Mobile systems for monitoring Parkinson's disease
To my family
Studies from the School of Science and Technology at Örebro University 22

Mevludin Memedi

Mobile systems for monitoring Parkinson's disease
Abstract
This thesis presents the development and evaluation of IT-based methods and systems for supporting assessment of symptoms and enabling remote monitoring of Parkinson’s disease (PD) patients. PD is a common neurological disorder associated with impaired body movements. Its clinical management regarding treatment outcomes and follow-up of patients is complex. In order to reveal the full extent of a patient’s condition, there is a need for repeated and time-stamped assessments related to both patient’s perception towards common symptoms and motor function. In this thesis, data from a mobile device test battery, collected during a three year clinical study, was used for the development and evaluation of methods. The data was gathered from a series of tests, consisting of self-assessments and motor tests (tapping and spiral drawing). These tests were carried out repeatedly in a telemedicine setting during week-long test periods.

One objective was to develop a computer method that would process traced spiral drawings and generate a score representing PD-related drawing impairments. The data processing part consisted of using the discrete wavelet transform and principal component analysis. When this computer method was evaluated against human clinical ratings, the results showed that it could perform quantitative assessments of drawing impairment in spirals comparatively well. As a part of this objective, a review of systems and methods for detecting the handwriting and drawing impairment using touch screens was performed. The review showed that measures concerning forces, accelerations, and radial displacements were the most important ones in detecting fine motor movement anomalies.

Another objective of this thesis work was to design and evaluate an information system for delivering assessment support information to the treating clinical staff for monitoring PD symptoms in their patients. The system consisted of a patient node for data collection based on the mobile device test battery, a service node for data storage and processing, and a web application for data presentation. A system module was designed for compiling the test battery time series into summary scores on a test period level. The web application allowed adequate graphic feedback of the summary scores to the treating clinical staff. The evaluation results for this integrated system indicate that it can be used as a tool for frequent PD symptom assessments in home environments.

Keywords: Home assessments, Parkinson’s disease, mobile computing technology, e-diary, discrete wavelet transform, principal component analysis.
Included papers

This thesis is a summary of the following three papers, which are referred to in the text as Paper I, Paper II and Paper III.


Comments on my contribution:

**Paper I** – I planned the literature review in the paper together with Jerker Westin. I conducted the review and wrote the first version of the manuscript.

**Paper II** – I was partly involved in the development of the methods and I built the framework for the collection of clinical ratings of spirals. I wrote parts of the manuscript and reviewed the rest of it.

**Paper III** – I was involved in method development, analysis of the results, and writing of the first version of the manuscript.
# Table of contents

1 INTRODUCTION ........................................................................................................................ 11
   1.1 INFORMATION TECHNOLOGY IN HEALTHCARE ................................................................. 11
   1.1.1 MOBILE COMPUTING TECHNOLOGY .............................................................................. 13
   1.1.2 INFORMATION PROCESSING ............................................................................................. 13
   1.2 PARKINSON’S DISEASE ......................................................................................................... 15
   1.2.1 CLINICAL FEATURES .......................................................................................................... 15
   1.2.2 TREATMENT .......................................................................................................................... 16
   1.3 MOTIVATION ........................................................................................................................... 16
   1.4 OBJECTIVES ............................................................................................................................. 17
   1.5 THESIS OUTLINE ..................................................................................................................... 17

2 BACKGROUND ........................................................................................................................... 19
   2.1 PD SYMPTOM ASSESSMENT IN CLINICAL SETTINGS ............................................................. 19
   2.2 QUANTITATIVE ASSESSMENT OF FINE MOTOR IMPAIRMENT .............................................. 20
   2.3 SYSTEMS FOR MONITORING OF PD ......................................................................................... 21
   2.4 SUBJECTS AND DATA COLLECTION ........................................................................................... 22
   2.4.1 SUBJECTS ............................................................................................................................... 22
   2.4.2 DATA COLLECTION WITH THE MOBILE DEVICE TEST BATTERY ........................................ 22

3 METHODOLOGY ............................................................................................................................... 27
   3.1 SYSTEMS FOR ASSESSING THE FINE MOTOR DYSFUNCTION (PAPER I) .................... 27
   3.2 DEVELOPMENT OF THE COMPUTER-BASED METHOD (PAPER II) ............................... 27
   3.2.1 MANUAL APPROACH .......................................................................................................... 27
   3.2.2 COMPUTER-BASED APPROACH ........................................................................................ 27
       DISCRETE WAVELET TRANSFORM .......................................................................................... 28
       PRINCIPAL COMPONENT ANALYSIS ...................................................................................... 29
       THE COMPUTER METHOD ....................................................................................................... 30
       DATA ANALYSIS ....................................................................................................................... 31
   3.3 THE WEB-BASED SYSTEM (PAPER III) ............................................................................... 33
   3.3.1 CALCULATION OF SUMMARY SCORES .............................................................................. 33
       DATA SUMMARY ...................................................................................................................... 33
       CALCULATION OF TEST BATTERY DIMENSIONS ................................................................ 34
       CALCULATION OF OVERALL TEST SCORE .......................................................................... 35
   3.3.2 SYSTEM DESCRIPTION ........................................................................................................ 35
## Glossary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADL</td>
<td>Activities of Daily Living</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>ATA</td>
<td>American Telemedicine Association</td>
</tr>
<tr>
<td>CSUQ</td>
<td>IBM Computer System Usability Questionnaire</td>
</tr>
<tr>
<td>DAPHNE</td>
<td>Duodopa in Advanced Parkinson’s: Health Outcomes &amp; Net Impact</td>
</tr>
<tr>
<td>DPSS</td>
<td>Data Processing Sub System</td>
</tr>
<tr>
<td>DSS</td>
<td>Decision Support System</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transform</td>
</tr>
<tr>
<td>E-MOTIONS</td>
<td>Evaluation of a Motor/Non-Motor-Test Intelligent Online System</td>
</tr>
<tr>
<td>GDI</td>
<td>Graphics Device Interface</td>
</tr>
<tr>
<td>HHT</td>
<td>Hilbert Huang Transform</td>
</tr>
<tr>
<td>HP</td>
<td>High-Pass filter</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>LP</td>
<td>Low-Pass filter</td>
</tr>
<tr>
<td>M-code</td>
<td>Matlab code</td>
</tr>
<tr>
<td>MDS</td>
<td>Movement Disorders Society</td>
</tr>
<tr>
<td>MDS-UPDRS</td>
<td>Movement Disorders Society’s revision of UPDRS</td>
</tr>
<tr>
<td>MSD</td>
<td>Mean Squared Deviation</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Digital Assistant</td>
</tr>
<tr>
<td>PDQ-39</td>
<td>Parkinson’s Disease Questionnaire</td>
</tr>
<tr>
<td>RDM</td>
<td>Remote Device Manager</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SMR</td>
<td>Standardized Manual Rating</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>UPDRS</td>
<td>Unified Parkinson’s Disease Rating Scale</td>
</tr>
<tr>
<td>UPW</td>
<td>UPDRS weighted overall test score</td>
</tr>
<tr>
<td>WA</td>
<td>Web Application</td>
</tr>
<tr>
<td>WAST</td>
<td>Microsoft’s Web Application Stress Tool</td>
</tr>
<tr>
<td>WSTS</td>
<td>Wavelet Spiral Test Score</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Information Technology in healthcare

Over the past several decades, Information Technology (IT) has produced major breakthroughs in healthcare and has had a great impact on transforming it from in-hospital to more advanced home healthcare [Koch, 2006, Chaudhry et al., 2006]. The factors that contribute to this transformation include: the nature of new diseases and their treatments; demographic changes in the population; demand for healthcare cost containment; increased availability of complex healthcare medical equipment and services in the home; increase in rehabilitation services and increased focus on self-care and quality of life [Pepe et al., 2004]. With this trend of moving healthcare from the hospital to the patient’s home, caregivers would need to remotely monitor and treat their patients [Stanberry, 2000, Hebert et al., 2006] and patients would also need to improve their ability for self-managed care through constant doctor-patient consultations [Chin, 2003]. In the case of patient groups suffering from chronic diseases, the information flow between the patient and caregiver and also within the caregivers is complex and challenging. It is for this reason that IT can efficiently improve information flow [Young et al., 2007]. Information processing and communication are centrally involved in healthcare activities such as: patient data collection; information sharing with patients; communication among healthcare professionals; decision making in diagnostics and therapeutics; interpretation of laboratory results; collection of clinical research data, to name just a few [Balas et al., 1996, Georgiou, 2002].

The application of modern IT-based systems that are developed to support the clinical management of diseases in home healthcare provide an opportunity to reduce medication and diagnostic errors, increase efficiency, and support healthcare professionals [Ammenwerth et al., 2003]. Information Technology (IT) can be defined as “the use of electronic machines and programs for the processing, storage, transfer and presentation of information” [Alter, 1996, Björk, 1999]. According to Alter [1996], the three characteristics that make the use of IT efficient and effective are modularity, compatibility, and reusability. Modularity refers to the separation of the system into a set of independently developed, tested and understood subsystems. Compatibility is the extent to which the technology works with other complementary technologies. And finally, reusability means that system modules can be designed and used in different situations without major changes. Multiple studies have demonstrated that IT-based systems can improve the quality and efficiency of healthcare while reducing its costs, improve the physicians’ performance and patient outcomes, increase patient compliance and make a significant difference in family medicine [Blum,
1986, Balas et al., 1996, Hunt et al., 1998, Poissant et al., 2005, Yi et al., 2006, Chaudhry et al., 2006, Hebert et al., 2006].

Recent advances in a variety of disciplines like wireless communications, mobile computing, sensing technology, decision support systems and the web technology enable patients to be monitored remotely while offering reliable and cost-effective home healthcare [Bellazzi et al., 2001, van Halteren et al., 2004, Lu et al., 2005, Chen et al., 2011]. The process of applying the above technologies to healthcare is known as telemedicine.

The American Telemedicine Association (ATA) defines telemedicine as “the use of medical information exchanged from one site to another via electronic communications to improve patients’ health status. Closely associated with telemedicine is the term "telehealth,” which is often used to encompass a broader definition of remote healthcare that does not always involve clinical services. Videoconferencing, transmission of still images, e-health including patient portals, remote monitoring of vital signs, continuing medical education and nursing call centers are all considered part of telemedicine and telehealth” [ATA]. One of the telemedicine services is remote patient monitoring which uses devices to remotely collect and send data to a monitoring station for interpretation. The application of telemedicine to the home environment can also be described as telehomecare, home telehealth or home based eHealth [Koch, 2006]. In 2003, the ATA produced guidelines (Home Telehealth Clinical Guidelines) to establish a set of universal principles in order to regulate the development and deployment of telemedicine for homecare. These principles include criteria for patients, health providers, and technology. The patient criteria include guidelines on study ethics, patient privacy and confidentiality, patient education and satisfaction. The healthcare provider criteria define how the healthcare professionals can improve delivery of care by education and administration. The technology criteria recommend the type of technology to be used, its maintenance and support.

Strict evaluation of IT in healthcare is recommended and of high importance for decision makers and users [Ammenwerth et al., 2003]. Koch [2006] has conducted a systematic review of existing scientific literature with a focus on the home healthcare domain. The main finding of this study is that a majority of technical papers were focused on vital sign monitoring and virtual visits (using audio/video consultations). However, a sizeable minority of papers were focused on the improvement of information access and communication as well as decision support. Kaplan [2001] reviewed clinical decision support systems (DSS) literature focusing on changes in clinical performance. His study showed that most of the studies that use randomized clinical trial design can assess the system performance but fail to investigate the factors that determine the systems’ usefulness. Few studies use naturalistic design in routine clinical settings with real patients. Ammenwerth et al. [2003] discusses three problems during the evaluat-
tion of IT in healthcare. These are the complexity of the evaluation object, complexity of an evaluation project, and motivation for the evaluation. Rahimi and Vimarlund [2007] reviewed published articles in the area of evaluations of IT-based systems applied in healthcare settings. The findings show that there is no standard framework for evaluation of IT in healthcare. Moreover, most of the articles focus on aspects concerning user satisfaction, financial benefits and improved organizational work. Carson et al. [1998] proposed the stakeholder matrix analysis for evaluation of IT systems in healthcare by including a number of dimensions for stakeholders and criteria for the evaluation process.

1.1.1 Mobile computing technology
Mobile computing refers to technologies that employ small portable devices and wireless communication networks that allow user mobility by providing access to data “anytime, anywhere” [Burley et al., 2005]. Mobile computing technology improves healthcare in a number of ways such as by providing healthcare professionals access to reference information and electronic medical records and improving communication among them. It also offers computerized monitoring of clinical information and provides clinical decision support to give healthcare professionals further feedback [Ruland, 2002, Bates et al., 2003]. An instance of mobile computing technology is the Personal Digital Assistants (PDA). These are light-weight handheld computers and one of their medical applications is decision support which provides real-time information access, clinical computational programs, and diagnostic data management [Lu et al., 2005]. The most common method for assessing symptoms in clinical settings is paper home diaries. Major disadvantages of paper home diaries are poor patient compliance for the timing of completion and inflexible data storage and analysis [Stone et al., 2003, Broderick, 2008]. Electronic diaries (e.g. PDAs) overcome these issues by including functions that remind patients to complete diary entries at the proper time, allow just one answer per entry, and stamp the time and the date of the entry [Drummond et al., 2005, Nyholm et al., 2004, Lyons and Pahwa, 2007]. However, other factors such as privacy and security issues, system and information quality, and limitations of mobile devices still need to be investigated [Wu et al., 2007].

1.1.2 Information processing
According to Van Bemmel and Musen [1997], healthcare professionals go through three stages to complete the diagnostic-therapeutic cycle. These steps include observations, diagnosis and therapy. The computerized systems can be used to analyze large collections of data derived from many patients in order to draw conclusions by method of induction. During information processing by a computer, we observe a process similar to the diagnostic-therapeutic cycle where
the stages are measurement and data entry, data processing and output generation. Van Bemmel and Musen furthermore define a six-layer model for structuring computer healthcare applications with increased complexity and a correspondingly increased dependence on human interaction with respect to each layer. The layers of the model are arranged from a lower to higher complexity in the following order: communication and telematics; storage and retrieval; processing and automation; diagnosis and decision making; therapy and control; and research and development.

The successful introduction of DSS into clinical use was based on the support that they extended to data acquisition, data reduction, and data validation [Van Bemmel and Musen, 1997]. Moreover, Van Bemmel and Musen suggest that with any decision support tool, a user-friendly presentation is of paramount importance.

The foundation of clinical DSS can be based upon statistical pattern recognition and machine learning methods such as artificial neural networks (ANN). These methods can provide predictions, in terms of diagnostic or therapeutic outcomes, by using clinical information in a systematic way thus supporting clinicians in their decision-making processes. They are useful for modelling relationships between a set of independent observations (predictors) taken at different time intervals and a set of dependent outcomes (predicted scores). In the presence of multiple and long-term captured observations, there is a need for a multivariate analysis that considers several random observations or variables simultaneously, each of which are considered equally important at the start of the analysis [Manly, 1994]. Multivariate time series analysis accounts for dependencies between variables and also indicates which ones significantly do and do not add any predictive information.

Additionally, these methods should be evaluated for their metrics such as validity and reliability [Kudyba, 2004]. Validity is the extent to which an instrument measures what it intends to measure and nothing else [van de Ven-Stevens et al., 2009]. It is usually assessed by statistical tests such as concordance correlation coefficients and factor analysis. Reliability refers to the extent to which an instrument is free from measurement error including the internal consistency of its items and the reproducibility of its scores [van de Ven-Stevens et al., 2009]. Internal consistency is often assessed by Cronbach’s Alpha coefficient whereas reproducibility is assessed by the intra-class correlation coefficient or the kappa statistic.

The type of method selected depends on the type of outcome which is desired. For numeric outcomes, numeric prediction can be applied whereas for nominal outcomes, classification is often used. For any predictive method, its performance can be determined by looking at its accuracy or equivalently, at its errors. For numeric prediction situations, the common measure of error is the correla-
tion coefficient. For classification, the common measure of error is the proportion of errors made over a whole set of instances, called the error rate. The data is usually divided randomly into two sets. The part which is used for building the method is called the training set. However, the performance of the method should be evaluated by assessing the error on the dataset that played no part in its formation [Witten and Frank, 2005]. This independent dataset is called the test set. When the amount of data for training and testing is limited, a more general technique, known as cross-validation, is used. This technique repeats the whole process, training and testing, several times with different random samples. In order to ensure that random sampling is done in a way that it guarantees the proper representation of each outcome in both training and test sets, the stratification procedure is employed. In cross-validation, the fixed number of folds or partitions of data should be decided. If it is decided that ten folds would be used then 90% of the data is randomly selected to be the training set whereas the remaining 10% is selected to be the testing set. This procedure is then repeated ten times, with different choices of training and testing sets, and then the results of these ten trials are averaged [Witten and Frank, 2005]. This is called ten-fold cross-validation but if stratification is adopted as well, it is stratified ten-fold cross-validation.

1.2 Parkinson’s disease

Parkinson’s disease (PD) is a progressive hypokinetic movement disorder which is caused by degeneration of dopamine producing nerve cells in the region of the brain called substantia nigra. This region of the brain is important for control of body movement and coordination. The prevalence of PD increases with age; approximately 2% of people over the age of 65 have this disease. Although the etiology of PD still remains unclear, the disease probably results from an interaction between genetic and environmental factors [Warner and Schapira, 2003].

1.2.1 Clinical features

Clinical symptoms develop with a substantial variability among patients once there is at least 50% degeneration of dopaminergic neurons [Grosset et al., 2009]. The four cardinal motor symptoms comprise of bradykinesia (slowness of initiating voluntary movements), rigidity (increased muscle tone), tremor (a 3-5 Hz hand tremor) and impaired postural stability. The motor symptoms are often accompanied by non-motor symptoms such as fatigue, sleep disorders, cognitive impairment and psychotic features [Poewe, 2008]. The diagnosis of PD is usually made by clinical assessments. Thus the diagnostic criteria include the presence of bradykinesia in combination with one or more of the other three motor symptoms plus a positive response to levodopa administration which works as a
dopamine replacement and helps in reduction of symptoms and prolongation of life expectancy.

1.2.2 Treatment
Levodopa is a dopamine precursor and several decades after its introduction, it remains the “gold standard” oral treatment for PD [Fahn, 2003, Schapira et al., 2009]. In the early stages of the disease, the therapeutic effect of levodopa is very good and helps in improving the patient’s motor function. However, with disease progression and long-term levodopa therapy, patients start to experience motor complications or fluctuations. Their motor condition fluctuates between the “off” state (as a result of insufficient levodopa levels) and the “on” state (in which levodopa levels are enough for the patient to respond as a non-parkinsonian person). In addition to these two motor conditions, patients in the “on” state may develop abrupt involuntary movements, also known as dyskinesias, in response to peak levels of medication.

The side-effects of levodopa therapy are not only related to motor symptoms but to non-motor symptoms as well. Non-motor fluctuations appear both in the “on” and “off” states [Gunal et al., 2002]. In the long term, these fluctuations related to motor and non-motor symptoms may contribute to severe disability amongst patients. It has been found that fluctuations to a large extent result from the short half-life and irregular absorption of oral levodopa therapy [Kurlan et al., 1986, Dijaldetti et al., 1996]. Continuous delivery of a levodopa/carbidopa gel (Duodopa®, Abbot Laboratories) into the duodenum has been shown to control motor fluctuations in advanced PD [Nyholm et al., 2003, Nyholm et al., 2005]. In general, medications must be fine-tuned to the individual needs with regard to the timing and quantity of each levodopa dose and with regard to the food intake, mood and daily physical activities.

1.3 Motivation
The research described in this thesis is part of the “Evaluation of a Motor/Non-Motor-Test Intelligent Online System” (E-MOTIONS) project. The main objective of the project is to employ mobile technology and Artificial Intelligence in order to develop and evaluate IT-based methods that provide support for treatment and the evaluation of treatments for PD patients with motor fluctuations.

The complexity associated with the clinical management of PD is high as a result of random and frequent changes of symptoms during the day. In the presence of symptom fluctuations, detailed and frequent reporting of multiple measurements related to motor and non-motor symptoms is necessary [Weaver et al., 2005]. This is useful in order to reveal the full extent of a patient’s condition and avoid bias in measuring treatment effects [Isacsson et al., 2008]. Reimer et al. [2004] highlighted that when data was collected through paper home diaries for
at least one week, it was adequate and representative for interpretation. The adoption of IT tools can be applied to support the management of PD. A computerized system that enables PD patients to be monitored periodically and remotely offers a reliable and cost-effective screening of those patients.

The research group of the Department of Computer Engineering, Dalarna University has developed a test battery, consisting of both self-assessments and fine motor tests, for a handheld computer with touch screen in a telemedicine setting [Westin et al., 2010a]. This mobile device was designed as a tool for repeated use in a patient’s home environment.

1.4 Objectives

The main objective of this thesis is to develop and evaluate IT-based methods and systems for supporting assessment of symptoms and enabling remote monitoring of PD patients. This includes a review of methods and approaches found in literature suitable for quantifying fine motor dysfunction; method development useful for automatic assessment of PD-related drawing impairments; and finally addressing the problem of processing, summarizing and combining frequent and time-stamped time series, concerning subjective and objective measurements, into scores that relate to conceptual symptom dimensions and an overall health profile of the patient during a particular test period.

- **Paper I** – to survey techniques, measurements and data acquisition schemes which can be used for the development of assessment tools for quantifying fine motor skills in individuals suffering from movement disorders.
- **Paper II** – to develop and evaluate a computer-based method that automatically rates the drawing impairment in spiral drawings during event-based data acquisition.
- **Paper III** – to develop and evaluate a web-based system that processes time series of self-assessments and motor tests and delivers assessment support information to the treating clinical staff.

1.5 Thesis outline

The thesis is organized in the following format. In the *background* section, a description is provided for (a) PD symptom assessment in clinical settings, (b) quantitative assessment of fine motor impairment, (c) systems for enabling PD patient monitoring, and (d) data collection with a mobile device test battery. The *methodology* section describes the review strategy for Paper I, development of a computer-based method for Paper II, and development of the system for Paper III. The general results for the three papers are summarised in the *results* section.
followed by the discussion sections. And finally, the conclusions for all the three papers are presented.
2 Background

2.1 PD symptom assessment in clinical settings
The severity of each of the cardinal symptoms can be scored quantitatively using clinical rating scales. Rating scales are used as instruments by observers to evaluate PD-related disability and impairment in order to provide a comprehensive clinical picture. According to the World Health Organization, impairment is defined as an abnormality of body or organ structure or function; and disability is defined as a reduction of a person’s ability to perform a basic task [Simeonsson et al., 2000]. In PD, impairment relates to major symptoms as the underlying cause of a patient’s disability or inability to perform within the range of normal.

A common rating scale to describe the progression of symptoms is the Hoehn and Yahr scale [Hoehn and Yahr, 1967]. It is a five-point scale that describes disability and impairment in PD. Weaknesses include the scale’s mixing of impairment and disability, its non-linearity, strong emphasis on postural instability over other motor symptoms, and its lack of information delivery on non-motor problems [Goetz et al., 2004]. This scale has been largely supplanted by the Unified Parkinson’s Disease Rating Scale (UPDRS), which is much more complicated [Wolters et al., 2007].

UPDRS [Fahn et al., 1987] is a multi-modular scale and is the most widely used rating scale for the assessment of PD motor impairment and disability in published clinical trials [Mitchell et al., 2000, MDSTFRSPD, 2003]. The UPDRS is increasingly used as a gold standard reference scale. The scale is made up of four parts covering mentation, behaviour and mood (Part I); activities of daily living (ADL), (Part II); motor performance (Part III); and complications of therapy (Part IV). Parts I, II and IV are assessed by interviewing the patient whereas part III is assessed by physical examination. Parts I, II and III contain 44 questions, each of which are scored on a five-point scale rating from normal to severe corresponding to 0-4 respectively. Part IV contains 11 questions and each of these are scored either on a 0-4 scale or as yes/no responses. A “Total UPDRS” score is a combined sum of the four parts used to represent the global severity of impairment and disability. In the study performed by Ramaker et al. [2002], it was found that UPDRS is the most thoroughly studied scale with overall better validity and reliability compared to other scales. In 2008, the Movement Disorder Society (MDS) sponsored a revision of the current version of the UPDRS, resulting in a new version called MDS-UPDRS [Goetz et al., 2008]. This was based on recommendations from a previously published critique [MDSTFRSPD, 2003].

Non-motor symptoms are evaluated using non-motor scales such as Mini Mental State Examination, Dementia Rating Scale, PD Questionnaire (PDQ-39),
and others. The PDQ-39 is the most widely used disease specific measure of subjective health status that is completed by patients [Jenkinson et al., 1995, Peto et al., 1998].

A major disadvantage of clinical rating scales is their subjectivity; i.e. the scoring is dependent not only on the motor performance of the patient but also on the examiner’s interpretation [Hagell, 2000]. Shulman et al. [2006] showed that a patient’s subjective ratings of function may not correlate well with objective performance-based measurements. Hagell [2000] concluded furthermore that combining objective and quantitative measurements such as timed tests with clinical rating scales provides a good basis for clinical assessments of PD.

2.2 Quantitative assessment of fine motor impairment

Fine motor control can be defined as the ability to perform small and precise movements requiring hand-eye coordination. PD affects the fine motor control of an individual by slowing his/her movements and decreasing reaction time leading to the occurrence of involuntary movements. In addition, the disease may have an effect in diminishing handwriting abilities, also known as micrographia, characterized by reduced letter size [van Gemmert et al., 1999]. The most common procedure for assessing the severity of fine motor symptoms in movement disorders is through clinical rating scales based on observations and judgements by specialists such as the UPDRS motor disability score (Part III). Additionally, the level and nature of the fine motor impairment can be measured by computer-based analysis of handwriting and drawing movements.

The approaches of using accelerometers and digitizing tablets in quantifying essential tremor during writing and drawing tasks have been compared in a study from Elble et al. [1996]. The study showed that the digitizing tablet had several advantages over the accelerometer and provided accurate measures of tremor frequency and amplitude during fine motor movements. Touch screen devices not only record the x and y coordinates but also the pressure exerted by the instrument thus providing a rich source of information about movement dynamics. Digitizing graphic tablets have been widely utilized in studies where both healthy subjects and patients with different movement disorders have participated. Digitizing tablets have been mainly employed for recording digitized movement information of individuals. Off-line analysis and interpretation of the said information is done by employing digital signal processing techniques.

In most of the studies, spatial and spectral analysis of digitized drawing/handwriting specimens was performed to detect fine movement anomalies. Auto- and cross-spectral analyses of accelerometer data, digitized handwriting and spiral drawing were employed in tremor evaluation [Elble et al., 1996]. Non-parametrical kernel estimation of digitized x and y coordinates of handwriting movements was used in determining quantitative parameters, such as velocity.
and acceleration [Mergl et al., 1999]. The Fourier transform was used in the work done by Rudzinksa et al. [2007] for measuring three tremor intensity coefficients for displacement, velocity and acceleration respectively. In another study performed by Liu et al. [2005], the drawing velocity signals in horizontal, vertical, radial and tangential directions were bandwidth filtered using a four-pole Butterworth filter. The filter uses a frequency in the range of 1-5 Hz to extract dyskinetic movements and a frequency of >5 Hz to extract action tremor. Cross-correlation function along with Fourier transform was used to derive features suitable for quantification of scanned specimens of Archimedes spiral drawings [Miralles et al., 2006].

2.3 Systems for monitoring of PD

There have been a number of initiatives from different research groups to address the development of integrated systems that enable PD patient monitoring. Most of the approaches mainly target a specific aspect of the disease such as gait, posture, tremor, and cognition, and thus address a one-dimensional construct. An outline of objective measurements described in scientific literature is given as follows.

Keijsers et al. [2000] evaluated the feasibility of ANNs in assessing the severity of dyskinesia and distinguishing levodopa-induced dyskinesias from voluntary movements in which the neural network used data gathered from accelerometers mounted on the upper and lower limbs of the PD patients. The motor impairment in early to moderate PD was measured in a study done by Brewer et al. [2009] in which commercially available sensors were used to develop a protocol called Advanced Sensing for Assessment of Parkinson’s disease. This protocol measures the grip force as an individual tracks a sinusoidal or pseudorandom target force under certain cognitive-load conditions. Various modified regression techniques were used to predict an individual’s score on the UPDRS scale based on accelerometer data. Salarian et al. [2004] developed and validated an ambulatory gait analysis system for PD patients treated with Subthalamic Nucleus Deep Brain Stimulation based on body-attached sensors. Patel et al. [2009] presented results from a pilot study to assess the feasibility of using accelerometer data to estimate the severity of symptoms and motor complications in patients with PD. Accordingly, a web-based system was developed to allow clinicians to access analyzed sensor data and estimate UPDRS scores in their patients through two-way video communication [Chen et al., 2011]. Wearable sensors were also used in quantitative posture analysis in subjects with early PD to allow easy monitoring of their balance maintenance and analysis of their disease progression and fluctuations [Palmerini et al., 2011].

Motor function was tested in other studies by employing appliances such as computer keyboards [Giovannoni et al., 1999, Hoffmann et al., 2008] and cus-
tom hardware testing devices [Goetz et al., 2009]. Objective assessments of hand and finger movements in PD patients were done by using a computer-based monitoring and assessment tool developed and deployed on a standard computer [Cunningham et al., 2011]. The study showed that the data collected by this monitoring tool could distinguish between “on” and “off” states for each patient. Rigas et al. [2010] developed a decision support tool, dependent on a set of wearable sensors, for providing real time levodopa dose adjustments.

2.4 Subjects and data collection

2.4.1 Subjects
Sixty-five patients diagnosed with advanced PD at nine clinics around Sweden participated in an open longitudinal study DAPHNE (Duodopa in Advanced Parkinson’s: Health Outcomes & Net Economic Impact, Eudract No. 2005-002654-21). On inclusion, they were either treated with intestinal levodopa infusion (Duodopa), or they were candidates for receiving this treatment. UPDRS ratings were performed in afternoons at the start of the weeklong test periods. Hoehn and Yahr ratings are also available. Baseline characteristics are shown in Table 1.

<table>
<thead>
<tr>
<th>Patients (gender)**</th>
<th>Oral levodopa*</th>
<th>Levodopa infusion</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>37 (23m; 14f)</td>
<td>40(27m; 13f)</td>
<td>77(50m; 27f)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>63.6 ± 7.6</td>
<td>65.9 ± 6.9</td>
<td>64.8 ± 7.3</td>
</tr>
<tr>
<td>Duration of levodopa (years)</td>
<td>10.8 ± 4.5</td>
<td>15.7 ± 5.9</td>
<td>13.3 ± 5.7</td>
</tr>
<tr>
<td>Duration of duodenal levodopa (years)</td>
<td></td>
<td>2.2 ± 1.5</td>
<td></td>
</tr>
<tr>
<td>Hoehn and Yahr at best</td>
<td>2.2 ±1.1</td>
<td>2.5 ± 0.7</td>
<td>2.4 ± 0.9</td>
</tr>
<tr>
<td>Hoehn and Yahr at worst</td>
<td>3.8 ± 1.0</td>
<td>3.9 ± 0.8</td>
<td>3.8 ± 0.9</td>
</tr>
<tr>
<td>Total UPDRS at visit (afternoon)</td>
<td>50.1 ± 15.3</td>
<td>46.9 ± 16.7</td>
<td>48.6 ± 16.1</td>
</tr>
</tbody>
</table>

Table 1: Clinical features at baseline (mean ± one SD) in the DAPHNE study. * 27 of these 37 patients later started levodopa infusion treatment. ** 12 of these 77 patients did not use the test battery (7 were not Swedish-speaking and 5 were thought unable to handle the device or unwilling to use it) [Westin, 2010b].

2.4.2 Data collection with the mobile device test battery
A test battery was implemented in a handheld computer (PDA) with touch screen and mobile communication for enabling monitoring of fluctuating movement disorders, e.g. PD, in home environment conditions [Westin et al., 2010a]. The
The test battery includes a synchronized collection of self-assessments of common symptoms and fine motor tests. Self-assessments are designed to target more subjective aspects of a symptom whereas motor tests give a more objective view, being connected to physiological functioning of a patient. Assessments are done several times per day in the patient’s home during test periods of one week in duration. Patients answer questions and perform motor tests by touching the touch screen in a certain position with a stylus. The test battery includes a function for reminding the patient to enter, e.g. at 08:00, 12:00, 16:00 and 20:00 and the data entry is only possible during a limited time slot. Test results are transmitted via a mobile net to a central server for analysis and processing. The test battery was used by patients in the DAPHNE study and in total there were 379 test periods and 10439 test occasions.

Generally the test battery should provide information on patient state during a test period in order to (a) evaluate treatment effects in clinical practice and research, (b) follow-up treatments and disease progression, and (c) predict outcome based on patient state to optimise treatment strategy [Westin et al., 2010a].

Selection of self-assessment questions was based on questions from electronic diaries used in two previous studies [Nyholm et al., 2004, Nyholm et al., 2005]. Tables 2a and 2b show questions and their answer alternatives. Most of the diary questions (questions 1-6) relate to the previous 4 hours or ‘this morning’, depending on the actual time-of-day, and are of verbal descriptive scale type with answer alternatives ranging from 1 (worst) to 5 (best). In question 2, patients are asked to mark their portion of time spent in “off”, “on” and dyskinetic states during the last 4 hours or this morning. In addition to these questions, another question (question 7) relates to ‘right now’ and allows seven categories of motor states on a global Treatment Response Scale (TRS) ranging from -3 (very ‘off’) to 0 (‘on’) to +3 (very dyskinetic) [Nyholm et al., 2005].

Tables 3a and 3b provide descriptions and illustrations of motor tests. The objective of the tapping tests is to assess tapping speed, hand dexterity, reaction time and attention to visual stimuli. There are different tapping tests (tests 8-11) with and without visual cueing and are of 20 seconds long. Square tapping areas or “buttons” are displayed within a distance of approximately 15 mm in between, where at least one button will be active (red colour in the illustrations). For each tapping test, the average tapping speed (taps per 20 seconds) and accuracy (proportion of correct taps) were calculated and used in subsequent analyses. In addition to the tapping tests, spiral drawing tasks (tests 12-14) were added in which the patient traces a pre-drawn Archimedean spiral (defined by $r = a \cdot \theta$ in polar coordinates) from the center and out as accurate and fast as possible until completing three revolutions. The objective of the spiral test is to assess drawing impairment, which is most likely related to the handwriting ability, mainly affected by involuntary movements, such as tremor, bradykinesia and drug-
induced dyskinesias. The drawing patterns, concerning x, y coordinates and time, are recorded when a new pixel on the screen is touched (event-based data acquisition). This test is repeated three times per test occasion. As a result, a computerized method for assessing the degree of the drawing impairment of spiral drawings in a scale from 1 (mild impairment) to 10 (extremely severe impairment) was developed and evaluated against trained clinical raters, as presented in Paper II.

<table>
<thead>
<tr>
<th>Question#</th>
<th>Text</th>
<th>Answer alternatives</th>
</tr>
</thead>
</table>
| 1         | Have you had difficulty to walk (app. 100 meters) during the last four hours/this morning? | • Unable to walk  
• Difficult  
• With effort  
• Fairly well  
• Walks well |
| 2         | Mark the proportion of time you have been off, on and dyskinetic, during the last four hours/this morning. Tap to place two lines in the figure and move them to adjust percentages. | • Extremely off  
• Quite a bit  
• Moderately  
• Slightly  
• Not at all off |
| 3         | How much off have you been at worst during the last four hours/this morning?                      | • Extremely off  
• Quite a bit  
• Moderately  
• Slightly  
• Not at all off |
| 4         | How much dyskinetic have you been at worst during the last four hours/this morning?                | • Extremely dyskinetic  
• Quite a bit  
• Moderately  
• Slightly  
• Not at all dyskinetic |

Table 2a: Self-assessment questions and their answer alternatives [Westin et al., 2010a].
<table>
<thead>
<tr>
<th>Question#</th>
<th>Text</th>
<th>Answer alternatives</th>
</tr>
</thead>
</table>
| 5         | Have you had painful cramps (dystonia) during the last four hours/this morning? | • All the time  
• Most of the time  
• Half the time  
• A little of the time  
• No cramps at all |
| 6         | Have you been satisfied with your functioning during the last four hours/this morning? | • Not at all satisfied  
• Slightly  
• Moderately  
• Quite a bit  
• Completely satisfied |
| 7         | How is your condition right now?                                      | • Very dyskinetic  
• Moderately dyskinetic  
• A little dyskinetic  
• On  
• A little off  
• Moderately off  
• Very off |

Table 2b: Self-assessment questions and their answer alternatives [Westin et al., 2010a].

<table>
<thead>
<tr>
<th>Test#</th>
<th>Description</th>
<th>Test area</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Alternately tapping two fields (un-cued) using right hand. Two active buttons (red) are displayed. The buttons become inactive (black) 20s after the first button was pressed.</td>
<td><img src="image1" alt="Test 8" /></td>
</tr>
<tr>
<td>9</td>
<td>Same as 8 but using left hand.</td>
<td><img src="image2" alt="Test 9" /></td>
</tr>
<tr>
<td>10</td>
<td>Two buttons, where one is active (red) and the second is inactive are displayed. The buttons alternately become active with increasing speed. Alternations start with 2s per tap and speed increases with constant acceleration to 0.5s per tap after 20s. A sound is heard and a small filled circle is briefly (timeout) shown after a tap. Sounds on correct and missed taps are different. Dominant hand is used.</td>
<td><img src="image3" alt="Test 10" /></td>
</tr>
</tbody>
</table>

Table 3a: Definitions of fine motor tests [Westin et al., 2010a].
<table>
<thead>
<tr>
<th>Test#</th>
<th>Description</th>
<th>Test area</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Tapping an active field that changes location when tapped. Four buttons, where one is active (red) and the other three are inactive, are displayed. When an active button is tapped, one (randomly selected) of the other three buttons become active. Cueing is supplied as in test 10 above. Dominant hand is used.</td>
<td><img src="image" alt="Test area" /></td>
</tr>
<tr>
<td>12-14</td>
<td>Tracing a pre-drawn Archimedes spiral from the centre and out using dominant hand without supporting it. The test is repeated three time and patients are instructed that it should be completed in approximately 10s per drawing.</td>
<td><img src="image" alt="Test area" /></td>
</tr>
</tbody>
</table>

*Table 3b: Definitions of fine motor tests [Westin et al., 2010a].*
3 Methodology

3.1 Systems for assessing the fine motor dysfunction (Paper I)
Paper I comprises of a review study of invented systems and methods that detect and assess fine motor dysfunction in patients with movement disorders. These systems use data gathered from touch screen devices. The study discusses requirements for design and implementation of such systems in telemedicine settings. The reviewed patents explain the application of digital signal processing methods in analysis and interpretation of digitized handwriting/drawing data from individuals. The patents that were selected were applicable in home environments and were focused on assessing movement disorders such as PD where the data was collected through touch screens. Some of the keywords used in patent search tools for quantifying the fine motor dysfunction were: “handwriting”, “drawing”, “motor function”, “digitizing tablet”, “fine motor movement” and “Parkinson’s”. These keywords were put in patent search engines, namely, WIPO, EPO and Google Patents.

3.2 Development of the computer-based method (Paper II)

3.2.1 Manual approach
A separate web application was constructed to display spiral drawings and to allow users (PD specialists) to rate observed drawing impairment. The application retrieved paired x and y coordinates of the drawn spiral from Structured Query Language (SQL) database table that store the spiral information. The Graphics Device Interface (GDI+) was used to generate an image and to draw the retrieved pixels on the screen. Along with the spiral drawing image, the drawing completion time in seconds and the patient’s self-assessment of motor state (Table 2b, question 7) at the time of the particular test occasion were retrieved and displayed. The web application was organized in four “tracks”: “preliminary rating”, “training”, “standardised rating”, and “rater agreement”. In all these tracks, users could display spirals, rate them and change previously given ratings.

Two raters studied the examples in the clinical handbook for assessment of tremor in spiral drawings in ten categories [Bain and Findley, 1993]. The same ordinal severity scale (1-3 = mild impairment, 4-6 = moderate, 7-9 = severe and 10 = extremely severe) was applied for rating of drawing impairment in spiral drawings, not necessarily caused by tremor, but as well by other causes, such as dyskinesia or bradykinesia. First a rating of the overall drawing impairment was done and secondly a probable cause for the given impairment was marked. The general impression of the shape of the spirals determined the level of severity;
homogenous and symmetrical spiral shapes were rated as mild drawing impairment, larger deviations from the pattern of spiral shape were rated as moderate impairment and spirals with large interruptions, skewed or incomplete shapes were rated as severe. Drawings without signs of a spiral shape were rated as extremely severe. Within the levels of mild, moderate and severe, there were three scales steps to allow further refinement of the rating. Small and/or few deviations from the expected lines rendered low points on the scale. Completion time could assist in determining the cause of impairment-bradykinesia causes long completion times whereas on-phase choreatic dyskinesia likely is associated with shorter time.

In the “preliminary rating” track, one rater first browsed through spiral drawings and rated at least 10 representative examples of each of the 10 drawing impairment categories. This was done in order to achieve a common basis for assessment among the two raters. The examples were chosen from test occasions where all three spirals had similar degree of drawing impairment.

Then, in the “training” track, both raters observed these preliminary ratings and used them as templates for rating of drawing in other tracks.

In the “standardised rating” track, the web application displayed all three spiral drawings from three randomly selected (by computer program) test occasions per patient. The users were instructed to rate each test occasion based on the global impression of all the three spirals, drawn at the same occasion. Although drawn at the same occasion, sometimes drawings in this group of three appeared too different and could not be given a common rating. These groups were replaced by drawings from other occasions. A ‘standardised manual rating’ score (SMR) was defined as the mean of the two raters’ assessments.

In the “rater agreement” track, single spiral drawings from three randomly selected test occasions (different ones than those in “standardised rating”) from each patient were displayed in random order. These ratings were performed blinded to patient’s ID and to the other user’s rating.

3.2.2 Computer-based approach
A new method, using the discrete wavelet transform and principal component analysis, was developed to process the spiral drawings and generate a score to represent drawing impairments. In this section, a brief overview of the discrete wavelet transform and principal component analysis will be given followed by the methodology explaining the development of the method.

Discrete Wavelet Transform
Signal processing methods based on the Fourier transform are not suitable for analyzing non-stationary or transitory signals [Akay, 1995]. However, the Discrete Wavelet Transform (DWT) addresses this problem where the original sig-
nal is passed through a series of complementary High-Pass (HP) and Low-Pass (LP) filters [Daubechies, 1988]. In this iterative process, the signal is broken down into several frequency sub-bands in order to provide more discriminating features and is thus known as multi-resolution decomposition. The scale of the signal is changed by down sampling operations and the resolution is changed by the filtering operations by using wavelet functions. This procedure is depicted schematically in Figure 1. The original signal is decomposed into approximation (A) and detail (D) coefficients. The approximations represent high-scale, low frequency components and the details represent low-scale, high frequency components of the signal. The approximations are further decomposed in other levels as shown in Figure 1. As a result, the signal is decomposed into the details D1-D3 and one final approximation, A4. Statistics over the set of wavelet coefficients can be computed in order to reduce their dimensionality. In some applications, the mean, variance and ratio of the mean values for adjacent frequency bands were calculated [Kandaswamy et al., 2004, Cvetkovic et al., 2008]. In this work, the dimensionality of the wavelet coefficients was reduced by a multivariate statistical method which is further explained in the section below.

![Figure 1: Three level wavelet decomposition tree; HP is the high-pass filter; LP is the low-pass filter; D is details and A is approximates.](image)

**Principal Component Analysis**

Principal Component Analysis (PCA) is a technique that is widely used for applications such as dimension reduction, lossy data compression, feature extraction and selection, and data visualization [Jolliffe, 2002]. The purpose of the analysis is to take \( n \) variables \( X_1, X_2, ..., X_n \), find combinations of these and transform them into a new set of non-correlated variables \( Z_1, Z_2, ..., Z_n \) called principal components [Chatfield and Collins, 1980]. These components are linear combinations of the original variables and are derived in decreasing order of eigenvalues so that the first principal component accounts for the largest possible variation in the original data. The effect of dimension reduction is achieved when the first two or three components account for most of the variation in the original
data. When any original variables are highly correlated, they measure the “same concept”. In this case, they can be adequately represented by the first two or three components, a situation that helps better understand the data and operate with a smaller number of variables in subsequent analyses. An important step in applying PCA is identifying and retaining the important components that account for a large proportion of the total variance. The most frequently used cut-off criterion for determining the appropriate number of “significant” components is to apply the Kaiser-Guttman rule (i.e. eigenvalue>1) [Jackson, 1993]. Another approach is to select a cumulative percentage of total variation for which it is desired that the selected principal components should contribute more than 70% of the total variation.

The computer method
The steps involved in the development of the computer method are given as follows:

- **Calculate spiral radius** – the spiral data consists of x (horizontal position) and y (vertical position) coordinates. In order to perform quantitative evaluation of the drawing impairment in spirals, the raw coordinates were transformed into polar coordinates. The radius (i.e. the square root of the sum of squares of x and y coordinates) represents the degree to which the spiral drawing deviates from the pre-drawn spiral. It was subsequently used in analysis and processing. Validation of input data is performed so that only those spiral drawings that contain >50 data points are processed.

- **Feature extraction using DWT** – three level decomposition using Daubechies (db10) wavelet family was performed on the signal of the spiral radius to obtain approximations and detail coefficients [Daubechies, 1988]. A feature vector (reconstructed signal), containing 256 features was obtained after the details (Dn, n = 1, 2, 3) at the first, second and third levels (128+64+32 coefficients) and the approximations (A3) at the third level (32 coefficients) were appended to each other in order of descending levels, i.e. A3, D3, D2 and D1 (Figure 2).

- **Dimension reduction using PCA** – since the analysis of high-dimension data sets is often complex and time-consuming, PCA was performed due to the large number of features extracted by DWT. At first, PCA was applied on a subset of data, preselected on the basis of the 10% worst and 10% best tapping results. Tapping results were assessed based on both speed and accuracy (by summation of normalised results). By doing this, the first principal component of the radius was assumed to provide a de-
sired direction in multidimensional feature space well representing PD symptom severity. Secondly, after having both the feature vectors from all the spirals (full dataset) and the obtained PCA coefficient matrix from the subset of extreme tapping results, the score was obtained as the first element after multiplying the DWT feature vector by the coefficient matrix.

- **Scaling of the score** – the score was calibrated using logarithmic and linear transformations to bring it to a roughly linear interval scale between 0 and 10, comparable to the manual rating scale. Linear regression was performed between the manually rated spirals from a “training” track and scaled log values in order to reduce systematic errors, such as over-prediction of low impairment or under-prediction of high impairment. The resulting spiral score is hence denoted “wavelet spiral test score” (WSTS).

**Data analysis**
Test-retest reliability was assessed by Spearman rank correlations. Since there were three spirals per test occasion (A, B, C), the mean of all three possible correlations (AB, AC, BC) was taken. Between-rater agreement comparisons were assessed by Spearman rank correlations. In order to avoid the problem of multiple test occasions per individual, 200 random samples of single test occasions per patient were drawn and mean correlations in this sample was calculated. Bland-Altman analysis of difference [Bland and Altman, 1986] was used to estimate the prediction error of the WSTS versus the SMR. Correlations between the different spiral scores and tapping test results were assessed on test occasion level, first taking mean values of the three spiral drawings, while correlations with UPDRS were assessed after taking mean values over all (approximately 3 × 28) spiral drawn during each weeklong test period.
Figure 2: Example of spiral drawings (left column) and corresponding wavelet coefficients (right column) with drawing impairment rated 1 (first row), 5 (middle row) and 10 (bottom row).
3.3 The web-based system (Paper III)

3.3.1 Calculation of summary scores

In order to analyze the test battery data and obtain test period summary scores from it, time series of self-assessments and motor tests were summarized and processed using statistical and machine learning methods. This procedure is schematically depicted in Figure 3 and further described in the following sections.

Figure 3: Data analysis and processing of test battery time series. In the data summary stage, the raw data was summarized into 28 statistical features. PCA was applied to these features to obtain six conceptual dimensions and least squares regression analysis maps these dimensions to the total UPDRS score. Linear transformations were then used to scale dimensions and predicted overall score to a scale from 0 (worst) to 1 (best) score.

Data Summary

Initially, time series of test battery items were summarized by calculating statistical features, such as the level (MEAN), fluctuation (standard deviation, SD) and the mean squared deviation (MSD) from “the best” answer alternative for test battery questions: question 1, 3, 4 and 6, on a test period level. For instance, MSD for question 1 was calculated using the following formula:

\[
MSD = \frac{\sum_{i=1}^{n}(5 - x_i)^2}{n}
\]

(4.1)

Where \(n\) is the total number of test occasions in a test period, 5 is the best answer alternative meaning “Walking without a problem” and \(x_i\) is the patient’s test response. In total, for each patient and test period, there were obtained 28 features.
The rationale for selection of these statistical features was to define scores taking into account:

- The intensity
- The frequency and
- The importance of occurring symptoms

Mean values are the obvious choice to represent levels and standard deviations are obvious to represent overall variation. The MSD feature was also used aiming at combining both level and variation.

**Calculation of test battery dimensions**

Some of the test battery items are highly correlated indicating that they measure the same concept. Because of this redundancy, it is possible to combine these items and reduce them into a smaller number of principal components without much loss of information. To achieve this, PCA was applied. The information content of a test period with the test battery can be described by six dimensions: walking (based on question 1), satisfaction (question 6), dyskinesia (question 2-dyskinetic and question 4), off (question 2-off and question 3), tapping (all tapping tests) and spiral (all spirals). For each dimension, PCA was performed using the correlation matrix method applied to the statistical features for the questions or motor tests that the dimension was based on by retaining the first principal components. The motivation for using correlations instead of covariances was based on the fact that features might have different scales. Before applying PCA, the statistical features were standardized to have zero means and unit variances to ensure that they all have equal importance in the analysis in terms of variability and also to overcome the scale disparity problem. For example, the scores for the first principal component of the walking dimension are calculated with a linear equation of the type:

\[
Walking = 0.658 \times q_{1_{MSD}} + 0.37 \times q_{1_{SD}} - 0.656 \times q_{1_{MEAN}} \quad (4.2)
\]

where MEAN, SD and MSD are standardized variables and constants are variable weighting factors. This procedure was repeated for the remaining five dimensions. In order each dimension to be represented by a single value, only the first principal components were retained. The total variation (%) accounted for by first principal components of the six dimensions were as following, walking: 71%, satisfied: 77%, dyskinetic: 71%, off: 61%, tapping: 51% and spiral: 69%.

Linear transformations were used to scale first principal component scores of the dimensions to a scale from 0 (worst) to 1 (best) score based on their minimum and maximum values. To evaluate whether dimensions measure the same concept of patient’s health status, Cronbach’s Alpha for dimensions was calculated.
Calculation of overall test score
Using the six derived dimensions, the overall score can be calculated by using a regular hexagon where each dimension becomes a corner on it. The geometrical placement of dimensions in the hexagon was decided based on conceptual relationships between them and in collaboration with domain experts (Figure 4). Using the side-angle-side method the areas of each triangle of the hexagon can be calculated separately and then added up to form an unweighted overall score.

A problem with this approach is that all dimensions, concerning self-assessments and motor tests, have the same weight in the assessment of the overall condition of the patient which is in contradiction with the common clinical rating scales, e.g. UPDRS, used in clinical practice where the highest weight is given to the evaluation of motor symptoms.

To overcome this problem, an alternative way to define the overall score was to weight symptom dimensions using the UPDRS, since it is the most widely used scale for assessing PD today [MDSTFRSPD, 2003]. The standard least-squares regression was used for examining the relationship between first components of dimensions and the patient’s UPDRS score. The weighted overall score (UPW) was defined as a linear combination of the first principal components, with numerical weights estimated by regression technique to fit simultaneous clinical ratings on the UPDRS scale. This was done on the basis of the following equation:

\[
UPW = 0.0265 \times \text{Walking} + 0.0933 \times \text{Satisfaction} + 0.0513 \times \text{Dyskinetic} + 0.0007 \times \text{Off} + 0.119 \times \text{Tapping} + 0.203 \times \text{Spiral}
\]  

(4.3)

The UPW was then scaled to a scale from 0 (worst) to 1 (best) score using linear transformations based on its minimum and maximum values. The dataset used for optimizing this statistical procedure was collected from DAPHNE study. Correlations between UPW and other clinical rating scores were assessed by Spearman rank correlations.

3.3.2 System description
DSSs comprise components for (a) database management capabilities with access to internal and external data, information knowledge, (b) modelling functions accessed by a model management system, and (c) user interface designs that enable interactive queries, reporting and graphing functions [cf. e.g. Shim et al., 2002]. In the developed system, these processes are handled separately where data collection, storage and processing takes places before presentation. The
The former two processes are handled by the so-called data processing sub system and the later one is handled by the web application.

**Data Processing Sub System**
For each test occasion, raw test data are sent from the PDA hand unit over the mobile net to the so-called remote device manager (RDM). The RDM is a commercial product responsible for collecting and storing the data from the hang unit. There is a communication protocol between these two systems which handles the transfer of the data. The data processing sub system (DPSS) is a stand-alone application, which incorporates knowledge to analyze and interpret the raw test battery data. It parses, processes and stores the data into relational database tables and at the same time calculates and stores test scores on test period level.

A connection with RDM is first established followed by receiving and parsing XML data from files. Once files are received, they can be directly interpreted by the DPSS which runs during a specified time interval, e.g. every single hour. The data messages consist of patient identification, test period, hand unit identification, starting and ending time of the test occasion and responses to the test battery items. The main reason for processing data centrally instead of locally is the risk of losing raw data in the remote devices if it is not uploaded regularly. Access to raw data centrally also facilitates future research and method developments. The data collection in hand units is designed to minimize upload bandwidth [Westin et al., 2010a]. For computing the spiral score according to the computer method explained in Paper II a single piece of M-code was written in Matlab® (Mathworks Inc.). An error handler was designed so that the application can recover from possible run-time errors without terminating by rolling back all the information and saving the error information to a log file.

**Web Application**
The web application (WA) is a feedback system comprising a secure web server and a database with web-based access for medical staff. The main role of the WA in the overall system is to present test results to the end-user clients. It is designed in line with users’ requirements and needs for decision support by displaying the test results per patient and test period in graphical and tabular format. Some important user requirements specifications for the WA were: easy to understand and use, user-friendly design, easy navigation, fast response, rapid and convenient screening of patients, a comprehensive overview of a patient’s condition on a single page. They were drafted in collaboration with experts and different prototype versions were developed in several iterations.

Data security is assured at three levels: forms-authentication level; web-server level; and database-server level. Forms authentication is an authentication model in which users, in this case the treating clinical staff and administrators, access
the WA by providing their unique usernames and passwords in a web page form. Depending on the user credentials and their access permissions, the application limits what resources or functionalities are accessible (e.g. it enables access only to data belonging to a specific clinic). On a web-server level, the Integrated Windows Authentication mechanism is used for connecting clients to the server. It uses a hash algorithm for sending credentials before sending them across the network. On a database-level, database administrators have their Active Directory accounts where they have full access to the data and can manage users, groups and passwords, whereas the users of WA that are not part of this domain can access the data by a SQL-specific account, also known as SQL login, incorporated in the configuration file of the WA.

To enable rapid patient status assessment, the information in WA is displayed and ordered using a top-down approach where the general overview of the patient’s performance per test period is given (Figure 4). Raw data and other detailed information may be accessed in more “advanced” displays. After the user logs in and selects a relevant patient, the main page displays the graphs of the patient scores on a test period level, focusing on the overall test score, dimensions and daily summaries. These graphs can be updated interchangeably.

3.3.3 System’s usability evaluations
The potential usefulness of the overall system was evaluated by an advisory board consisting of 14 neurologists after presentation and demonstration of its functionality. They were based in the following countries: USA 3, Germany 3, Italy 2, Spain 1, Netherlands 1, UK 1, Sweden 1, Denmark 1 and Finland 1. They responded to questions by pressing keypads.

The WA was demonstrated to fifteen nurses from the nine clinics in Sweden who had experience using the mobile test battery device in the DAPHNE study, but they had not previously seen the WA. The evaluation was performed as a presentation session for the nurses where they asked the presenter to display certain data from a particular patient in whom they had an interest. The presenter took notes about the reactions from the group. At least one patient per clinic was shown.

One year later, the IBM Computer System Usability Questionnaire (CSUQ) was administered to evaluate the nurses’ satisfaction with the WA. They were asked to perform a series of tasks using the WA, such as (a) login, (b) select a patient, (c) check the patient’s performance by looking at the graphs of summarized scores and compare these results with own clinical observations, (d) select and check the other patients’ results repeating steps (b) and (c), (e) complete the survey and (f) logout. The questionnaire was web-based and more than one person per clinic involved in the study could check and update the answers to the questionnaire giving a consensus response. CSUQ consists of 19 items on a
seven-point Likert [Likert, 1932] rating scale ranging from 1 (strongly agree) to 7 (strongly disagree). Four usability scores were derived after responses to the CSUQ items were averaged: Overall Satisfaction, System Usefulness, Information Quality and Interface Quality [Lewis, 1995]. Data about gender, age and previous experience with computer applications were also noted.

3.3.4 System requirements
DPSS was written in the VB.NET programming language to allow data parsing and processing. It is designed to run with a PC-compatible machine with a minimum requirement of a Pentium processor with 256 of RAM. The M-function to calculate the spiral score was encrypted and wrapped into a C# interface by using the Matlab Builder for .NET to be accessed by the DPSS. This pre-built interface requires the Matlab Compiler Runtime to be installed in the running machine.

The WA was developed in ASP.NET as a code-behind model and ADO.NET was used as a standard way to connect to a database. Structured Query Language (SQL) was employed as a query language to access the data stored in Microsoft SQL Server. Three-tier architectures ensure a high-level of availability where the different application components can be easily replicated to increase the overall performance [Casteleyn et al., 2009]. WA supports this architecture in which a web browser sends HTML requests using HTTP to a web server, which in turn passes the request to a common gateway interface application program. This application server then sends requests to a database server, which generates the

Figure 4: Patient status report in WA with graphical visualization of time series of test measures.
query result set and sends it back in order to be formatted and presented to the end-user clients as an HTML page.

Microsoft Web Application Stress Tool (WAST) was used to stress the web server by realistically simulating a large number of users accessing the WA at the same time. The main measurements include response time and throughput [Meier et al., 2004]. The tests were run from a separate machine, running at 100 Mbits/sec with varying load levels (concurrent connections) of 1, 10, 50, and 100. The web server, IIS 6.0, runs on Windows Server 2003 with CPU speed of 2.93 GHz and 1 GB of RAM. A script was developed, capturing a test scenario with a typical web browser. The test variability of the metrics was measured by repeating each test 10 times per load level of 1 minute run time. The results showed that as the load level increased from 50 to 100, the number of requests that can be successfully served by the application per unit time starts to saturate. The average response time varied from page type to page type. For example, the login page introduced more delay compared to other pages; this was because it used the POST method to send data to the server. All average request times were in the range of 0.5 seconds indicating the users will not wait for too long for each page. However, memory requirements on the web server side tend to increase over time as the number of users accessing the database table storing spiral drawings data (x, y and time) increases. This is because this table contains a large number of records (around 8 million in the current dataset).
4 Results

4.1 Systems for assessing the fine motor dysfunction (Paper I)
From the review of systems reported in Paper I, it was found that application of time and frequency domain signal processing methods is useful in digitized analysis of handwriting/drawing movement data. Furthermore, measures concerning forces, accelerations and radial displacements were the most important ones in detecting fine movement anomalies. The study showed that for a system to be applicable in a patient’s home environment, the choice of the data collection schemes and signal processing approaches would be crucial. The study also suggested that event-based data acquisition should be employed in combination with off-line non-stationary signal processing.

4.2 The computer-based method (Paper II)
Test–retest reliability coefficients were as follows: 0.77 for WSTS, 0.71 for first rater and 0.70 for second rater. In general, correlations (absolute values) between the WSTS and tapping results were low to moderate (Table 4). Correlations within the tapping results were 0.58 or lower.

The correlations between WSTS and UPDRS scores were as follows: 0.41 to total UPDRS, 0.38 to UPDRS subsection III (motor exam) and 0.51 to UPDRS subsection II (ADL). The correlation between WSTS and SMR for the spiral drawings was (0.89) strong. The 95% confidence interval for the prediction error for the WSTS was ±1.5 and the mean value (bias) was 0.39 scale units. Rater agreements were as follows: 0.87 between raters, 0.91 between WSTS and mean rating.

<table>
<thead>
<tr>
<th>SMR</th>
<th>0.89</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total UPDRS</td>
<td>0.41</td>
</tr>
<tr>
<td>Average speed (Table 3a, tests 8 and 9)</td>
<td>-0.40</td>
</tr>
<tr>
<td>Average accuracy (Table 3a, tests 8 and 9)</td>
<td>-0.45</td>
</tr>
<tr>
<td>Speed (Table 3a, test 10)</td>
<td>-0.56</td>
</tr>
<tr>
<td>Speed (Table 3b, test 11)</td>
<td>-0.52</td>
</tr>
<tr>
<td>Accuracy (Table 3b, test 11)</td>
<td>-0.50</td>
</tr>
</tbody>
</table>

Table 4: Mean Spearman correlations after repeated random sampling of one occasion per patient.
4.3 The web-based system (Paper III)

The Cronbach’s Alpha test showed that there was a good internal consistency between dimensions with a coefficient of 0.81. The correlations between UPW and total UPDRS (-0.6, p<0.001) were adequate whereas between UPW and Hoehn and Yahr (-0.44, p<0.001) were medium.

Eleven neurologists had a positive impression regarding the overall impression of the integrated system, one had a neutral impression and two had negative impression. The most important benefits they could see were an increased ability to identify patients who are not doing well and facilitated follow-up optimization of an individual’s treatment. The system was seen as most important for complicated patients and for regional patients, that is patients living in regions far away from a clinic. The general conclusion of the board was that the system was recognized as a tool that will assist in management of patients.

At the presentation session, the responses from nurses are summarized in a qualitative manner as follows: (a) the WA is very useful, (b) the results during test periods showed agreement with qualitative observations of the patient during that test period, for example, “one patient was in a bad condition in baseline, he improved after starting Duodopa, then he became worse again, 24-hour infusion started and the patient became better again; we can clearly follow this in the web application”, (c) comparisons between patients are possible (one patient is in better/worse condition than another).

Responses to the CSUQ were obtained from seven of the nine clinics and the results were mixed. A majority of the clinics were quite satisfied with the usability although a sizeable minority were not. All evaluators were female and ages ranged from 38 to 61 (mean value 49). Two out of seven asserted that they had much previous experience with computers, whereas four had some experience and one had little experience.
5 Discussion

5.1 Assessment of fine motor impairment

Quantitative assessment of fine motor impairment has been tried using tracking devices [Hacisalihzade et al., 1986], sensing technology [Pradhan et al., 2010], scanning devices [Kraus and Hoffmann, 2010], and computer-interfaced musical keyboards [Taylor Tavares et al., 2005]. Elble et al. [1996] compared the feasibility of using touch screen devices and accelerometers in objectively assessing tremor during handwriting and drawing. It was concluded that touch screen devices had several advantages over accelerometers and they could be considered as useful and reliable inexpensive tools for clinical studies.

The review reported in Paper I found that invented systems and methods using touch screens as inputs were able to digitally analyze the handwriting/drawing movement data. It was also found that no methods were able to differentiate between various motor states. In other words, there was a lack of methods that can discriminate between voluntary and involuntary movements such as tremors and drug-induced dyskinesias as well as between abstention from voluntary movements and decreased movement ability also known as hypokinesia (as a result of insufficient medication).

The WSTS method presented in Paper II automatically assessed the drawing impairment in traced spirals. One limitation was that it over-predicted small impairments and under-predicted moderate impairments, compared to SMR. Therefore, further optimization and transformation of the data is required. The method could not exactly classify the drawing impairment in spiral drawings into the correct category on the scale 1-10. Using an ordinal scale with fewer categories would improve the classification performance of the method.

Another weakness of the method is that it only detects disability but lacks the mechanism to distinguish between specific symptoms causing the particular disability. This problem will be addressed in future research work.

In the study performed by Wang et al. [2008], it was found that the detection of the optimal spiral center improves the classification rate for spiral analysis during free drawn (no templates to trace) spirals. A new approach for analyzing the spiral data and estimating its varied origin in the polar coordinate system has been recently proposed by Wang et al. [2011]. However, in the case of the WSTS method, the data is gathered by tracing a pre-drawn spiral template and its scoring mechanism is dependent on the screen resolution of the handheld PDA device. Further development of the WSTS method is needed so that it would be applicable in cases of spiral center variation and different screen resolutions.
In summary, the WSTS method can be considered as a simple, feasible, and useful tool in clinical trials and/or patient home environments for evaluating and assessing the degree of PD-related drawing impairments.

5.2 PD Monitoring

Most of the existing approaches to remote monitoring of PD symptoms have been based on the use of wearable sensors. Typically, these approaches do not address the important issue of defining scores that combine different aspects of patient function in order to facilitate monitoring of the patients. In evaluating treatments and motor performance, both the physician and patient-oriented outcomes offer complementary information [Hobart et al., 1996, Gijbels et al., 2010]. The mobile device test battery used for data collection in this thesis emphasized the need for repeated measurements by taking into account both the patient’s perception of common symptoms and also their ability to perform motor tests [Westin et al., 2010a].

The web-based system presented in Paper III compiled the test battery raw data, concerning subjective and objective measurements, into summary scores for conceptual symptom dimensions and an overall test score per patient and test period. From Eq. (4.3), motor test results which consist of tapping tests and spirals were given the highest weight of 65%, just as motor section (Part III) has the highest weight in UPDRS.

When using a multivariate analysis for dimension reduction many criteria are used for determining the adequate number of principal components to be retained and subsequently used in model development. In this work, the justification of retaining only the first principal components for each dimension was based on the fact that a single value was needed to represent them. The total variation (%) accounted for by first principal components of the six dimensions were as following, walking: 71%, satisfied: 77%, dyskinetic: 71%, off: 61%, tapping: 51% and spiral: 69%.

The summary scores have been validated in a separate study (in Italy, 30 patients), with the objective of assessing test-retest reliability, correlations to other assessment methods, and the ability to detect differences between patient groups at different stages [Westin et al., 2011]. All equations and constants for calculation of the summary scores were defined in Paper III and clinimetric properties of the test battery’s overall score were determined based on new data from the Italian study. The compliance and reliability of the test battery was good. Correlations between the overall test score to the rating scales were adequate. Moreover, differences in overall test score and conceptual dimensions between stable and fluctuating patient groups were detected.

When developing and applying web-based systems for telemedicine applications several issues should be taken into consideration such as (a) architectural
and technical, (b) security, privacy and confidentiality, and (c) usability and user acceptance [Bellazzi et al., 2001]. The system presented in Paper III is at a prototype stage and its architecture and technology allows data collection, retrieving and visualization. The WA relies on password-based user authentication but lacks data encryption. However, the current usability evaluation is missing feedback from PD physicians and this limitation will be taken into account in future usability tests for the final version of the system.

The main advantage of the system is that it allows easy access to relevant symptom information from PD patients’ home environment.

5.3 Future work
To analyze the potential benefits of using the system, presented in Paper III, in everyday clinical practice, a “market acceptance test” is planned (in the framework of the E-MOTIONS project) in at least three countries. The goal of the test will be to determine whether compliance and usability in everyday clinical practice is similar to that in clinical studies. Another question that will be addressed is whether there are country-specific issues posing different sets of requirements for its introduction into clinical practice. Other important issues that would be addressed are whether the usage of the system would improve treatment of patients during clinical practice and whether a cost benefit analysis is of any advantage.

Ongoing research work includes the development of a computer-based method for assessing the cause of drawing impairment in traced spirals that were collected during the DAPHNE study. The idea of the method is to be able to categorize spirals in two different classes for being drawn in the “off” state and the dyskinetic state for every test occasion. The development will include calculating various geometrical entities from spirals and using them as inputs to non-stationary signal processing methods, e.g. DWT and Hilbert-Huang Transform [Xie and Wang, 2006], for extracting features at different frequencies. The evaluation of the method will be done against clinical ratings.

Another, more specific issue that will be addressed in the future is employing data fusion techniques on test battery items to generate a test occasion score equivalent to the test period overall score.

Future research will also focus on designing a cognitive test battery for repeated use at home that would provide basic information on the cognitive status of test persons on test occasions. The development and evaluation of the system will be done in collaboration with psychology researchers.
6 Conclusions

The result of the review presented in Paper I suggests that event-based data collection and signal processing techniques suitable for non-stationary signals would be useful in systems for patient-home monitoring. This offers tolerance to variations in speed during unsupervised handwriting/drawing tasks and less post-processing of data. The review reveals that measures concerning forces, accelerations and radial displacements were the most relevant ones in detecting fine movement anomalies.

The WSTS method presented in Paper II could assess PD-related drawing impairments well comparable to trained raters. The WSTS was applicable during event driven sampling, in the way that data were usable even if spirals were drawn slowly, which is often the case in advanced PD. The WSTS resulted in a strong correlation with the SMR and a medium correlation with the UPDRS scores. Its approach for classifying the impairment in spirals through multi-resolution representation of the radius and dimension reduction with a selected subset of data is a novel and powerful one.

The web-based system presented in Paper III was designed to provide PD patient status evaluation through convenient access to a patient’s current symptom profile and history. The system processed and summarized various time series of self-assessments and motor tests into summary scores for six conceptual symptom dimensions and an overall test score, UPW. The UPW reflected the overall condition of the patient during week-long test periods. The method to take the first principal component of mean, SD and MSD for each feature, is a data-driven general way to automatically obtain weights for combining level and variation. The internal consistency between the six conceptual symptom dimensions was good. Correlations between the UPW and total UPDRS were adequate and significant. The derived UPW may be helpful in (a) facilitating the process of screening patients, (b) avoiding sub-optimization of treatments, and (c) deciding if a treatment change leads to an improvement of a patient’s general condition or not.
Acknowledgements

After finally approaching the end of the degree of Licentiate of Technology, I would like to express my gratefulness to a number of individuals and institutions who contributed to the success of the thesis in a variety of different ways. These individuals are my supervisors Mark Dougherty, Silvia Coradeschi, Peter Funk and Jerker Westin. However, a special thanks goes to Jerker Westin who, being my supervisor, was responsible for creating the fantastic research environment within which I was accepted and thoroughly supported throughout my time at the Computer Engineering department at Dalarna University.

It would certainly be an injustice to neglect Torgny Groth and Dag Nyholm from Uppsala University with whom I shared co-authorship of the two research papers that are included in this thesis work. Within this context, Torgny and Dag’s comments regarding methodology and results were valuable due to their decision support system and neurological (Parkinson’s) backgrounds respectively.

On the industry side, Stefan Åsberg from Abbott and Ulf Bergqvist from Nordforce facilitated important aspects such as data collection and clinical studies.

Also, my appreciation goes to Taha Khan for the almost five years we have spent together at Dalarna University, first studying and now working together as colleagues sharing office, as well as to Fawad Syed for proof-reading the thesis.

The institutions that need to be appreciated are the Swedish Knowledge Foundation (KK-stiftelsen), Abbott Product Operations AG, Nordforce Technology AB, and Dalarna University for the financial support they have extended within the framework of the E-MOTIONS project. I also thank the Teknikdalen Foundation in Borlänge for providing me the scholarship for the best degree project.

Last but not the least, my deepest gratitude goes to my wife, Gzime, for her love and constant support. You made this possible!

Borlänge, Mevludin Memedi
November 2011
Bibliography


