuracy for results of CART using model #1 of pilot study (patient #1 on-off)

Training set:

	predicted "off"	predicted "on"
actual "off"	35	4
actual "on"	21	117

Test set:

	predicted "off"	predicted "on"
actual "off"	19	7
Actual "on"	23	70

Test set accuracy = 0.748

Table 10b. Confusion matrix and accuracy for results of CART using model #2 of pilot study (patient #2 on-off)

Training set:

	predicted "off"	predicted "on"
actual "off"	77	7
actual "on"	16	91

Test set:

	predicted "off"	predicted "on"
actual "off"	42	4
actual "on"	15	68

Test set accuracy = 0.853

Table 10c. Confusion matrix and accuracy for results of CART using Model #3 of pilot study (patient #1 moving average dyskinesia)

Training set:

	predicted "not dyskinesic"	predicted "dyskinesic"
actual "not dyskinesic"	138	20
actual "dyskinesic"	2	17

Test set:

	predicted "not dyskinesic"	predicted "dyskinesic"
actual "not dyskinesic"	94	7
actual "dyskinesic"	3	15

Test set accuracy = 0.916

Table 10d. Confusion matrix and accuracy for results of CART using Model #4 of pilot study (patient #2 moving average dyskinesia)

Training set:

	predicted "not dyskinesic"	predicted "dyskinesic"
actual "not dyskinesic"	95	11
actual "dyskinesic"	6	79

Test set:

	Predicted "not dyskinesic"	Predicted "dyskinesic"
Actual "not dyskinesic"	54	15
Actual "dyskinesic"	10	50

Test set accuracy = 0.806

2.5.5.4 Discussion of pilot data

Machine learning

The results obtained demonstrate the feasibility of using accelerometric readings to classify hypokinesia / bradykinesia (“on-off”) and dyskinesia. Neural networks appeared to perform better than the classification tree algorithm (table 11). This is to be expected because neural networks are more flexible in creating decision boundaries between classes (neural nets can use an arbitrary hyperplane to separate classes, whereas classification trees can only divide the feature space using one dimension at a time). The advantage of classification trees is that the tree that is generated is far easier for a human to interpret and hence easier for a human to trust. This may become relevant when the results are used to influence the decisions of physicians.

Table 11. C-index results for neural network models (pilot study)

Model	Training Set c-index (average for 5 fold cross-validation)	Test Set c-index
#1 (Patient1, on_off)	0.87965	0.88084
#2 (Patient2, on_off)	0.85376	0.92142
#3 (Patient1, dyskinesia)	0.84994	0.91144
#4 (Patient2, dyskinesia)	0.88299	0.94106

Sensors

It is relevant to note that, in general, detecting hypokinesia / bradykinesia (“on-off”) required the input of more sensor sites than detecting dyskinesia. The level of “on-off” represents the level of voluntary activities. Voluntary activities may be more focal and have a broader range of magnitudes than dyskinetic activities. More focal

movements (e.g. just one arm or leg) would require more sensors to detect it. In contrast, movements that involve several limbs may require sampling from just one of those limbs. If a particular state (e.g. “off”, “on”, dyskinetic, not-dyskinetic) has broad range of possible acceleration values, then it may require more sensors to arrive at a classification. That is because the magnitude reading in one sensor is not specific enough to that state and comparisons with different sensors would be needed.

Surprisingly, despite the fact that it was not well secured to the patient, the hip sensor seemed to yield very important information. It was found to be an important factor in 3 of the 4 neural network models, despite the technical problems with the sensor in patient #2. Perhaps the hip was important because it is a measure of truncal movement. In the future, it would be reasonable to have better measurements of truncal movement.

We must note that tremor was not found to any clinical relevance in these two patients. That will not be true of future subjects and this must be accounted for in future models.

3. Final study

The pilot study showed that it is possible to classify the movement states of a Parkinsonian patient using a model based on other recordings made on the same patient. A truly useful classifier would be able to classify the movement states of a patient without any prior access to any accelerometric data for that patient. This could not be attempted using only the two patients of the pilot study.

3.1 Major Changes to Data Collection for the Final Study

Certain items of information that had not been collected in the pilot study and were collected in this final study include:

1. Use of both physician-based scoring and patient diaries. They were used to create separate classification models. Since the patient diary is the commonly used scheme against which this device would be compared, this project also attempted to classify movements based on patient diaries.
2. Use of more standardized metrics.
3. Baseline Hoehn & Yahr and MMSE scores (in order to gauge generalizability).

3.2 Major Changes to Data Processing for the Final Study

One of the major goals of the final study was to demonstrate that classification could be done on a patient even without the use of training data from that same patient. This could be a difficult problem because patients vary so much from each other. For instance, the cutoff above which I felt patient #1 was “on” was a score of 1.5, whereas the cutoff I used for patient #2 was 3.5 (see table 2). These cutoffs were based on my clinical observations, which is information that the classifying algorithms will not have access to. Therefore, it would be difficult for the algorithms to classify, if the cutoffs for dichotomization are not known. There are other problems too. For instance, the value of features may vary widely across patients. An algorithm such as CART which relies on fixed values of individual features to differentiate classes is likely to make errors. Conceptually, it seems more likely that algorithms such as logistic or

linear regression or neural networks which use combinations of features would be more robust. The following is a list of major changes in data processing as well as their rationales:

1. Arbitrary cutoffs were not used to dichotomize data. Instead a regression was performed and then a series of cutoffs were applied. The effectiveness of the algorithms was judged by how well they classified using all the dichotomization cutoffs. In this way, no clinical knowledge would be needed in order to choose the “right” cutoff for the patient and a general assessment could be obtained of how well the algorithms performed at all the possible classification tasks.
2. Cutoffs were based on percentile for the particular patient. Using cutoffs based on fixed numbers does not take into account what is considered a high score or a low score for that particular patient. Using given percentiles as cutoffs for the patient in question helps remedy this problem. This, however, was not applied to cutoffs used for dichotomizing diary scores. Diary scores are different because they inherently take into account what is high or low (“good” or “bad”) for that particular patient. That is because in diary scoring, the patient is asked to subjectively assess how “good” or “bad” they are doing and that would be based on the patient’s specific thresholds.
3. CART was not used. CART cutoffs are based on the value of only one feature and therefore were felt to be too susceptible to variations between patients. For instance, if a patient was 50% greater acceleration in all accelerometers than most other patients, then CART may well misclassify that patient. An algorithm that uses several features may be able to use a ratio between features in order to compensate for variation between patients.
4. Since regression algorithms were to be used, it seemed most appropriate to assess goodness of fit using error measures based on deviation of the predicted value from the actual value. These include mean squared error, mean absolute error and the R² statistic. It is true that Hosmer Lemeshow could be applied to each of the many dichotomizations that will be used, but then those values would then have to be combined into a single value. This would seem to be unnecessarily complicated and arbitrary. The standard error functions were therefore used.

5. Basic analysis was done on very short segments of accelerometry data and then the results of these basic analyses were aggregated over the entire 10-minute period of analysis. In the pilot study, features were derived from processing the entire 1-minute period as a whole. Observing the patients, I noticed that many actions occurred more in fits and starts than as continuous activity. This could lead to small burst of perhaps irrelevant activity “drowning out” more important subtler actions that are present for a large fraction of the time, but are not as dramatic. Using small segments to do basic analyses on and then aggregating these analyses (e.g. by taking covariance) makes short bursts of activity less relevant.
6. Frequency analyses are to be used. This is because of the importance of frequency as noted in the literature.

3.3 Method and Materials

Patients were recruited for the study from the movement disorders clinic at Memorial Hospital in Pawtucket, Rhode Island. The collaborating investigator at that hospital was Dr. Hubert Fernandez, who is a board-certified neurologist with a subspecialty in movement disorders. All participating patients were determined by Dr. Fernandez to have the diagnosis of Parkinson's disease and to have significant fluctuations in their movements, either fluctuations between bradykinesia and eukinesia ("on" vs. "off") and/or fluctuations in their degree of dyskinesia. All participating patients signed consent forms to participate in the study as well as consent forms to allow themselves to be videotaped. The study was approved by the internal review board of the hospital.

Five new patients participated in the final study. Additionally, the two patients from the pilot study were also included in the analysis. Since some types of data were only collected in the final study, some aspects of the analysis could only be performed on the 5 patients from the final part of the study.

All patients were tested using a Folstein mini-mental status examination (a common screening test for dementia) and required to have at least a score of 24/30. Additionally, a Hoehn and Yahr staging was performed on each patient to gauge the level of their Parkinsonism.

All patients in the final study were observed in the Parkinson's day room at the hospital. There they were observed by a neurologist (myself) and videotaped for later review by the same neurologist. Clinical observations

and scorings were recorded every 10 minutes. Additionally, patients were asked to complete a diary every 30 minutes noting the state of their movements. Simultaneous to the observations and scorings, the patients wore 5 accelerometers identical to those described in the pilot study. As in the pilot study, they were placed distally on each extremity as well as on the right hip (attached to belt or trousers). At a later time, all patients had their video recordings reviewed and a final determination of the scorings was determined.

The two patients in the pilot study did not have this systematic diary information recorded. Additionally, since the scoring scheme done in the room at the time differed for the two parts of the study, the videotapes of the two pilot study patients needed to be reviewed and re-scored.

Tables 12,13 and 14 contain list of the clinical scores that were obtained on the study patients.

Table 12. Final study clinical scores based on observations (recorded every 10 minutes). [This was obtained on the 5 patients of the final study]

Number	Label's name	Description
1	AIMS_overall	Level of dyskinesia overall for the whole body, based on AIMS ² (0=none, 1=minimal, 2=mild, 3=mild, 4=severe)
2	AIMS_UE	Level of dyskinesia for the upper extremities, based on AIMS ² (0=none, 1=minimal, 2=mild, 3=mild, 4=severe)
3	AIMS_LE	Level of dyskinesia for the lower extremities, based on AIMS ² (0=none, 1=minimal, 2=mild, 3=mild, 4=severe)
4	AIMS_trunk	Level of dyskinesia for the trunk, based on AIMS ² (0=none, 1=minimal, 2=mild, 3=mild, 4=severe)
5	Dyskinesia_old	Dyskinesia scoring scheme used in the pilot study (0=none, 1=mild, does not appear to impair patient at all, 2=moderate, appears to cause mild impairment of activity, 3= significant, appears to cause moderate impairment of activity, 4=severe, appears to cause severe impairment of activity)
6	BBH	Body bradykinesia and hypokinesia (item #31 of the Unified Parkinson's Disease Rating Scale ²⁷). Combining slowness, hesitancy, decreased arm swing, small amplitude and poverty of movement in general score as follows: (0=none, 1=minimal slowness giving movement a deliberate character; could be normal for some persons. Possibly reduced amplitude, 2=Mild degree of slowness and poverty of movement which is definitely abnormal. Alternatively, some reduced amplitude, 3=Moderate slowness, poverty or small amplitude of movement, 4=Marked slowness, poverty or small amplitude of movement)
7	On_off	Scoring scheme used in the pilot study to gauge "on" vs."off" state (0=significantly off, 1=mildly off, 2=ambiguous or intermediate, 3=mildly on, 4=definitely on)
8	Tremor_RUE	Rest tremor score for right upper extremity (based on item #20 of the UPDRS ²⁷). (0=absent, 1=slight and infrequently present, 2=mild in amplitude and persistent or moderate in amplitude but only intermittently present, 3=moderate in amplitude and present most of the time, 4=Marked in amplitude and present most of the time.
9	Tremor_LUE	Rest tremor score for left upper extremity (scored as above).
10	Tremor_RLE	Rest tremor score for right lower extremity (scored as above).
11	Tremor_LLE	Rest tremor score for left lower extremity (scored as above).

Table 13. Final study patient diary scores (recorded every 30 minutes). [This was obtained on the 5 patients of the final study]

Number	Label's name	Description
1	Diary	Patient notes how the patient believes he or she has been over the past 30 minutes (0=asleep, 1=off, 2=on without dyskinesia, 3=on with non-troublesome dyskinesia, 4=on with troublesome dyskinesia)

Table 14. Pilot study clinical scores based on observations (recorded every 10 minutes). [These were obtained also on the 2 pilot patients]

Number	Label's name	Description
1	On_off	Same as #7 above
2	BBH	Same as #6 above
3	Dyskinesia_old	Same as #5 above
4	AIMS_overall	Same as #1 above

3.3.1 Analyses

All accelerometry accelerometer data was off-loaded from the device's flash card and processed off-line. C language code was used to convert the recordings into ASCII format. Subsequent data processing and analyses were performed with the help of custom-written code in Java (Sun Microsystems), Matlab (Matlab12, by Mathworks), SAS(SAS institute) and Neurosolutions (by Neurodimension). SAS was used for linear regression and Neurosolutions was used for neural networks.

Because of the limited number of patients in the study, it was felt that there would not be enough patients for a true validation set. Without a true validation set, it would not be possible to adjust the features and parameters used in the linear regression and neural network models after the analysis has begun. Adjusting the features and parameters for the models in order to optimize the results, in the absence of a true validation set would likely lead to results that are unreliable and likely better than they would be in reality.

In order to avoid this problem, all the features that would be used were determined before analysis. When constructing the models, only default settings were used (no adjustment of parameters). The (3) features that were used in all the models were chosen based on experience from the pilot study, as well as from information obtained from the literature (results on the pilot study patients were later compared with those of the final study patients to determine whether using information from the pilot study to design the analysis led to inappropriately better results for the pilot study patients).

Each of the five 3-axis accelerometers consisted of two 2-axis accelerometers aligned perpendicularly to each other. Two of the four readings were for the same axis and were therefore averaged together (mean) to form a single reading .The readings from the three axes were combined to form a single reading corresponding to magnitude of the overall vector (using the Pythagorean equation: $\text{magnitude} = \sqrt{x^2+y^2+z^2}$).

The magnitude thus obtained was subject to a fast Fourier transform (FFT). The FFTs were obtained over 800 samples at a time. Since the device sampled at slightly less than 40Hz, this corresponded to slightly more than 20 seconds of recordings. The FFT values were then converted to real (non-imaginary) values by obtaining the absolute value. The sum of all values (area under the curve) corresponding to the following frequency ranges were obtained:

1. Sum of values 0.25Hz - 3Hz
2. Sum of values 4Hz - 6Hz

A ratio of the two sums was obtained. Since the unit of analysis was the 10-minute time period (corresponding to a single set of clinical scores), these ratios were combined to obtain a single value for the whole 10 minute time period. This was done by obtaining the covariance of this ratio in one accelerometer versus that of another accelerometer. There were 10 possible pairs of accelerometers for which covariance could be obtained, but based on the results of the 2 pilot study patients, only 3 were chosen:

1. Covariance of frequency ratio between hip and right upper extremity
2. Covariance of frequency ratio between hip and right lower extremity
3. Covariance of frequency ratio between hip and left lower extremity

Linear regression was performed by SAS version 8 (using the "analyst" program). Neural network models were constructed using Neurosolutions. All default parameters were used, including the following:

1. Model: multilayered perceptron
2. 1 hidden layer
3. regression
4. tanh transfer function
5. 1000 epochs

3.4 Results

The five final study patients had accelerometry recordings for a total of 13 hours, 38 minutes and 43 seconds. The break down is shown in table 15.

Table 15. Final study accelerometry recordings

Patient	Accelerometry recordings
#1	One sequence of 2:39:09 in length
#2	Two sequences. One 1:54:35 in length another 1:10:34 in length
#3	Two sequences. One 30:11 in length another 1:50:00 in length
#4	One sequence 2:30:22 in length
#5	One sequence 3:03:52 in length

All data were divided into 10-minute time blocks corresponding to the periods of time for clinical observations. If any part of that time period corresponding to a set of clinical scores had accelerometry data, then that period was analyzed. Since time blocks do not necessarily have data recorded for the entire 10-minute time period, it is possible for a patient to have more 10-minute time blocks than it might seem possible at first glance. For instance, if a patient had accelerometry recordings from 12:05 to 12:15 then that would be counted as two 10 minute time blocks (i.e. 12:00-12:10 and 12:10-12:20). In the end a total of 121 labeled ten-minute blocks were analyzed. This break down is shown in table 16.

Table 16. Time blocks by patient

Patient	Number of 10-minute blocks
Patient #1	17
Patient #2	20
Patient #3	12
Patient #4	16
Patient #5	19
Pilot Patient #1	32
Pilot Patient #2	15

The labels had the attributes as shown in table 17. General information about the final study patients is shown in table 18.

Table 17. Mean and Standard deviation of clinical labels for each patient

Patient	BBH (mean)	BBH (std)	AIMS_overall(mean)	AIMS_overall (std)	Diary (mean)	Diary (std)
#1	0.94	0.6587	0.47	0.7174	1.35	0.7859
#2	1.30	1.4179	1.55	0.9987	2.25	0.9105
#3	0.67	0.4924	1.25	1.4848	2.50	0.9045
#4	1.06	0.9287	0.19	0.5439	2.00	0.8165
#5	0.89	1.1496	1.37	1.1648	2.21	0.7133
Pilot 1	1.50	1.3912	0.91	0.9625	N/A	N/A
Pilot 2	1.47	1.3558	1.33	1.4960	N/A	N/A

(std: standard deviation from mean)

Table 18. General features of the final study patients

Patient	Age	Gender	Hoehn & Yahr	Handedness
#1	62	Male	Stage 3	Right
#2	62	Female	Stage 4	Right
#3	77	Female	Stage 4	Right
#4	52	Male	Stage 3	Right
#5	62	Male	Stage 3	Right

3.5 Analyses

Because of the small amount of observed tremor and because most dyskinesia appeared to be generalized, the analysis was focused on only 3 target variables as shown in table 19.

Table 19. Clinical labels used in analysis

1. body bradykinesia and hypokinesia (BBH)
2. AIMS overall (AIMS_overall)
3. diary (Diary)

Since the diary was only recorded every 3 time blocks, the patient's scoring was applied to all 3 previous time blocks (i.e. the past 30 minutes). This was appropriate because, when completing the diary, the patients were instructed to assess how they were "over the last 30 minutes."

The two scores initially used in the pilot study (on_off and dyskinesia_old) attempted to measure the same characteristics as target variables #1 and #2 above and were therefore felt to be redundant.

For both linear regression and neural network (regression), a leave-1-out method was used to compile a series of training and test sets. For instance, a model would be constructed using 6 patients and would then be tested on the patient not used in constructing the model. In the case of the diary, the model would be constructed based on only 4 patients and

then tested on the remaining patient. Since different patients had different numbers of time blocks, the training set for each model was obtained by randomly resampling the time blocks of each patient so that each patient would end up with 50 time blocks to be used to construct the model. This way, patients with more data would not be over-represented in the models.

As discussed earlier in the thesis, the time relation of target values should be taken into account. This could have been done using a hidden Markov model, but a very simple technique was used instead. The predicted value for each (10 minute) time block was substituted by the median value of the current time block, the previous time block and the time block that follows. The intention of this was to screen out predictions that were outliers and were not in line with the surrounding predictions.

The overall results were obtained as shown in tables 20 and 21. A discussion of the meanings of the various statistics is given below.

Table 20. Linear regression results overview

Target	Average correlation	Average c-index	Mean absolute error	R ²
BBH	0.6441	0.8219	0.7905	0.1220
AIMS (overall)	0.5289	0.7552	0.8301	0.2730
Diary	0.6143	0.8799 (0.8815)	0.6853	0.2262

Table 21. Neural network results overview

Target	Average correlation	Average c-index	Mean absolute error	R ²
BBH	0.6356	0.8043	0.8203	0.1885
AIMS (overall)	0.4495	0.6398	0.7717	0.3133
Diary	0.4125	0.7374 (0.7243)	0.6851	0.1563

Description of statistics:

Average correlation:

This was obtained by obtaining the correlation of the measured target value with the predicted target value for each of the patients. These correlations were then averaged (mean) to obtain a single value for "average correlation"

Average C-index:

C-index (equivalent to the area under the receiver operator characteristics curve) requires a dichotomous variable in order to be calculated. Clearly, the c-indices would be different if different cut-points would be used to

dichotomize the variables. Here, several different cut-points were used and c-index results for the different cut-points were averaged for each patient. Then the average of all the patients was calculated (i.e. the average c-index).

Since it was felt that the absolute value of the AIMS score or BBH score for a particular patient would not be as relevant as whether it is low or high for that particular patient, cutoffs were obtained based on percentiles for that patient. Nine cut-off were obtained (10 percentile, 20 percentile, 30 percentile, 40 percentile, 50 percentile, 60 percentile, 70 percentile, 80 percentile, 90 percentile). In contrast to the AIMS and BBH scores, the actual value of the diary score should be relevant clinically because it is a direct measure of how the hypokinesia, bradykinesia and dyskinesia affects the individual. Therefore, cut-offs were not obtained using percentiles for that particular patient, but rather were obtained by fixed cutoffs (0.5, 1.5, 2.5, 3.5). The average c-index obtained using the percentile method is included in parentheses for comparison.

Mean Absolute Error:

This was obtained by obtaining the mean absolute error for each patient and averaging it over all patients.

R2:

This is a statistic used to assess goodness-of-fit. A value of 1 corresponds to perfect prediction of the target value. A value of zero corresponds to a fit that is no better than simply guessing that the value is the same as the mean (of the data that were used to build the model).

More detailed statistics on all models are shown in tables 22-33.

Table 22. Linear regression BBH model: c-indices using different percentile cutoffs

Percentile as cutoff	Pat#1	Pat#2	Pat#3	Pat#4	Pat#5	Pilot1	Pilot2
10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
20	N/A	N/A	N/A	N/A	N/A	N/A	N/A
30	0.8750	N/A	N/A	N/A	N/A	N/A	N/A
40	0.8750	0.8132	1.0000	0.5818	N/A	0.4909	0.8796
50	0.8750	0.8132	1.0000	0.5818	N/A	0.4909	0.9018
60	0.8750	0.8132	1.0000	0.5818	0.8056	0.6412	0.9018
70	0.8750	0.8690	1.0000	0.8909	0.8056	0.6412	0.7000
80	0.8750	0.9531	1.0000	0.8909	0.9214	0.6412	0.7000
90	0.8571	0.9412	1.0000	0.8909	0.9375	0.7704	0.7000
Mean	0.8724	0.8672	1.0000	0.7364	0.8675	0.6126	0.7972

Table 23. Linear regression AIMS_overall model: c-indices using different percentile cutoffs

Percentile as cutoff	Pat#1	Pat#2	Pat#3	Pat#4	Pat#5	Pilot1	Pilot2
10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
20	N/A	0.8672	N/A	N/A	N/A	N/A	N/A
30	N/A	0.8672	N/A	N/A	N/A	N/A	N/A
40	N/A	0.8229	N/A	N/A	0.7679	N/A	N/A
50	N/A	0.8229	0.9000	N/A	0.6818	0.5992	N/A
60	N/A	0.8229	0.9000	N/A	0.6818	0.5992	0.9722
70	0.6136	0.8229	1.0000	N/A	0.6818	0.5992	0.9722
80	0.6136	0.8229	1.0000	N/A	0.6818	0.6594	0.9722
90	1.0000	0.6569	1.0000	0.4643	0.7708	0.6594	0.9722
Mean	0.7424	0.8132	0.9600	0.4643	0.7110	0.6233	0.9722

Table 24. Linear regression Diary model: c-indices (using fixed cutoffs)

Patient	0.5 cutoff	1.5 cutoff	2.5 cutoff	3.5 cutoff	Mean
Pat #1	N/A	0.9672	0.9672	N/A	0.9672
Pat #2	N/A	1.0000	0.8788	N/A	0.9394
Pat #3	N/A	N/A	1.0000	1.0000	1.0000
Pat #4	N/A	0.9273	0.4364	N/A	0.6818
Pat #5	N/A	0.9375	0.6667	N/A	0.8021
Mean	N/A	0.9602	0.7916	1.0000	

(note: there is no diary information for the 2 pilot patients)

Table 25. Neural Network regression BBH model c-indices using different percentile cutoffs

Percentile as cutoff	Pat#1	Pat#2	Pat#3	Pat#4	Pat#5	Pilot1	Pilot2
10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
20	N/A	N/A	N/A	N/A	N/A	N/A	N/A
30	0.8269	N/A	N/A	N/A	N/A	0.6208	N/A
40	0.8269	0.7363	1.0000	0.5455	N/A	0.6208	0.8796
50	0.8269	0.7363	1.0000	0.5455	N/A	0.7285	0.9018
60	0.8269	0.7363	1.0000	0.5455	0.8111	0.7285	0.9018
70	0.8269	0.8214	1.0000	0.7636	0.8111	0.7285	0.7000
80	0.8269	0.9063	1.0000	0.7636	0.9214	0.7971	0.7000
90	0.6667	0.8529	1.0000	0.7636	0.8750	0.8269	0.7000
Mean	0.8040	0.7982	1.0000	0.6545	0.8547	0.7216	0.7972

Table 26. Neural Network regression AIMS_overall model: c-indices using different percentile cutoffs

Percentile as cutoff	Pat#1	Pat#2	Pat#3	Pat#4	Pat#5	Pilot1	Pilot2
10	N/A	N/A	N/A	N/A	N/A	N/A	N/A
20	N/A	0.7578	N/A	N/A	N/A	N/A	N/A
30	N/A	0.7578	N/A	N/A	N/A	N/A	N/A
40	N/A	0.7031	N/A	N/A	0.8333	N/A	N/A
50	N/A	0.7031	0.4143	N/A	0.7670	0.6619	N/A
60	N/A	0.7031	0.4143	N/A	0.7670	0.6619	0.9722
70	0.4394	0.7031	0.7500	N/A	0.7670	0.7386	0.9722
80	0.4394	0.7031	0.7500	N/A	0.7670	0.7386	0.9722
90	1.0000	0.5490	0.7000	0.0714	0.8854	0.7386	0.9722
Mean	0.6263	0.6975	0.6057	0.0714	0.7978	0.7080	0.9722

Table 27. Neural Network regression Diary model: c-indices (using fixed cutoffs)

Patient	0.5 cutoff	1.5 cutoff	2.5 cutoff	3.5 cutoff	Mean
Pat #1	N/A	0.9808	0.9808	N/A	0.9808
Pat #2	N/A	0.9400	0.8385	N/A	0.8893
Pat #3	N/A	N/A	1.0000	1.0000	1.0000
Pat #4	N/A	0.1250	0.3750	N/A	0.2500
Pat #5	N/A	0.5208	0.6131	N/A	0.5670
Mean	N/A	0.6417	0.7615	1.0000	

(note: there is no diary information for the 2 pilot patients)

Table 28. Linear Regression: BBH model

Patient	Mean squared error	Mean absolute error
Pat #1	0.2774	0.4107
Pat #2	1.2878	0.8110
Pat #3	0.2639	0.4412
Pat #4	0.7444	0.6874
Pat #5	0.9631	0.8474
Pilot1	2.4257	1.2618
Pilot2	1.4208	1.0737

Table 29. Linear Regression AIMS_overall model

Patient	Mean squared error	Mean absolute error
Pat #1	0.4576	0.6457
Pat #2	0.7960	0.7346
Pat #3	1.4052	0.8851
Pat #4	0.4138	0.5921
Pat#5	1.4147	1.0292
Pilot1	0.8526	0.7733
Pilot2	1.8036	1.1510

Table 30. Linear Regression Diary model

Patient	Mean squared error	Mean absolute error
Pat #1	1.0761	1.0289
Pat #2	0.4006	0.5505
Pat #3	0.7095	0.6059
Pat #4	0.6863	0.6972
Pat #5	0.4477	0.5439

Table 31. Neural Network BBH model

Patient	Mean squared error	Mean absolute error
Pat #1	0.3768	0.5291
Pat #2	1.2784	0.8767
Pat #3	0.3592	0.5889
Pat #4	1.1745	0.8822
Pat #5	0.7146	0.7284
Pilot1	2.1900	1.1556
Pilot2	1.2043	0.9811

Table 32. Neural Network AIMS_overall model

Patient	Mean squared error	Mean absolute error
Pat #1	0.2743	0.5202
Pat #2	1.1487	0.7806
Pat #3	1.9626	1.1259
Pat #4	0.3299	0.4908
Pat #5	0.8402	0.7094
Pilot1	0.6586	0.6954
Pilot2	1.8442	1.0794

Table 33. Neural Network Diary model

Patient	Mean squared error	Mean absolute error
Pat #1	1.0977	1.0338
Pat #2	0.3993	0.5208
Pat #3	0.4238	0.4628
Pat #4	0.8618	0.8178
Pat #5	0.5453	0.5904

3.6 Discussion

Linear regression performed better than neural network models. This may have happened because I was unable to adjust the parameters of the neural network in order to optimize results, which was a necessary restriction to avoid overfitting. Linear regression appeared to perform reasonably well for both the BBH (body bradykinesia / hypokinesia) model and the Diary model (average c-indices of 0.8219 and 0.8719, respectively). Evaluation data shows a quite remarkable performance of linear regression in classifying the diary score. Clinically, the most important (i.e. relevant) information for management of Parkinsonism is:

1. Whether the patient feels on or off
2. Whether the patient has troublesome dyskinesia or not

The clinician's observations are generally felt to be less relevant. In addition, non-troublesome dyskinesias are not nearly as relevant as troublesome dyskinesias. These two most relevant pieces of information are discerned nearly perfectly by the linear regression model (for diary). The model is able to discern off (diary scores 0,1) from on (diary scores 2,3,4) with a c-index of 0.9602 and to discriminate troublesome dyskinesias (diary score 4) from all others with a c-index of 1.

The AIMS model does perform less well than all the rest (average c-index 0.7552). In the pilot study, dyskinesia had actually been easier to predict than on_off. The reason why the models performed less well across patients is not clear.

I have chosen to use c-indices for dichotomized data rather than mean absolute error, mean squared error or the R^2 statistic as the main determinant of success or failure of a model because such dichotomization will likely be necessary in order to produce a report that the managing clinician could readily understand. As can be seen in tables 20 and 21, there is generally an inverse relationship between average c-indices and mean absolute error (with the exception of the neural network model for BBH). The R^2 statistic, which uses the squared errors in its calculation, does not increase with the better models as might have been expected. This is likely because using the square of errors makes it particularly susceptible to a few predicted values that are far off from their target values. This would also be true of the mean absolute error, but to a lesser degree. When the data are going to be dichotomized anyway, these error measures would not be that relevant.

Since no true validation set could be constructed, a cross validation approach was used, but all the features and parameters used in model construction were fixed before analysis was performed. Since the pilot patients were included in most of the analysis and the lessons learned from the pilot study were used in constructing models, it could be argued that the pilot study patients may receive an unfair advantage by having the model specifically tailored to them. While this can not be entirely dismissed, it is possible to demonstrate that the models did not perform grossly better on these patients. Table 34 below does not show a dramatic difference between the pilot study patients and all the patients as

a whole. In some models they performed slightly better and in some models slightly worse.

Table 34. Performance of pilot study patients as compared with all patients in the study

Model	Pilot patient #1 average c-index	Pilot Patient #2 average c-index	Mean average c-index of pilot patients	Mean average c-index of all 7 patients	Mean average c-index of all patients excluding pilot study patients
BBH (linear regression)	0.6126	0.7972	0.7094	0.8219	0.8687
AIMS_overall (linear regression)	0.6233	0.9722	0.7977	0.7552	0.7382
BBH (neural network)	0.7216	0.7972	0.7594	0.8043	0.8223
AIMS_overall (neural network)	0.7080	0.9722	0.8401	0.6398	0.5597
Mean	0.6664	0.8847	0.7766	0.7553	0.7472

4. Conclusion

The results that were obtained in this study appear to be quite promising. A clinically useful classifier would need to be constructed using far more patients than were used here. This likely would yield even more accurate models. If higher sensitivity and specificity would be desired, readily available data about the patient might be integrated into the models to yield even better results. For instance, age, gender, handedness, and Parkinson's stage are easily available and may help fine-tune the models for specific patients.

This research was designed to demonstrate that a device that uses accelerometers and its respective classifier is feasible. A study designed to actually develop such a device would need to be run differently. Recordings used to build the models should be done in the patients' natural environment at home. In that way, it would correspond better to the environment patients will be in when they are using the device clinically. Likely, the initial work should use only patient diaries and not clinical observations, for two reasons: Firstly, observations would limit the patient's natural movement (and therefore limit the usefulness of the models constructed from that data); additionally, diaries have the most clinically relevant information.

5. References and Notes

1. Aminian K, Robert P, Buchser EE, Rutschmann B, Hayoz D, Depairon M. "Physical Activity Monitoring Based on Accelerometry: Validation and Comparison with Video Observation." *Med & Biol Eng & Comp* 37 (1999): 304-308.
2. "Available Dyskinesia Clinical Rating Scales." *Mov Disord* 14 supplement 1 (1999): 75-80.
3. Bergen JA, Carter NB, Craig J, Macfarlane D, Smith EF, Beumont PJV. "AIMS Ratings – Repeatability." *British Journal of Psychiatry* 152 (1988): 670-673.
4. Beuter A, Edwards R. "Using Frequency Domain Characteristics to Discriminate Physiologic and Parkinsonian Tremors." *Jour of Clin Neurophys* 16(5) (1999): 484-494.
5. Block G, Liss C, Reines S, et al, for the CR First Study Group. "Comparison of Immediate-Release and Controlled Release Carbidopa/Levodopa in Parkinson's Disease: a Multicenter 5-Year Study." *Eur Neurol* 37(1) (1997): 23-7.
6. Brown P, Manson A. "Dyskinesias Assessment and Ambulatory Devices." *Mov Disord* 14 supplement 1 (1999): 67-68.
7. Burkhard PR, Shale H, Langston JW, Tetrud JW. "Quantification of Dyskinesia in Parkinson's Disease: Validation of a Novel Instrumental Method." *Mov Disord* 14(5) (1999): 754-763.
8. Damier PD, Jaillon C, Clavier I, Arnulf I, Bonnet AM, Bejjani BP, Agid Y. "Dyskinesia Assessment in Phase 2 Studies." *Mov Disord* 14 supplement 1 (1999): 54-59.
9. Dunnewold RJW, Hoff JI, van Pelt HCJ, Fredrikze PQ, Wagemans EAH, van Hilten BJJ. "Ambulatory Quantitative Assessment of Body Position, Bradykinesia, and Hypokinesia in Parkinson's Disease." *Jour of Clin Neurophys* 15(3) (1998): 235-242.
10. Dunnewold RJW, Jacobi CE, van Hilten JJ. "Quantitative Assessment of Bradykinesia in Patients with Parkinson's Disease." *Jour of Neurosci Meth* 74 (1997): 107-112.
11. Edwards R, Beuter A. "Sensitivity and Specificity of a Portable System for Measuring Postural Tremor." *Neurotoxicol and Teratol* 19(2) (1997): 95-104.
12. Fahn S. "The Spectrum of Levodopa-Induced Dyskinesias." *Ann of Neurol* 47 supplement 1 (2000): s2-s11.
13. Foerster F, Smeja M. "Joint Amplitude and Frequency Analysis of Tremor Activity." *Electromyog and Clin Neurophys* 39(1) (1999): 11-19.

14. Goetz, CG. "Rating Scales for Dyskinesia in Parkinson's Disease." *Mov Disord* 14 supplement 1 (1999): 48-53.
15. Haines J. "Ultrasound Apparatus for the Measurement of Facial Movement." *Med Biol Eng Comput* 21(1) (1983): 105-6.
16. Hauser RA, Friedlander J, Zesiewica TA, Adler CH, Seeberger LC, O'Brien CF, Molho ES, Factor SA. "A Home Diary to Assess Functional Status in Patients with Parkinson's Disease with Motor Fluctuations and Dyskinesia." *Clin Neuropharmacol* 23(2) (2000): 75-81.
17. Hoff JJ, van Hilten BJ, Roos RA. "A Review of the Assessment of Dyskinesias." *Mov Disord* 14(5) (1999): 737-743.
18. Hoff JJ, v/d Plas AA, Wagemans EAH, van Hilten JJ. "Accelerometric Assessment of Levodopa-Induced Dyskinesias in Parkinson's Disease." *Mov Disord* 16(1) (2001): 58-61.
19. Hoff JJ, Wagemans EA, van Hilten BJ. "Ambulatory Objective Assessment of Tremor in Parkinson's Disease." *Clin Neuropharmacol* 24(5) (2001): 280-283.
20. Katayama S. "Actigraph Analysis of Diurnal Motor Fluctuations During Dopamine Agonist Therapy." *Europ Neurol* 46 supplement 1 (2001): 11-17.
21. Keijers NLW, Horstink MWIM, van Hilten JJ, Hoff JJ, Gielen CCAM. "Detection and Assessment of the Severity of Levodopa-Induced Dyskinesia in Patients with Parkinson's Disease by Neural Networks." *Mov Disord* 15(6) (2000): 1104-1111.
22. Kiani K, Snijders C, Gelsema ES. "Recognition of Daily Life Motor Activity Classes Using an Artificial Neural Network." *Arch of Physic Medic and Rehabil* 79 (1998): 147-154.
23. Manson AJ, Brown P, O'Sullivan JD, Asselman P, Buckwell D, Lees AJ. "An Ambulatory Dyskinesia Monitor." *Jour of Neurol, Neurosurg and Psych* 68 (2000): 196-201.
24. Marsden CD, Parkes JD. "Success and Problems of Long-Term Levodopa Therapy in Parkinson's Disease." *Lancet* 1 (1977): 345-9.
25. Matsumoto JY, Dodick DW, Stevens LN, Newman, RC, Caskey PE, Fjerstad W. "Three-Dimensional Measurement of Essential Tremor." *Mov Disord* 14(2) (1999): 288-294.
26. McClelland HA, Fairbairn A, McDonald M, Nicholas TR, Cox GA. "The Evaluation of an Ultrasound Detector (UD) in the Measurement of Oro-facial Dyskinesia." *Int Clin Psychopharmacol* 2(2) (1987): 159-64.

27. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. "The Unified Parkinson's Disease Rating Scale (UPDRS): Status and Recommendations." *Mov Disord* 18(7) (2003):738-50.
28. Norman KE, Edwards R, Beuter A. "The Measurement of Tremor Using a Velocity Transducer: Comparison to Simultaneous Recordings Using Transducers of Displacement, Acceleration and Muscle Activity." *Jour of Neurosci Meth* 92 (1999): 41-54.
29. Nutt John. "A Unified Dyskinesia Rating Scale for L-DOPA Induced Dyskinesias?" *Mov Disord* 14 supplement 1 (1999): 74.
30. Resek G, Haines J, Sainsbury P. "An Ultrasound Technique for the Measurement of Tardive Dyskinesia." *Br J Psychiatry* 138 (1981):474-8.
31. Rowland Lewis P, ed. *Merritt's Neurology*, 10th ed. Philadelphia: Lippincott Williams & Wilkins (2000): 679.
32. Scholz E, Bacher M, Diener HC, Dichgans J. "Twenty-Four-Hour Tremor Recordings in the Evaluation of the Treatment of Parkinson's Disease." *Jour of Neurol* 235 (1988): 475-484.
33. Schrag A, Quinn N. "Dyskinesias and Motor Fluctuations in Parkinson's Disease a Community-Based Study." *Brain* 123 (2000): 2297-2305.
34. Shoulson I et al. (the Parkinson Study Group). "Evaluation of Dyskinesias in a Pilot, Randomized, Placebo-Controlled Trial of Remacemide in Advanced Parkinson Disease." *Arch of Neurol* 58 (2001): 1660-1668.
35. Someran EJW, van Gool WA, Vonk BFM, Mirmiran M, Speelman JD, Bosch DA, Swaab DF. "Ambulatory Monitoring of Tremor and Other Movements Before and After Thaladotomy: a New Quantitative Technique." *Jour of Neurolog Sci* 117(1-2) (1993): 16-23.
36. Spieker S, Boose A, Breit S, Dichgans J. "Long-Term Measurement of Tremor." *Mov Disord* 13 supplement 3 (1998): 81-4.
37. Spieker S, Jentgens C, Boose A, Dichgans J. "Reliability, Specificity and Sensitivity of Long Term Tremor Recordings." *Electroenceph and Clin Neurophys* 97 (1995): 326-331.
38. Starner T., Auxier J., Ashbrook D., Gandy M. "The Gesture Pendant: a Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring." *Proceed of the Fourth Internat Symp on Wearable Comp* (2000).
39. van Hilten JJ, Hoff JJ, Middelkoop HAM, Roos RAC. "The Clinimetrics of Hypokinesia in Parkinson's Disease: Subjective Versus Objective Assessment." *Jour Neural Transm [P-D Sect]* 8 (1994): 117-121.

40. van Hilten JJ, Hoogland G, van der Velde EA, Middelkoop HAM, Kerkhof GA, Roos RAC. "Diurnal Effects of Motor Activity and Fatigue in Parkinson's Disease." *Jour of Neurol, Neurosurg and Psych* 56 (1993): 874-877.
41. van Hilten JJ, Hoogland G, van der Velde EA, van Dijk JG, Kerkhof GA, Roos RAC. "Quantitative Assessment of Parkinsonian Patients by Continuous Wrist Activity Monitoring." *Clin Neuropharmacol* 16(1) (1993): 36-45.
42. van Hilten JJ, Kabel JF, Middelkoop HAM, Kramer CGS, Kerkof GA, Roos RAC. "Assessment of Response Fluctuations in Parkinson's Disease by Ambulatory Wrist Activity Monitoring." *Acta Neurol Scand* 87 (1993): 171-177.
43. van Hilten JJ, Middelkoop HAM, Kerkhof GA, Roos RAC. "A New Approach in the Assessment of Motor Activity in Parkinson's Disease." *Jour of Neurol, Neurosurg and Psych* 54 (1991): 976-979.
44. van Hilten JJ, Middelkoop HAM, Kuiper SIR, Kramer CGS, Roos RAC. "Where to Record Motor Activity: an Evaluation of Commonly Used Sites of Placement for Activity Monitors." *Electroenceph and Clin Neurophys* 89 (1993): 359-362.
45. Walker DJ, Heslop PS, Plummer CJ, Essex T, Chandler S. "A Continuous Patient Activity Monitor: Validation and Relation to Disability." *Physiolog Meas* 18 (1997): 49-59.
46. Waters CH. "Advances in Managing Parkinson's Disease." *Hospit Practice* 36(6) (2001): 27-32, 41.
47. Webster DD. "Critical Analysis of the Disability in Parkinson's Disease." *Mod Treat* 5(2) (1968): 257-282.
48. WeMove. Retrieved July 22, 2003: www.wemove.org/par/par_faq.html.
49. Westerterp KR. "Physical Activity Assessment with Accelerometers." *Internat Jour of Obesity* 23 supplement 3 (1999): s45-s49.
50. Widner H, Defer GL. "Dyskinesia Assessment: From CAPIT to CAPSIT." *Mov Disord* 4 supplement 1 (1999): 60-66.
51. Wilson AD. "Adaptive Models for the Recognition of Human Gesture." Ph.D. M.I.T. thesis (2000).