Mobile systems for monitoring Parkinson’s disease
To my parents
(Dedikuar prindërve të mi)
Mobile systems for monitoring Parkinson’s disease
Abstract

A challenge for the clinical management of Parkinson's disease (PD) is the large within- and between-patient variability in symptom profiles as well as the emergence of motor complications which represent a significant source of disability in patients. This thesis deals with the development and evaluation of methods and systems for supporting the management of PD by using repeated measures, consisting of subjective assessments of symptoms and objective assessments of motor function through fine motor tests (spirography and tapping), collected by means of a telemetry touch screen device.

One aim of the thesis was to develop methods for objective quantification and analysis of the severity of motor impairments being represented in spiral drawings and tapping results. This was accomplished by first quantifying the digitized movement data with time series analysis and then using them in data-driven modelling for automating the process of assessment of symptom severity. The objective measures were then analysed with respect to subjective assessments of motor conditions. Another aim was to develop a method for providing comparable information content as clinical rating scales by combining subjective and objective measures into composite scores, using time series analysis and data-driven methods. The scores represent six symptom dimensions and an overall test score for reflecting the global health condition of the patient. In addition, the thesis presents the development of a web-based system for providing a visual representation of symptoms over time allowing clinicians to remotely monitor the symptom profiles of their patients. The quality of the methods was assessed by reporting different metrics of validity, reliability and sensitivity to treatment interventions and natural PD progression over time.

Results from two studies demonstrated that the methods developed for the fine motor tests had good metrics indicating that they are appropriate to quantitatively and objectively assess the severity of motor impairments of PD patients. The fine motor tests captured different symptoms; spiral drawing impairment and tapping accuracy related to dyskinesias (involuntary movements) whereas tapping speed related to bradykinesia (slowness of movements). A longitudinal data analysis indicated that the six symptom dimensions and the overall test score contained important elements of information of the clinical scales and can be used to measure effects of PD treatment interventions and disease progression. A usability evaluation of the web-based system showed that the information presented in the system was comparable to qualitative clinical observations and the system was recognized as a tool that will assist in the management of patients.

Keywords: automatic assessments, data visualization, data-driven modelling, home assessments, information technology, mobile computing, objective measures, Parkinson’s disease, quantitative assessments, remote monitoring, spirography, symptom severity, tapping tests, telemedicine, telemetry, time series analysis, web technology.
Acknowledgements

This work was carried out at the department of Computer Engineering, School of Technology and Business Studies, Dalarna University, Sweden.

Swedish Knowledge Foundation, Abbott Product Operations AG (now AbbVie), Nordforce Technology AB and Animech AB are gratefully acknowledged for the financial support they have extended within the frameworks of the E-MOTIONS and PAULINA projects.

I would like to express my gratitude to people who contributed to the success of the thesis in a variety of different ways and encouraged me during my time as a PhD student.

I would like to start by thanking my four supervisors: Mark Dougherty, Silvia Coradeschi, Peter Funk and Jerker Westin. Mark thank you for your support, constructive comments and guidance as well as for your sense of humour throughout this fruitful and exciting journey. Silvia and Peter thank you for your constructive comments and for your support. Jerker thank you for your patience, flexibility, caring and for placing your trust and confidence in my professional abilities. You created a fantastic research environment within which I was accepted and thoroughly supported throughout my time as a student. You were not only my formal supervisor, but my friend too. All my dear supervisors, I appreciate your great scientific suggestions which taught me to be grown up as an independent researcher that I am today.

My co-authors: Torgny Groth, Anders Johansson, Peter Grenholm, Samira Ghiamati, Dag Nyholm, Sven Pålhagen, Taha Khan, Thomas Willows, Håkan Widner and Jan Linder thank you for all your inputs and efforts. A special thank you to: Torgny for your contributions to Paper I and Paper V such as methodology and design of the method and web-based system, Dag for your contributions to all papers such as study designs, results interpretation from neurological perspective and helping in revising the papers, Taha for your contributions to Paper II regarding data processing and thank you for the fun time spent together both as students and colleagues sharing office and good luck on your upcoming dissertation.

Stefan Åsberg from Abbott/AbbVie and Ulf Bergqvist from Nordforce, thank you both for your help regarding clinical studies and data collection.

Lars Rönnegård, thank you for reviewing the first version of the thesis. I also thank anonymous reviewers of the published papers for their comments and acknowledge that they have helped me significantly in improving the papers.
I would like to express my gratitude to Teknikdalen Foundation in Borlänge for providing me the scholarship for the best degree project.

Finally, to my family I dedicate this thesis. I would like to give my deepest gratitude to my parents in the language they understand that is Albanian: “Prindër të dashur! Fjalët janë të pakta për ta shprehur falemenderimin për gjithë mbështetjen dhe sakrificat e juaja që nga hapat e para të jetës time. Ju më ndihmuat që të realizoj ëndrrat e mi ja që nga fëmijëria, njëra nga ato ishte edhe dëshira për tu bërë doktor shkence dhe ja që më përkrahjen tuaj ia arrita. Të gjitha sukseset e mi ju takojn juve të dashurit e mi. Uroj që Zoti të ju jep shëndet dhe jetë të gjatë dhe shpresoj se në të ardhmen do të kemi mundësi të kalojm më shumë kohë së bashku.”

Last but not the least, I thank you my wonderful wife, Gzime, for your love and constant support. During this journey, there were very difficult moments for me but you were always there to stand by my side and cheering me up whenever I felt down. My son, Anis, I owe you so many hours of fun. You have firmly planted yourself in my heart. I cannot imagine my life without you.

Thank you (Faleminderit)!
December 2013
Borlänge, Sweden
Included papers

This thesis is based on the following papers, referred to by Roman numerals in the text.


Reprints for Papers I – V were made with permission from the respective publishers.
My contributions to the papers were as follows:

**Paper I** – development of the framework for collecting visual ratings, partly involved in method development, writing parts of the manuscript and reviewing the rest.

**Paper II** – development of the framework for collecting visual ratings, method development, data analysis, results interpretation, writing the first version of the manuscript and revising it.

**Paper III** – data analysis, results interpretation, writing the first version of the manuscript and revising it.

**Paper IV** – planning the literature review, conducting the review, writing the first version of the manuscript and revising it.

**Paper V** – method development, development of the custom software, data analysis, results interpretation, writing the first version of the manuscript and revising it.

**Paper VI** – data analysis, results interpretation and writing the first version of the manuscript.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Approximations coefficients</td>
</tr>
<tr>
<td>A-ACCURACY</td>
<td>Automated Accuracy score</td>
</tr>
<tr>
<td>A-ARRHYTHMIA</td>
<td>Automated Arrhythmia score</td>
</tr>
<tr>
<td>ADL</td>
<td>Activities of Daily Living</td>
</tr>
<tr>
<td>A-FATIGUE</td>
<td>Automated Fatigue score</td>
</tr>
<tr>
<td>A-GTS</td>
<td>Automated Global Tapping Severity score</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ApEn</td>
<td>Approximate Entropy</td>
</tr>
<tr>
<td>A-SPEED</td>
<td>Automated Speed score</td>
</tr>
<tr>
<td>ATA</td>
<td>American Telemedicine Association</td>
</tr>
<tr>
<td>ATP</td>
<td>Alternating Tapping Performance</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the receiving operating characteristics Curve</td>
</tr>
<tr>
<td>CDSS</td>
<td>Clinical Decision Supports Systems</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>Cross-ApEn</td>
<td>Cross Approximate Entropy</td>
</tr>
<tr>
<td>CSUQ</td>
<td>Computer System Usability Questionnaire</td>
</tr>
<tr>
<td>CV</td>
<td>Coefficient of Variation</td>
</tr>
<tr>
<td>D</td>
<td>Details coefficients</td>
</tr>
<tr>
<td>DDM</td>
<td>Data-Driven Modelling</td>
</tr>
<tr>
<td>DPSS</td>
<td>Data Processing Sub System</td>
</tr>
<tr>
<td>DTW</td>
<td>Dynamic Time Warping</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transform</td>
</tr>
<tr>
<td>GTS</td>
<td>Global Tapping Severity score</td>
</tr>
<tr>
<td>HE</td>
<td>Healthy Elderly</td>
</tr>
<tr>
<td>HP</td>
<td>High-Pass filter</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra-Class Correlation coefficient</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>LCIG</td>
<td>Levodopa-Carbidopa Intestinal Gel</td>
</tr>
<tr>
<td>LME</td>
<td>Linear Mixed-Effects models</td>
</tr>
<tr>
<td>LP</td>
<td>Low-Pass filter</td>
</tr>
<tr>
<td>LR</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td>MEAN</td>
<td>Mean value</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple Linear Regression</td>
</tr>
<tr>
<td>MRA</td>
<td>Multi-Resolution Analysis</td>
</tr>
<tr>
<td>MSD</td>
<td>Mean Squared Deviation</td>
</tr>
<tr>
<td>MTS</td>
<td>Mean Tapping Speed</td>
</tr>
<tr>
<td>MTSPC</td>
<td>Mean Tapping Speed Per Cycle</td>
</tr>
<tr>
<td>OTS</td>
<td>Overall Test Score</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</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>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PDQ-39</td>
<td>Parkinson’s Disease Questionnaire 39-item</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of Life</td>
</tr>
<tr>
<td>RDM</td>
<td>Remote Device Manager</td>
</tr>
<tr>
<td>SA</td>
<td>Self Assessed</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation value</td>
</tr>
<tr>
<td>SMR</td>
<td>Standardized Manual Rating</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>TOSS</td>
<td>Test Occasion Spiral Score</td>
</tr>
<tr>
<td>UPDRS</td>
<td>Unified Parkinson’s Disease Rating Scale</td>
</tr>
<tr>
<td>WA</td>
<td>Web Application</td>
</tr>
<tr>
<td>V-ACCURACY</td>
<td>Visually-assessed Accuracy score</td>
</tr>
<tr>
<td>V-ARRHYTHMIA</td>
<td>Visually-assessed Arrhythmia score</td>
</tr>
<tr>
<td>V-FATIGUE</td>
<td>Visually-assessed Fatigue score</td>
</tr>
<tr>
<td>V-GTS</td>
<td>Visually-assessed GTS score</td>
</tr>
<tr>
<td>V-SPEED</td>
<td>Visually-assessed Speed score</td>
</tr>
<tr>
<td>WSTS</td>
<td>Wavelet Spiral Test Score</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
# Table of contents

1 INTRODUCTION .......................................................................................... 15  
1.1 Motivation ............................................................................................ 15  
1.2 Research questions ................................................................................ 17  
1.2.1 Quantification and analysis of fine motor performance ............... 17  
1.2.2 Methods and systems for remote and long-term assessment of  
    symptoms ............................................................................................... 18  
1.3 Research approach ................................................................................ 19  
1.4 Thesis outline ....................................................................................... 22  

2 BACKGROUND ............................................................................................ 24  
2.1 Overview of applied IT in healthcare .................................................. 24  
2.1.1 Telemedicine .................................................................................. 25  
2.1.2 Mobile computing technology ....................................................... 26  
2.1.3 Information processing ................................................................... 26  
2.1.4 Evaluation of IT-based systems ..................................................... 28  
2.2 Parkinson’s disease ............................................................................ 29  
2.2.1 Clinical features ............................................................................ 29  
2.2.2 Treatment ...................................................................................... 29  
2.2.3 Symptom assessment in clinical settings ....................................... 30  
2.3 Related work ....................................................................................... 31  
2.3.1 Quantification of fine motor performance .................................... 31  
2.3.2 Systems and methods for monitoring PD symptoms ................. 33  
2.4 Subjects and data ............................................................................... 34  
2.4.1 Subjects ....................................................................................... 34  
2.4.2 Symptom data collection via a telemetry device ......................... 35  

3 METHODS .................................................................................................. 37  
3.1 Discrete Wavelet Transform ............................................................... 37  
3.2 Principal Component Analysis ............................................................ 39  
3.3 Approximate Entropy .......................................................................... 40  
3.4 Dynamic Time Warping ..................................................................... 41  
3.5 Multiple Linear Regression ................................................................ 42  
3.6 Logistic Regression ............................................................................. 42  
3.7 Mixed-Effects Models ......................................................................... 43  

4 QUANTIFICATION AND ANALYSIS OF FINE MOTOR  
PERFORMANCE ............................................................................................ 45  
4.1 Paper I – A new computer method for assessing drawing  
    impairment in Parkinson’s disease ....................................................... 45  
4.2 Paper II – Automatic and objective assessment of alternating tapping  
    performance in Parkinson’s disease .................................................... 47  
4.3 Paper III – Spiral drawing during self-rated dyskinesia is more  
    impaired than during self-rated off .................................................... 51
5 METHODS AND SYSTEMS FOR REMOTE AND LONG-TERM ASSESSMENT OF SYMPTOMS ............................................................. 53
5.1 Paper IV – Combined fine-motor tests and self-assessments for remote detection of motor fluctuations .................................................... 53
5.2 Paper V – A web application for follow-up of results from a mobile device test battery for Parkinson’s disease patients.............................. 53
5.3 Paper VI – Self-assessments and motor tests via telemetry in a 36-month levodopa-carbidopa intestinal gel infusion trial ....................... 56

6 RESULTS .............................................................................................. 58
6.1 Paper I ................................................................................................ 58
6.2 Paper II ............................................................................................... 58
6.3 Paper III ............................................................................................. 59
6.4 Paper IV ............................................................................................. 61
6.5 Paper V ............................................................................................. 62
6.6 Paper VI ............................................................................................. 64

7 CONCLUSIONS ................................................................................... 67
7.1 Summary and Discussion ................................................................... 67
7.2 General implications .......................................................................... 72
7.3 Limitations ......................................................................................... 72
7.4 Future prospects ................................................................................ 74
7.5 Concluding remarks .......................................................................... 75

REFERENCES: ........................................................................................ 76
1 Introduction

1.1 Motivation

Measuring symptoms and treatment-related complications in advanced Parkinson’s disease (PD) is complex and challenging. This complexity is highly associated with the significant between- and within-patient variability in the manifestation of symptoms as well as with the emergence of motor fluctuations as a result of chronic treatment.

In a clinical setting today, the state of the art is to use clinical rating scales such as the Unified Parkinson’s Disease Rating Scale (UPDRS) (Fahn et al., 1987) and the 39-item PD Questionnaire (PDQ-39) (Jenkinson et al., 1995), which are mainly based on observations and judgments by clinicians. During the evaluation of symptoms and treatments, both clinician- and patient-oriented outcomes offer complementary information (Chrischilles et al., 1998). Patient paper diaries targeting self-assessments are usually used to support the clinical evaluation in the patients’ home environment. Patients record the time they spend in ‘Off’ (a motor state in which PD symptoms reappear as a result of insufficient levels of medication), in ‘On’ (in which medication levels are sufficient for good motor symptom control) and in ‘On with dyskinesias’ (the appearance of hyperkinetic movements related to excessive levels of medication).

However, the use of these rating scales is not suitable for long-term, repeated and remote follow-up of the symptoms since they are relatively time consuming (Martinez-Martin et al., 1994), may need to be filled out at a clinical visit, require considerable clinical experience (Taylor Tavares et al., 2005) and some of their items have poor inter-clinician reliability (MDSTFRSPD, 2003; Hagell et al., 2003). Furthermore, the clinical visit may not accurately represent the patients’ activities in their home environment and may influence patient outcomes (Stocchi et al., 1986). Patient diaries capture symptom fluctuations better, but even these are often not filled out the correct time (Stone et al., 2003). 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). Since the use of clinical scales provides only a snapshot of symptom severity during the clinical visit, repeated measurements are useful in revealing the full extent of the patient’s condition and avoiding bias while measuring the effects of treatment (Isacson et al., 2008).

Therefore, there is a need to combine the clinical scales with frequent subjective and objective, observer-independent measures before and after a treatment intervention in order to cover more aspects of the outcome than...
what can be achieved by utilizing the established clinical scales alone. In contrast to clinical scales for the assessment and follow-up of symptoms, Information Technology (IT)-based systems provide a means for remote, long-term and repeated symptom assessments. Additionally, these systems have better resolution than traditional clinical approaches thus providing more valid data which can be processed and potentially improve the accessibility and efficiency of care as well as increase patient compliance (Goetz et al., 2009). They are able to more accurately capture subtle symptom fluctuations, which is imperative when evaluating treatment interventions. In addition, the introduction of IT-based systems for the remote monitoring of symptoms may also help in reducing hospitalization costs as well as in overcoming barriers to patient participation in clinical studies such as frequent clinical visits, mobility impairment and the need to travel (Baig and Gholamhosseini, 2013).

This thesis addresses an issue of fundamental importance to remotely monitoring the severity of PD symptoms by using IT. It does so in the context of the use of telemetry assessments of subjective (patient-based assessments of symptoms) and objective measures of fine motor function (tapping and spirometry) to address the development and evaluation of computer-based methods for scoring symptoms in an objective, quantitative and automatic manner. The aim of the thesis is two-fold. First, it aims at investigating the use of methods for measuring the severity of symptoms being represented in fine motor tests and analysing the severity of these objective measures in relation to a patient’s subjective measures. Second, it aims at developing and validating a method for combining the subjective and objective measures into composite measures which provide a means for the follow-up of the severity of different symptoms and a more in-depth assessment of the patient’s general health. As part of the second aim, the thesis aims to develop custom software and web-based applications to support clinicians in treating their patients by providing them with easy access to relevant symptom information in a visual and an objective manner. Different metrics such as user satisfaction with the IT-based system, validity, reliability, sensitivity to treatment interventions and the natural progression of PD for its derived computed measures were assessed.

The work reported in this thesis was performed in the framework of two research projects: “Evaluation of a Motor/Non-Motor Test Intelligent Online System” (E-MOTIONS) and “Home assessment of Parkinson’s disease symptoms” (PAULINA) during the period of 2010-2015. The overall aim of the projects was to apply IT for remote data collection, data processing and the presentation of symptom status for advanced PD patients. This thesis mainly focuses on methods and systems for data pro-
cessing and data presentation. However, the term IT-based system is often used in this thesis to refer to the whole technology comprising components for collection, processing and presentation of the data. The data used for development and evaluation of the methods consisted of repeated measures of subjective and objective health indicators at different times spanning a period of a week, using a wireless telemetry test battery implemented on a touch screen device (Westin et al., 2010a).

1.2 Research questions

1.2.1 Quantification and analysis of fine motor performance

The ability to perform functional upper limb motor tasks is essential for most of activities of daily living (ADL). Fine motor control can be defined as the ability to perform small and precise movements requiring hand-eye coordination. Patients diagnosed with PD often have difficulties with timing control and coordination of upper limb movements (Almeida et al., 2002; Yahalom et al., 2004). 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. The development of these impairments is associated with the progression of the disease and can eventually reduce the patients’ overall Quality of Life (QoL). The most common procedure to assess the severity of fine motor symptoms is through clinical rating scales such as the UPDRS motor disability (part III). Given the fact that it is not feasible to use these scales for long-term and repeated assessments of symptoms, the nature and level of the fine motor impairment can be measured by computer-based analysis of digitized movement data to characterize kinematic and dynamic performance. The focus of the thesis is the quantification of fine motor performance using repeated measures data gathered through alternating tapping tests and spirography.

Based on the above mentioned issues, the following research questions were addressed in this thesis:

- **RQ 1**: How can we develop methods to quantitatively and objectively measure the severity of PD-related impairments during fine motor tests (alternating tapping tests and spiral drawing)?
- **RQ 2**: How do measures of fine motor function relate to patient-based assessments of motor conditions (On, Off and dyskinesia)?
1.2.2 Methods and systems for remote and long-term assessment of symptoms

During the process of evaluation of symptoms and treatments, both the clinician- and patient-oriented outcomes offer complementary information (Hobart et al., 1996; Chrischilles et al., 1998; Gijbels et al., 2010). However, PD patients have difficulties in assessing their disability in relation to assessments of daily function (Shulman et al, 2006) and executive functions (Koerts et al, 2011), as well as difficulties in recognizing their treatment-related motor complications (Vitale et al., 2001). These difficulties regarding the self-assessment of their perceived state of health may be influenced more by a patient’s mental health symptoms than physical symptoms (Chrischilles et al., 2002). In clinical settings, a treatment would be considered to have a positive clinical effect if it simultaneously improved both the motor and non-motor functions of the patient. In the case of longitudinal observational studies, motor performance may improve over time by learning, even if the actual physiological status is unchanged, whereas patient-based assessments may be affected by changed expectations, for instance at the beginning of a new treatment. Knowledge of differences between these two types of information allows a reliable assessment of the degree of a patient’s disability. Additionally, combining subjective and objective measures provides more data for analysis to identify the above mentioned problems during longitudinal studies as well as provide input for cross-evaluation. Currently, clinical rating scales such as UPDRS are designed in a way so that different aspects of PD are addressed, by gathering evaluations through patient-administered questionnaires on ADL and clinician-derived assessments on motor performance. In addition, telemedicine approaches to remote monitoring of PD symptoms include e-diaries (Papapetropoulos et al., 2012), wearable inertia sensor systems (Mera et al., 2012) various testing tools (Goetz et al., 2009) and video-based monitoring systems (Marzinzik et al., 2012). There has however been a lack of mechanisms that combine subjective and objective remote measures into scores that provide a more holistic representation of patients’ general health, their symptom fluctuations and treatment effects. In addition, given the multi-dimensional nature of the PD, assessment methods should address different aspects of the disease and they should be related to the underlying disease process. With any method used for automatic and quantitative assessment of symptoms, it is imperative to develop and introduce integrated IT systems which enable access to relevant data in a user-friendly manner to clinicians for helping them during decision making concerning evaluation of symptoms and treatments (van Bemmel and Musen, 1997).
Based on the above-mentioned issues, the following research questions related to methods for combining subjective and objective measures, and systems for presentation of summarized symptom information were identified and addressed in the thesis:

- **RQ 3:** What are the recent trends and developments in telemedicine applications for collecting and processing subjective and objective measures?
- **RQ 4:** How can we develop a method for combining subjective and objective measures into scores that represent the severity of a patient’s symptoms over week-long test periods?
- **RQ 5:** How can we develop and evaluate a web-based system which enables clinicians to access relevant symptom information in a user-friendly manner that will be of assistance to them during decision making?
- **RQ 6:** Are the computed scores feasible for remote monitoring of PD symptoms over time?

### 1.3 Research approach

This thesis organizes the development and evaluation of the methods with respect to the nature of the data and the identified system-based outcome measures, as illustrated in Figure 1. At the lowest level, telemetry measurements, consisting of subjective and objective measures of fine motor function, were gathered using a touch screen test battery designed for telemedicine. These measurements were performed repeatedly in the patients’ homes.

Method development consisted of two stages: time series analysis and data-driven modelling (DDM). Time series analysis methods were used to extract quantitative measures from raw time series data to represent meaningful information, both in time and frequency domains. These measures included statistical moments to represent the levels and fluctuations of symptoms, trend components to represent the long-term direction of symptoms, irregularity components to represent short-term and abrupt symptom changes, and similarity measures to identify progressive symptom impairment over time, among others.
In the next stage, the quantitative measures were used in combination with multivariate analysis methods to automate the process of symptom assessment. The aim of DDM is to find and model relationships between a set of independent quantitative measures and the dependent outcome usually obtained by clinical ratings. The type of method selected depends on the type of outcome which is desired. For numeric outcomes, numeric prediction (regression) can be applied whereas for nominal and ordinal outcomes, classification is often used. For any method, its performance can be determined by looking at its accuracy or equivalently, at its errors. For these methods despite the importance of having relatively good accuracy other properties such as transparency and interpretability have become desirable (Silipo et al., 2001). Approaches to modelling these methods can be either knowledge-driven or data-driven. The first approach to modelling the methods is theoretical and mainly based on domain knowledge. This knowledge is usually derived from the opinions of experts and is important for describing the structure and processes that govern the overall problem. The models built with this approach can also be referred to as white-box models since results derived from this way can be easily interpreted. The second approach to modelling is empirical and is mainly based on analysis of the data characterising the problem at hand, without having to make
assumptions about underlying physical processes of the problem. In contrast to knowledge-driven models, data-driven models are also known as black-box models since they lack interpretability and their results are difficult to be reproduced. Examples of the most common data-driven methods include methods from different disciplines such as data mining, machine learning and artificial intelligence (AI). Instances of these methods include clustering, principal component analysis (PCA), regression analysis methods, artificial neural networks, etc. According to Solomatine and Ostfeld (2008), data-driven models can be beneficial if i) there is a considerably large dataset, ii) the studied problem does not experience considerable changes during the time period covered by the model and iii) it is difficult to build knowledge-driven models due to the lack of expert knowledge in identifying the underlying processes of the problem.

In this thesis, the focus is on development, evaluation and application of data-driven methods for enabling quantitative assessment of PD symptoms, using data gathered by means of a telemetry test battery, as described by Westin et al. (2010a). The rationale for this narrowed focus only on data-driven methods is mainly based on the nature of the collected data and the targeted system-based outcomes. When designing the test battery, the choice of test items was based on results from two studies (Nyholm et al., 2004; Nyholm et al., 2005). Given the fact that PD is a multidimensional disorder associated with a wide range of motor and non-motor symptoms and that the test battery should be feasible in terms of patient compliance, satisfaction, ease of use and not time consuming during repeated measurements in a patient’s home environment, the aim was to capture a few symptoms which were considered to occur more frequently and important to the majority of patients in the advanced stage of PD. During fine motor tests (tapping and spiral drawing) that were performed in the test battery, position coordinates and timestamps in milliseconds were recorded. From these data it was difficult to identify natural physical phenomena of fine motor movements and translate them into mathematical equations. Therefore, the target outcomes were mainly based on quantification of the levels of symptom severity, e.g. from normal to extremely severe. Nevertheless, during the modelling process of the data-driven methods presented in this thesis domain knowledge was incorporated as follows. Domain experts (i.e. neurologists) were involved during the conceptual formulation of the symptom dimensions of the test battery (Paper V), determination of the most relevant quantitative measures of the test battery (Paper II) and decision making about what data should be used for method development and evaluation by dividing data into different datasets (Paper I and Paper II).
As it was stressed earlier, interpretability of data-driven methods is of paramount importance if they are to be applied to the practice of healthcare. For this reason, the choice of methods was based not only on their simplicity but also interpretability, aiming at providing means for intuitive explanation and interpretation of the derived results in the language of the domain experts. The focus was on multivariate data analysis methods, e.g. PCA and regression analysis methods, which report results as linear combinations of independent measures and are easy-to-visualize. In the presence of long-term and repeated measures, there is a need for a multivariate analysis that considers several random independent measures simultaneously, each of which are considered equally important at the start of method modelling (Manly, 1994). Multivariate data analysis accounts for dependencies between measures and also indicates which ones significantly do and do not add any useful information to the overall model. When repeated measures are done on the same patient over time, during modelling it is also imperative to employ statistical methods which model the within-patient variability often present in longitudinal data. Instances of these methods are mixed-effects models.

Finally, the methods for quantitative and automatic assessment of symptoms should be evaluated for their metrics such as validity and reliability (Kudyba, 2004), (Figure 1, Evaluation stage). Validity refers to the extent to which a method 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 correlation analysis, factor analysis and area under receiver operating curve (AUC). Reliability refers to the extent to which a method is free from measurement error in terms of internal consistency of its sub-items and the test-retest reliability of its results (van de Ven-Stevens et al., 2009). The internal consistency is commonly assessed by Cronbach’s $\alpha$ coefficient whereas test-retest reliability is assessed by intra-class correlation coefficients (ICCs) or Kappa statistics. In addition to having high validity and reliability, the methods should also have the ability to detect subtle symptom changes over time which are a result of treatment interventions as well as they should be able to reflect the expected natural progression of PD. Hence, the third metric is called sensitivity to change (also known as responsiveness).

### 1.4 Thesis outline
The rest of the thesis is organized in the following format. The second chapter presents a general overview of applied IT in healthcare. This is followed by background information, related work and description of the subjects and data. The third chapter summarizes the methods that were used in this thesis during method development and exploratory data analy-
sis. The fourth chapter summarizes the three papers (Paper I-Paper III) of the first research theme “Quantification and analysis of fine motor performance”, describing the motivation, objectives and methodology. The fifth chapter summarizes the remaining three papers (Paper IV-Paper VI) of the second research theme “Methods and systems for remote and long-term assessment of symptoms”, describing the motivation, objectives and methodology. The sixth chapter summarizes the results of the six appended papers. The seventh chapter provides a discussion in terms of contributions and answers to research questions followed by general implications of the work, future directions for research and concluding remarks of the thesis.
2 Background

2.1 Overview of applied IT in healthcare

Over the past several decades, IT has produced major breakthroughs in healthcare and has had a great impact on transforming it from in-hospital to more advanced in-home healthcare (Koch, 2006; Chaudhry et al., 2006). There are many factors which contribute to this shift of healthcare including the nature of emerging diseases and their treatments, demographic changes in population, societal demands for healthcare cost-containment, increased availability of complex healthcare medical equipment and services at home, increased amount of rehabilitation services, and an increased focus on self-care and quality of life, among others (Pepe et al., 2004). In routine clinical settings, information processing and communication are paramount and centrally involved in different healthcare activities including patient data collection, communication among patients, communication among healthcare professionals, decision making in diagnostics and therapeutics, interpretation of laboratory results, collection of clinical research data, etc. (Balas et al., 1996; Georgiou, 2002). With the trend of shifting healthcare from the hospital to the patient’s home, the need for the remote monitoring and treatment of patients emerges (Stanberry, 2000; Hebert et al., 2006). From a patient’s perspective, there would be a great need for improved ability for self-managed care through constant doctor-patient consultations (Chin, 2003). In the case of patient groups suffering from chronic and progressive diseases, information flow between the patient and healthcare professionals during remote monitoring becomes more complex and challenging compared to the in-hospital consultations. Application of IT has shown to efficiently improve and facilitate the information flow and the relationship between patients and healthcare professionals (Young et al., 2007; Miller, 2003).

The introduction and application of IT-based systems developed to support the clinical management of diseases in home healthcare provide a means of reducing medication and diagnostic errors, increasing efficiency and supporting healthcare professionals during the decision making process (Ammenwerth et al., 2003). The 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 main characteristics that make the application of IT effective in different disciplines 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 the application of IT can improve the quality and efficiency of healthcare while reducing its costs, improving clinicians’ performance, improving health outcomes as well as increasing patient compliance (Blum, 1986; Balas et al., 1996; Hunt et al., 1998; Chaudhry et al., 2006; Hebert et al., 2006). Recent advances in a variety of disciplines like wireless communications, mobile computing, sensing technology, clinical decision support systems (CDSS) and Web technology enable patients to be monitored remotely while offering reliable and cost-effective home healthcare solutions (Bellazzi et al, 2001; van Halteren et al., 2004; Lu et al., 2005; Chen et al., 2011). The process of applying the above mentioned technologies in healthcare is known as telemedicine.

2.1.1 Telemedicine
According to the American Telemedicine Association (ATA) telemedicine refers to “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, 2013). One of the telemedicine services is remote patient monitoring which refers to the use of devices to remotely collect and send data to a central station for further processing and interpretation. The process of measuring and transmitting data from remote sources, e.g. a patient’s home, to a central station for processing and analysis is known as telemetry, a term derived from the Greek words tele (remote) and metron (measure). In scientific literature, the application of telemedicine can also be known as telehomecare, home telehealth or home-based eHealth (Koch, 2006). In 2003, the ATA produced guidelines known as 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 data privacy and confidentiality, patient education and satisfaction. The healthcare provider criteria define ways of improved delivery of care through education and administration of healthcare profes-
sionals. The technology criteria recommend the type of technology to be used, its maintenance and support.

2.1.2 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 and anywhere” (Burley et al., 2005). The mobile computing technology improves healthcare in a number of ways such as providing healthcare professionals access to reference information and electronic medical records as well as improving communication. It also provides computerized monitoring of clinical information and clinical decision support, important for healthcare professionals during the decision making process (Ruland, 2002; Bates and Gawande, 2003).

An instance of mobile computing technology is Personal Digital Assistants (PDAs). 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). To date, in clinical settings, the most common approach to assessing symptoms is through paper home diaries. The 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). On the other hand, 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 date and time of the entry (Drummond et al., 1995; Nyholm et al., 2004; Lyons and Pahwa, 2007).

2.1.3 Information processing
According to Van Bemmel and Musen (1997), healthcare professionals go through three stages to complete the diagnostic-therapeutic cycle including observations, diagnosis, and therapy. IT-based data collection schemes combined with intelligent data analysis methods can be used to process, analyse and interpret large datasets derived from many patients in order to draw conclusions through inductive reasoning. During computerized information processing, we observe a process similar to the clinical diagnostic-therapeutic cycle where the stages are measurement and data entry, data processing and output generation. Van Bemmel and Musen (1997) furthermore defined a six-layer model for structuring IT-based systems intended for application in the practice of healthcare, with increased complexity and increased dependence on human interaction with respect to each layer. The layers of the model were 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.

An instance of IT-based systems is CDSS. According to Greenes (2006) computer-based clinical decision support can be defined as “the use of the computer to bring relevant knowledge to bear on the health care and well-being of a patient”. These systems relate to integrated systems combining different modules such as database management systems for data storage and knowledge representation, data mining and statistical pattern recognition for exploratory data analysis and prediction, and Web technologies for data presentation. According to Berner (1999), CDSS can either be knowledge-based or non-knowledge-based systems. The knowledge-based systems mostly consist of three parts: the knowledge base (e.g. experts’ knowledge coded in forms of if-then rules), the inference or reasoning engine for mapping the rules in the knowledge base to the actual patient data, and the communication mechanism for delivering the output of the system to the end-user who will make the actual decision. Unlike knowledge-based systems, non-knowledge-based systems are based on the application of the DDM methods such as machine learning and AI methods for learning from historical clinical data by finding patterns and constructing a model that can be utilized for predicting future data.

The most important part of the DDM process is learning which refers to mapping of dependent and independent measures. The data are usually divided randomly into two sets. The part which is used for building the method is called the training set. However, the validity of the method should be evaluated by assessing the error on another dataset that played no role in its formation in order to check its generalization abilities to unseen data (Witten and Frank, 2005). This independent dataset is called the test set. When the amount of data for modelling is limited, a more general technique known as cross-validation or rotation estimation, 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 guarantees proper representation of each outcome in both the training and testing sets, a stratification procedure is employed.

In this thesis the remote monitoring of PD symptom severity is central and is mainly achieved by applying methods for processing and analysing the telemetry data as well as for presenting symptom information to end users i.e. clinicians who in turn will use the information during decision making process concerning evaluation of symptoms and treatments. At the data processing and analysis stage, the main focus was on designing and implementing server-side IT modules for receiving, processing, storing and
interpreting the data saved in relational databases. These modules were based on methods for automating the process of symptom scoring, given their good metrics such as validity, reliability and sensitivity to treatment changes. During this stage, time series analysis methods were initially employed in order to derive semantic and quantitative symptom information from the data, using time-domain and frequency-domain methods like summary statistics, discrete wavelet transform, dynamic time warping, approximate entropy, and others. Next, the DDM methods were employed in order to map the quantitative measures to clinically derived measures. At the presentation stage, the focus was on developing custom web-based systems for enabling visual and objective representation of symptoms to clinicians. The idea was to develop systems that are user-friendly, provide a fast response, enable rapid and convenient screening of patients, and provide a comprehensive overview of patients on a single page. In addition, at the presentation stage, the thesis investigates the development of web-based frameworks for eliciting clinical knowledge about a patient’s motor performance by allowing clinicians to visualize the raw time series data.

2.1.4 Evaluation of IT-based systems

The general architecture of the IT-based systems includes separate software modules for collecting, transferring, processing and presenting data. Generally, strict evaluation of IT in healthcare is recommended and of high importance for decision makers and users (Ammenwerth et al., 2003). Ammenwerth et al. (2003) discussed three problems that may occur during evaluation of IT-based applications in healthcare; these are the complexity of the evaluation object, complexity of an evaluation project and motivation for the evaluation. When evaluating IT-based systems that will be applied in healthcare settings, generally there is a diverse number of evaluation approaches and there is a need for standard framework for performing the evaluation (Rahimi and Vimarlund, 2007). In the majority of cases, the focus is on aspects concerning user satisfaction, financial benefits and improved organizational work. In addition, there is a need for an evaluation approach which takes into account different dimensions such as those related to stakeholders and the evaluation process itself. Carson et al. (1998) proposed an approach based on a so-called stakeholder matrix analysis which takes into account the above mentioned dimensions. Finally, factors such as privacy and security issues, system and information quality and technical limitations of the systems still need to be investigated (Wu et al., 2007).
2.2 Parkinson’s disease
Parkinson’s disease (PD) is a progressive neurological disorder which is caused by degeneration of dopamine producing nerve cells in a region of the brain called the substantia nigra. These cells release dopamine which acts as a neurotransmitter essential for smooth control of movement. The prevalence of PD increases with age; approximately 2% of the population over the age of 65 have this disease. Although the etiology still remains unclear, the disease probably results from an interaction between genetic and environmental factors (Warner and Schapira, 2003).

2.2.1 Clinical features
Clinical symptoms develop with a substantial variability among patients once there is at least 50% degeneration of dopaminergic nerve cells (Grosset et al., 2009). The four cardinal motor symptoms of PD are comprised of bradykinesia (slowness of initiating voluntary movements), rigidity (increased muscle tone), tremor (a 3-5 Hz tremor at rest) 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 made by clinical assessments. It includes criteria for the presence of bradykinesia in combination with one or more of the other three motor symptoms plus a positive response to treatment.

2.2.2 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 stage 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 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 sufficient for the patient to respond as a non-parkinsonian person). In addition to these two motor states, patients in the On state may develop abrupt involuntary movements, also known as dyskinesias, in response to peak levels of levodopa.

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 Off and On states (Gunal et al., 2002). Over 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; Djaldetti et al., 1996). In order to reduce Off times and symptom fluctuations as well as to improve patients’ health-related QoL, an alternative to oral treatment is continuous intraduodenal administration of levodopa-carbidopa intestinal gel (LCIG, Duodopa®; AbbVie) (Nilsson et al, 2001; Nyholm et al., 2005). Generally, medications must be fine-tuned to the individual patient’s needs with regard to the timing and quantity of each dose and with regard to food intake, mood and daily physical activities.

2.2.3 Symptom assessment in clinical settings

In routine clinical settings, the severity of symptoms is scored quantitatively using clinical rating scales. The scales are used as instruments by observers to evaluate PD-related disability and impairment in order to provide a comprehensive clinical picture of patients. According to the World Health Organization, impairment is defined as an abnormality of body or organ structure or function whereas disability is defined as a global health picture related to a reduction of a person’s ability to perform a basic task (Simonsen et al., 2000). In PD, impairment usually relates to major symptoms, e.g. bradykinesia, dyskinesias, which act as underlying causes of a patient’s disability to perform ADL within the range of normal.

A common rating scale to describe progression of symptoms is the five-point Hoehn and Yahr scale (Hoehn and Yahr, 1967). Weaknesses of this scale include mixing of impairment and disability, the strong emphasis on postural instability over other symptoms and the 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). The UPDRS (Fahn et al., 1987) is a multi-dimensional scale and is the most widely used clinical scale for assessing PD motor impairment and disability (Mitchell et al., 2000; MDSTFRSPD, 2003). It is made up of four parts covering mentation, behaviour and mood (Part I); ADL (Part II); motor performance (Part III); and complications of therapy (Part IV). Parts I, II and IV are assessed by interviewing the patient or self-evaluation whereas Part III is assessed by physical examination. Part I, II and III contain 44 questions, each of which are scored on a five-point scale ranging from 0 (normal) to 4 (severe). 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 disability. In the study performed by Ramaker et al. (2002), it was found that UPDRS is the 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
UPDRS scale, resulting in a new version called MDS-UPDRS (Goetz et al., 2008). This was done based on the recommendation found in a previously published critique (MDSTFRSPD, 2003). Other scales designed to evaluate non-motor symptoms also exist such as Mini Mental State Examination, Dementia Rating Scale, PDQ-39, etc. 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).

2.3 Related work
There have been a number of initiatives from different research groups to address the development and evaluation of methods and systems that enable PD symptom quantification and remote monitoring. The majority of the approaches address a one-dimensional construct of the disease by targeting a specific symptom, e.g. gait. Since PD is a multidimensional disorder, there is a need to combine measurements of multiple symptoms with the aim to better reveal the clinical picture of the symptom severities and fluctuations. An outline of objective assessment methods described in the scientific literature is given below. This is done by first focusing on quantification of fine motor performance through tapping tests and spirography and secondly focusing on methods and systems that collect and measure other PD-specific symptoms.

2.3.1 Quantification of fine motor performance
Spirography is an objective method of evaluating the severity of PD-related symptoms by enabling time series analysis of data, which are usually gathered by a measurement tool (e.g. digitizing tablet), for extraction of detailed motor features from spiral drawing tasks. Digitizing graphic tablets have been widely utilized in studies where both healthy subjects and patients suffering from different movement disorders have participated. Digitizing tablets have been mainly employed for recording digitized movement data of individuals which in turn are used for off-line analysis and quantification through time- and frequency-domain methods to derive measures that describe the intensity and frequency of fine motor symptoms. When compared to accelerometry, digitizing tablets had several advantages and provided more accurate measures of tremor frequency and amplitude during fine motor movements (Elble et al., 1996). These devices not only record x and y coordinates but also the pressure exerted by the drawing instrument thus providing a rich source of information about movement dynamics. In most of the studies, spatial and spectral analysis of digitized drawing specimens during spirography was performed to detect fine movement anomalies. In the study performed by Elble et al. (1996), auto-
and cross-spectral analyses were performed on digitized data of Archimedes spiral drawings for evaluating essential tremor. An objective method to assess severity of PD symptoms through spirography has been developed and validated in a study performed by Pullman et al. (2008), by first extracting time domain quantitative measures and then using them in a neural network and regression analyses as independent measures to be mapped to UPDRS scores. Fourier transform was used in the study performed by Rudzinska et al. (2007) to measure three tremor intensity coefficients representing displacement, velocity, and acceleration. In another study performed by Haubenberger et al. (2011), Fourier transform was used to quantitatively measure tremor severity in patients with essential tremor. Using a four-pole Butterworth filter, drawing velocity signals in horizontal, vertical, radial, and tangential directions were extracted to derive dyskinetic movements and action tremor (Liu et al., 2005). An approach based on the use of argument-based machine learning was developed and evaluated by Groznik et al. (2013) to differentiate different types of tremors, using spirography data. A semiautomatic method of quantifying scanned specimens of Archimedes spiral drawings through the use of the cross-correlation function and Fourier transform was presented in the study by Miralles et al. (2006).

The other objective method to assess fine motor performance relies on analysis of time series data collected during different upper limb tapping tasks. Tapping tests have been widely investigated and in general they are useful in recognizing reduced tapping frequency as well as increased tapping variability over time. In clinical settings the most common approach to assess tapping performance of PD patients is to use item #23 (Finger Tapping) of the UPDRS scale (Fahn et al., 1987). However, the main challenge for the clinical management of upper limb motor impairments is that it is difficult to capture symptom fluctuations since patients exhibit different movements as a result of the nature of the disease, the large within-patient variability of symptom profiles and the unstable medication responses. As a result, objective assessment of fine motor performance during tapping tests has been previously tried using different technologies for sensing tapping movements followed by employing time series analysis methods. The study performed by Freeman et al. (1993) showed that the index finger tapping performance of PD patients improved with inclusion of external timing cues. Objective measurements of several kinematic measures during repetitive alternating finger-tapping tasks using computer-interfaced musical keyboards was shown to correlate well with clinical ratings and be sensitive to treatment interventions (Taylor Tavares et al., 2005; Bronte-Stewart et al., 2000). The use of magnetic sensing systems coupled with
Fourier transform and time-domain methods has also shown to be valid and reliable for quantitative assessment of finger tapping performance of PD patients (Kandori et al., 2004; Shima et al., 2009). Finger tapping measures in a computerized assessment battery showed to be valid in documenting clinical changes over time in PD patient performance (Goetz et al., 2009). In the study performed by Mera et al. (2012), it was shown that it is possible to detect treatment changes using data from repetitive finger tapping tasks recorded by a wireless motion sensor technology. Other approaches to quantitative assessment of fine motor performance through tapping tests include digitized switch boards (Yahalom et al., 2004), optoelectronic cameras (Konczak et al., 1997), infra-red emitting diodes (Ling et al., 2012), personal computer keyboards (Giovannoni et al., 1999) and accelerometry (Yokoe et al., 2009; Stamatakis et al., 2013).

2.3.2 Systems and methods for monitoring PD symptoms

There are a number of objective assessment tools reported in the scientific literature that aim to provide a means for continuous and at-home assessments of PD-related symptoms including bradykinesia, therapy-induced dyskinesias, akinesia, posture and gait deficits, hand grasping, dysarthria, and movement asymmetries, among others.

The most popular approach to objective assessment of PD symptoms involves the use of wearable sensor technology. In recent years, wearable sensors technology has enabled a detailed assessment of gait by documenting the subtle movement characteristics of patients (Zampieri et al., 2011; Sant’Anna et al., 2012). Gait deficits, such as freezing of gait, are usually considered as late symptoms of PD. Wearable sensors have also been used in quantitative posture analysis of early PD patients to allow easy monitoring of their balance maintenance (Palmerini et al., 2011). People with PD also develop difficulties during normal articulation of speech (also known as dysarthria) over time. Objective and remote assessments of dysarthria has been previously reported in Little et al. (2009) and Khan et al. (2013) and showed to correlate well with UPDRS speech ratings.

Commercially wearable sensors were used to develop a protocol to objectively measure motor impairment in early and moderate PD patients (Brewer et al., 2009). The severity of bradykinesias in PD patients has been continuously assessed using wrist watches (Griffiths et al., 2012) and inertial sensors (Zwartjes et al., 2010). The feasibility of a computerized approach to assess the severity of dyskinesias and to distinguish levodopa-induced dyskinesias from voluntary movements was evaluated in the study performed by Keijsers et al. (2000). This was done by gathering data from multiple accelerometers mounted on the upper and lower limbs of PD pa-
tients which in turn were processed and used in a neural network for classification. In addition, other approaches at distinguishing between the severity of therapy-related complications have also been tried using wearable sensors (Tsipouras et al., 2012) and force plates (Chung et al., 2010), each of which were coupled with time series and machine learning methods. In addition to wearable sensing technologies, recently other game motion sensing technologies such as the Nintendo Wii Remote (Synnott et al., 2012) and Xbox Kinect (Softronic AB) have also been used for quantitative assessment of PD symptoms.

2.4 Subjects and data

2.4.1 Subjects
The results presented in this thesis are based on data from two clinical studies, both of which were approved by the relevant agencies and written informed consent was given. In total, 95 patients in different clinical stages of PD and 10 healthy elderly subjects (HE) were assessed (Table 1). Sixty-five patients with advanced PD were recruited in an open longitudinal 36-month study (Duodopa in Advanced Parkinson’s: Health Outcomes & Net Economic Impact, EudraCT No. 2005-002654-21) at nine clinics around Sweden (Pålhagen et al., 2012). On inclusion, 35 of the patients were treated with LCIG (hereafter denoted as LCIG-non-naive) and 30 patients were candidates for switching from conventional oral PD treatment to LCIG (hereafter denoted as LCIG-naive) (Nyholm, 2012a). Thirty patients in Milan, Italy, who had a clinical diagnosis of idiopathic PD participated in a second study (Westin et al., 2012). The Italian study consisted of two patient groups: intermediate stage patients experiencing On-Off fluctuations (F group) and clinically stable patients (S group).

In the general study designs, the primary outcome measures included clinical rating scores and the secondary outcome measures included symptom severity scores measured using the telemetry test battery, as described below. In the Swedish study, the clinical evaluation included administration of UPDRS, PDQ-39, and Hoehn and Yahr staging scales. The evaluations were performed in the afternoon at the start of each test period (Pålhagen et al., 2012). LCIG-naive patients were evaluated at baseline and follow-up test periods whereas LCIG-non-naive patients were evaluated starting from month 0. In total, clinical assessments were done during 369 test periods. In the Italian study, assessments for both groups, i.e. F and S, were performed in the On state at the end of each test period (Westin et al., 2012).
2.4.2 Symptom data collection via a telemetry device

Both patients and HE subjects performed repeated, time-stamped and remote assessments of their subjective and objective health indicators using a wireless telemetry test battery implemented on a touch screen handheld computer (Westin et al., 2010a; Westin, 2010b). The test battery consisted of a patient diary section for collecting self-assessments of common PD symptoms, a motor test section for collecting objective measures of motor function through a set of upper limb motor tests, and a scheduler for restricting operation to a multitude of predetermined limited time slots.

The overall system was implemented in a client-server architecture for collecting and wirelessly transmitting the remote data to a central server for storage and off-line processing. On each test occasion, the data were transmitted from the handheld computer over a Universal Mobile Telecommunication System to the server where the Remote Device Manager (RDM, Nordforce Technology AB) software was executed. The RDM is a commercial software platform that aims to provide a high availability data communication link with a very high security level over wireless internet. Once the data were received at the server side, they were initially stored as separate Extensible Markup Language (XML) files which in turn were interpreted by custom designed software in .NET. This software parsed, processed and stored both the raw and summarized data into relational database tables for later use. The test battery software was implemented on a Qtek 2020i Pocket PC device having a screen size of 60mm × 80mm.

Measurements with the test battery were performed four times per day in patients’ homes during week-long test periods. On each test occasion, subjects were instructed to place the device on a table, be seated in a chair and use an ergonomic stylus to perform the tests. Self-assessments were designed to target more subjective aspects of the symptoms whereas motor tests were designed to represent a more objective view of the symptoms, being connected to the physiological functioning of the patient.

On each test occasion, patients were first asked to answer a set of PD-related questions and then to perform the fine motor tests. There were seven self-assessment questions (q) relating to the previous 4 hours or that morning, including “Ability to Walk” (q1), “Portion of time spent in Off, On and Dyskinetic” (q2), “Off at worst” (q3), “Dyskinetic at worst” (q4), “Painful cramps” (q5), “Satisfied with functioning” (q6), and “Momentary motor condition” (q7). The q1, q3, q4, q5 and q6 were of the verbal descriptive scale type with answer alternatives ranging from 1 (worst) to 5 (best). In q2, patients were asked to mark their portion of time spent in the Off, On and dyskinesia. The q7 allowed seven categories of momentary motor states ranging from -3 (very Off) to 0 (On) to +3 (very dyskinetic).
The motor tests included different tapping tests and tracing a pre-drawn Archimedes spiral. The tapping tests consisted of four 20 second long tests, including “Uncued alternate tapping of two fields using the right hand” (q8), “Uncued alternate tapping of two fields using the left hand” (q9), “Tapping with increasing speed using the dominant hand” (q10), and “Tapping with random chasing using dominant hand” (q11). For the spiral drawing test, patients were asked to trace a pre-drawn Archimedes spiral using the dominant hand, and the test was repeated 3 times per test occasion (q12-q14).

The raw test data, consisting of responses to questions and movement recordings during motor tests such as stylus position (x-y coordinates) and timestamps (in milliseconds), were collected and wirelessly transmitted to the central server for storage and off-line processing.

In the Swedish study, LCIG-naïve patients used the test battery at baseline (before LCIG), month 0 (first visit; at least 3 months after permanent intraduodenal LCIG surgery), and thereafter quarterly for the first year and biannually for the second and third years. The LCIG-non-naïve patients used the test battery from the first visit, i.e. month 0. In the Italian study, patients used the test battery for two test periods with a washout week in between. The HE subjects used the test battery for one test period. The total number of observations with the test battery were as follows: Swedish group (n = 10079), Italian F group (n = 822), Italian S group (n = 811), and HE (n = 299).

<table>
<thead>
<tr>
<th></th>
<th>Swedish study</th>
<th>Italian study (F group)</th>
<th>Italian study (S group)</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n, gender)</td>
<td>65 (43m; 22f)</td>
<td>15 (13m; 2f)</td>
<td>15 (13m; 2f)</td>
<td>10 (5m; 5f)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>65 ± 11</td>
<td>65 ± 6</td>
<td>65 ± 6</td>
<td>61 ± 7</td>
</tr>
<tr>
<td>Years with levodopa</td>
<td>13 ± 7</td>
<td>7 ± 8.5</td>
<td>5.5 ± 6</td>
<td>NA</td>
</tr>
<tr>
<td>Hoehn and Yahr stage at present</td>
<td>2.5 ± 1*</td>
<td>2 ± 0**</td>
<td>2 ± 0.5</td>
<td>NA</td>
</tr>
<tr>
<td>Total UP-DRS</td>
<td>49 ± 20.5*</td>
<td>33.5 ± 11.8**</td>
<td>26 ± 16.5</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of PD patients and of healthy elderly participants, presented as median ± interquartile range. * Assessments performed in the afternoon. ** Assessments performed in the On state. Abbreviation: HE, healthy elderly; NA, not applicable.
3 Methods

When analysing time series, the applied methods need to be carefully selected according to their statistical characteristics. This chapter summarizes the methods that were used in this thesis for both method development and exploratory data analysis. During method development both time- and frequency-domain methods were applied on time series data to derive meaningful quantitative measures representing the severity and fluctuation of symptoms. Going further, different data-driven methods were then used to map the quantitative measures to clinical measures. In addition, in this thesis appropriate statistical methods for analysing repeated measures and longitudinal data were employed. All the methods were used in a combined fashion and come from different disciplines from statistics to machine learning. The selection of the methods was mainly based on the nature of the collected data as well as on the problem at hand.

3.1 Discrete Wavelet Transform

Since the majority of the time series generated from clinical problems are highly non-stationary and transient in nature, using traditional frequency-domain methods such as the Fourier transform is not suitable to extract the time of frequency occurrence as well as the amount of frequency changes over time (Akay, 1995). In addition, real world problems generate time series which contain both low frequency components occurring for long durations and high frequency components for short durations. All these problems are addressed by using a Discrete Wavelet Transform (DWT) which enables a multi-resolution analysis (MRA) of time series by employing a series of complementary high-pass (HP) and low-pass (LP) filters (Daubechies, 1988).

The MRA approach relies on an iterative process which breaks the time series signal into several frequency sub-bands in order to provide more discriminative features. This procedure enables simultaneous localization of time and frequency and is also known as a “mathematical microscope” through which different parts of the time series signal can be observed by adjusting the focus (Akay, 1995), using down-sampling and filtering operations with wavelet functions. When analysing time series using MRA with DWT, the resolution is adjusted depending on the type of the frequency components. For instance, when analysing low frequency components, the frequency resolution is narrow whereas time resolution is coarse. On the other hand, when analysing high frequency components, the frequency resolution is coarse and time resolution is narrow. The original time series signal is decomposed into wavelet coefficients for low frequency compo-
nents or approximations (A) and high frequency components or details (D), using the following equations:

\[
y_{\text{high}}(k) = \sum_{n} x(n) \cdot g(2k - n)
\]

\[
y_{\text{low}}(k) = \sum_{n} x(n) \cdot h(2k - n)
\]

where \( y_{\text{high}} \) and \( y_{\text{low}} \) are the outputs of HP and LP filters, respectively and \( g(n) \) and \( h(n) \) are their corresponding filter impulse responses, after subsampling by 2. The A coefficients are further decomposed in other levels, as shown in Figure 2. The A coefficients represent main features of the time series signal whereas D coefficients represent short and fast fluctuations. The final result of the MRA includes \( D_1 - D_3 \) coefficients and one final set of \( A_3 \) coefficients. Considering that the time series signal in Figure 2 had a maximum frequency component of 10 Hz, then the frequency bandwidths would be as follows, \( A_3 \) (0-1.25 Hz), \( D_3 \) (1.25-2.5 Hz), \( D_2 \) (2.5-5 Hz) and \( D_1 \) (5-10 Hz).

![Figure 2. Three level wavelet decomposition tree. HP is the high-pass filter, LP is the low-pass filter, D is details and A is approximations.](image)

In different engineering problems, summary statistics can be computed over these coefficients in order to reduce their dimensionality. For example, in the work performed by Kandaswamy et al. (2004) and Cvetkovic et al. (2008) mean, variance and ratio of the mean values for adjacent frequency bands were calculated and used in subsequent analysis. However, in this thesis this was accomplished by applying the PCA method which will be further explained below.

In paper I, when quantifying the severity of drawing impairment during spiral drawing tasks, the DWT was applied. The DWT decomposed the spiral drawing signal into different frequency bands in order to extract both the overall severity level as well as fast and sharp spatial displacements.
3.2 Principal Component Analysis

Principal Component Analysis (PCA) is a method that is widely used in applications such as dimension reduction, lossy data compression, feature extraction and analysis, and data visualization (Jolliffe, 2002). The goal of PCA is to take \( n \) variables \( X_1, X_2, \ldots, X_n \), find combinations of these and transform them into a new set of non-correlated variables \( Z_1, Z_2, \ldots, Z_n \) called principal components (PC) (Chatfield and Collins, 1980), denoted as:

\[
Z_i = a_{i,1}(X_1 - \bar{X}_1) + a_{i,2}(X_2 - \bar{X}_2) + \cdots + a_{i,n}(X_n - \bar{X}_n)
\]

(3)

where \( \bar{X}_n \) represent the mean of the original variables and \( a_{i,n} \) represent their corresponding weights. From equation 3, we can notice that the PCs are linear combinations of the original variables and are derived in decreasing order of eigenvalues so that the first PC (PC1) accounts for as much of the joint variation in the data as possible. The effect of the dimension reduction process is achieved when original variables are highly correlated, meaning that they measure the same “concept” and they can be adequately represented by the first two or three PCs, a situation that helps in better understanding the data as well as in operating with a smaller number of variables in subsequent analyses. A further advantage of PCA is that the PCs are uncorrelated to each other and using them as independent variables in data-driven models (e.g. regression models) helps in avoiding the problem of multicollinearity.

The most important step when applying PCA is to identify and retain the most relevant components accounting a relatively sufficient proportion of variation in the data. There are two most frequently used approaches to determine the appropriate number of components. The first one relates to a cut-off criterion by applying the so-called Kaiser-Guttman rule which means to retain all PCs that have an eigenvalue > 1 (Jackson, 1993). In the second approach, a cumulative percentage of total variation is selected for which it is desired that the retained PCs should contribute more than 70% of the total variation in the data.

In this thesis, the PCA method was used in Paper I, Paper II and Paper V. The main aim was to summarize a set of quantitative measures, which were derived by time series analysis, as accurately as possible using a few PCs in subsequent analyses. The PCA was applied using the correlation matrix method. The motivation for using correlations instead of covariance was based on the fact that the quantitative measures had different scales.
3.3 Approximate Entropy

Physiological and clinical time series data are associated with chaotic behaviour which provides an insight on the status of the subject and reflects important physiological information. These time series contain complex dynamics and sequential irregularity (or aperiodicity). In order to quantify the degree of irregularity in the time series, in this thesis the Approximate Entropy (ApEn) method was applied (Pincus, 1991). The ApEn is a statistical and non-linear measure which reflects the similarity between a chosen window of time series of a given duration and the next set of windows of the same duration. A time series containing a single frequency component has a relatively small ApEn value whereas more complex time series containing multiple frequency components have high ApEn values, as a result of a high level of irregularity.

Given a time series $S$ containing $N$ data points, ApEn requires determination of two user-specified parameters: a length of the window $m$ and a measure of similarity $r$, each of which must remain fixed during all calculations. Initially, a set of time series is constructed and expressed as

$$X(N - m + 1) = \{x(N - m + 1), x(N - m + 2), ..., x(N)\}$$  \hspace{1cm} (4)

each of which composed of $m$ consecutive values of $S$. Next, a distance measure $d(X(i), X(j))$ between vectors $X(i)$ and $X(j)$ as:

$$d \left[ X(i), X(j) \right] = \max_{k=1,2,...,m} (|x(i + k - 1) - x(j + k - 1)|)$$  \hspace{1cm} (5)

For each vector $X(i)$, count the number of $j (j = 1, 2, ..., N - m + 1)$ such that $d \left[ X(i), X(j) \right] \leq r$, denoted as $k_{r,m,i}$. Then, for $i = 1, 2, ..., N - m + 1$, calculate a measure of frequency of patterns similar to a given window of length $m$, using:

$$C_{r,m}(i) = \frac{k_{r,m(i)}}{N - m + 1}$$  \hspace{1cm} (6)

To represent the average frequency of all patterns, the following term is defined:

$$\phi_{m}(r) = \frac{\sum_{i=1}^{N-m+1} \log C_{r,m(i)}}{N - m + 1}$$  \hspace{1cm} (7)

and the ApEn value of the time series can be calculated as:

$$ApEn(m, r, N) = [\phi_{m}(r) - \phi_{m+1}(r)]$$  \hspace{1cm} (8)

When dealing with two time series, an extended version of ApEn called Cross-ApEn can be used to quantify the regularity of patterns in those series. In contrast to ApEn, Cross-ApEn is applied to two signals and thus measures the dissimilarity between them. In addition, Cross-ApEn evaluates both spatial and temporal dissimilarities whereas the ApEn reflects only the temporal irregularity.
In Paper II, the ApEn and Cross-ApEn methods were applied to quantify the degree of severity of different symptoms in time series data generated during alternating tapping tests. The aim was to quantitatively characterize the continued demotion of tapping performance relative to the passage of the test trial as well as the serial irregularity and abrupt changes during tapping.

### 3.4 Dynamic Time Warping

Dynamic Time Warping (DTW) is a method for comparing two time series of different lengths and frequencies by first locally stretching or compressing them and then by “warping” their time axes so that a relationship between their data points is maintained.

Given two discrete time series, an input $A = (a_1, a_2, ..., a_N)$ where $i = 1 \ldots N$ and a reference $B = (b_1, b_2, ..., b_M)$ where $j = 1 \ldots M$, the DTW compares them as follows. The first step is to calculate absolute local dissimilarities between paired $ith$ data points of $A$ and $jth$ data points of $B$, leading to a construction of a cross-distance matrix ($d$). The matrix $d$ has small values if the data points are similar and large values if they are different. Next, an alignment path (or warping path) is created using a warping function $w(k) = (w_a(k), w_b(k))$ where $k = 1 \ldots T$, $w_a(k) \in \{1 \ldots N\}$ and $w_b(k) \in \{1 \ldots M\}$. This path remaps the data points of $A$ and $B$ by minimizing their distance following the condition that the first and last data points of the two time series are aligned. Other constraints such as monotonicity and step size are imposed on the function $w$ to ensure reasonable wraps. The mean normalized distance ($mnd$) is then calculated using the following formula:

$$mnd_w(A, B) = \sum_{k=1}^{T} \frac{d(w_a(k), w_b(k))c_w}{C_w}$$

where $c_w$ is a per-step weighting coefficient and $C_w$ is the corresponding normalization coefficient. Finally, to find the optimal path the minimum global dissimilarity ($mgd$) is calculated using Dynamic Programming which breaks the entire set of solutions in sub-solutions thus reducing the number of computations, using the following formula:

$$mgd(A, B) = \min_{w} mnd_w(A, B)$$

In Paper II, the DTW method was used as a similarity measure to compare different parts of the tapping time series in order to quantify the progressive reduction of speed and reaction time over the test trial.
3.5 Multiple Linear Regression

The Multiple Linear Regression (MLR) allows us to model the joint probability of two or more independent variables to a quantitative dependent variable by fitting a linear equation to observed data. In this case, the dataset would consist of a set of independent variables $x$ and their corresponding values of the dependent variable $y$. The goal is to fit a line through the data points so that the squared deviations of the observed data points from the line are minimized, a procedure known as least squares estimation. The regression line is defined as a linear combination of the $x$’s to explain the variation in $y$, as follows:

$$Y = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k + \epsilon$$

where $Y$ is the predicted value of $y$, $a$ is a constant term representing the intercept of the line, $x_i$ is the $ith$ dependent variable ($i = 1, 2, \ldots, k$), $\epsilon$ is the noise or unexplained part which is assumed to be normally distributed and $b_i$ are the estimated regression coefficients. These coefficients represent the expected changes in $y$ per every unit change in independent variables, for example $b_1$ represents the amount by which $y$ increases on average if we increase $x_1$ by one unit while keeping all the other independent variables constant. In order to evaluate the model fit, the most commonly used statistic is the coefficient of determination which represents the percentage of variation in dependent variable that is explained by the relationship of the independent variables.

In Paper V, the MLR method was used to define the overall test score to reflect the global health condition of the patient during a week-long test period. This was done by weighing quantitative measures representing six symptom dimensions of the test battery, using the UPDRS scale. The rationale for choosing the MLR method over other data-driven models was that the overall test scores model was defined as a linear combination of the quantitative measures which in turn was easy to interpret and visualize for domain experts.

3.6 Logistic Regression

Like MLR, the Logistic Regression (LR) relies on defining a model which attempts to relate independent variables with the dependent variable. In the case of LR, the dependent variable, $y$, is categorical and the derived model is used for classification of observations into classes. LR defines a function called logit which is modelled as a linear combination of the predictors and then mapped to probabilities $p$ that an observation will be classified into a particular class. The linear combination $\eta$ can be expressed as:

$$\eta = a + b_1x_1 + b_2x_2 + \ldots + b_kx_k$$
where $a$ is the intercept, $b_k$ are the regression coefficients and $x_k$ are the independent variables. In order to ensure that the right-hand side of equation 12 will lead to class values, initially a non-linear function of the predictors is defined as:

$$ p = \frac{1}{1 + e^{-\eta}} $$

followed by defining odds of belonging to a class, using:

$$ Odds = \frac{p}{1 - p} $$

The proportional relationship between the odds and the independent variables can be defined as:

$$ Odds = e^{a + b_1x_1 + b_2x_2 + \cdots + b_kx_k} $$

describing the percentage of expected changes in dependent variable per unit change in a particular independent variable while holding all other predictors constant. The final form of the LR model is then defined as below by taking the natural logarithm of both sides of equation 15:

$$ \log(odds) = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k $$

In Paper II, the LR method was used to classify the alternating tapping tests in five classes of global tapping severity ranging from 0 (normal) to 4 (extremely severe). The LR method mapped the first 5 PCs of 24 quantitative measures, derived by time series analysis, to corresponding visually and clinically assessed scores from tapping graphs.

### 3.7 Mixed-Effects Models

Longitudinal data is often collected in clinical studies with the objective of assessing the progress of symptoms and the disease itself as well as the effect of treatment over time. A longitudinal study design yields repeated measures data nested within subjects. In repeated measures, the lowest level represents the measurement level (e.g. assessments of patients’ symptoms with the telemetry test battery) grouped by a measurement unit (e.g. patients) which in turn has multiple measurements at multiple test occasions. The two main challenges when analysing repeated measures data relate to the fact that i) there may be unbalanced data design due to incomplete follow-up or drop-out of subjects in long-term studies and ii) they are correlated rather than independent. With longitudinal data, between-subject variability is to be non-ignorable compared to within-subject variability. For all these reasons, the employment of special statistical techniques which are robust to unbalanced study designs and can account for intra-subject correlation of measurements is necessary in order to derive valid and reliable conclusions from the data. An instance of mixed-effects models are linear mixed-effects (LME) models (Pinheiro and Bates, 2000). The
LME models make specific assumptions about within- and between-subject variations by including a special term called a *random effect* which represents subject-specific influences in his/her measurements which are not directly captured by *fixed effects* (or sample effects). In most of the longitudinal studies, the aim is to characterize the overall sample trend as well as the magnitude of between-subject variation.

Let $Y_{ij}$ denote the measured outcome for measurement $j$ on subject $i$ at time $t_{ij}$ where $i = 1, 2, \ldots N$ and $j = 1, 2, \ldots n$. Let $X_{ij}$ denote covariates associated with the outcome $Y_{ij}$. These covariates can be subject-dependent or time-dependent. In typical longitudinal datasets, there are both individual- and sample-specific intercepts and slopes. For example, let a line that characterizes a linear trend of a subject $i$ be defined as:

$$Y_{i,j} = \beta_{i,0} + \beta_{i,1} \cdot X_{i,j} + \varepsilon_{i,j} \quad (17)$$

where $\beta_{i,0}$ and $\beta_{i,1}$ represent the individual intercept and slope, respectively and $\varepsilon_{i,j}$ represents the error term assumed to be normally distributed. The within-subject variability is defined as the deviation of individual measurements to the individual’s line, using the following equation:

$$\varepsilon_{i,j} = Y_{i,j} - (\beta_{i,0} + \beta_{i,1} \cdot X_{i,j}) \quad (18)$$

On the other hand, the between-subject variability is defined by the variation among both the intercepts $\beta_{i,0}$ and slopes $\beta_{i,1}$. Going further, the LME model is then defined as:

$$Y_{i,j} = \beta_{0} + \beta_{1} \cdot X_{i,j} + b_{i,0} + \varepsilon_{i,j} \quad (19)$$

where $\beta_{0}$ represents the sample average intercept, $\beta_{1}$ represents the sample average slope, $b_{i,0}$ represents the deviation from the sample average intercept ($b_{i,0} = \beta_{i,0} - \beta_{0}$) assumed to be normally distributed.

The LME models were used in three papers (Paper II, Paper III and Paper VI) to assess changes in mean computed scores over time as well as among other covariates such as subject groups and self-assessment categories. In all these cases, the patient/subject ID variable was regarded as a random effect.
4 Quantification and analysis of fine motor performance

This chapter summarizes the three papers (Paper I-Paper III), illuminating the first research theme in terms of motivation, objectives and methodology.

4.1 Paper I – A new computer method for assessing drawing impairment in Parkinson’s disease

Patients suffering from PD may exhibit different types of impairments which can be very disabling during ADLs. These impairments are related to upper limb involuntary movements as a result of tremor and bradykinesia as well as therapy-related complications such as dyskinesias. A popular approach to assess the severity of the impairments is to ask patients to draw spirals (spirography) on papers which in turn are visually assessed by clinicians (Bain and Findley, 1993). In clinical examinations, UPDRS is widely used to measure the severity of these impairments. However, given the fact that the UPDRS is highly-subjective and time consuming (Martinez-Martin et al., 1994) and paper collection schemes have problems with compliance and reliability (Stone et al., 2003; Broderick, 2008), more objective methods of severity assessment are needed. These objective methods should be designed and tailored to assess symptoms based on remote data gathered in home environment settings where there is no clinical supervision.

The objective of Paper I was to develop and evaluate a method for quantifying the severity of the PD-related impairments being represented in spirography tasks of the telemetry device. The aim was to derive an objective measure which can be used to document the kinematics and dynamics of fluctuating symptoms during spirography. This measure related to the severity of impairments irrespective of their cause i.e. it just reflected the problem in upper limb motor function regardless of the cause.

- Visual assessment of drawing impairment

A separate web application was constructed to display static images of spiral drawings and to allow users (PD specialists) to rate observed drawing impairment. The application retrieved paired x and y coordinates of the drawn spirals from Structured Query Language (SQL) database tables which stored spiral information. Along with the spiral drawing image, the drawing completion time (in seconds) and the self-assessed motor state at the time of the particular test were retrieved and displayed. The application was organized in four “tracks”: “preliminary rating”, “training”, “standardised rating”, and “rater agreements”. Two raters studied the examples
in the clinical handbook for assessment of tremor in spiral drawings which was based on a 10-point scale (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 drawing impairment in spiral drawings.

First a rating of the overall drawing impairment was done and secondly a probable cause (tremor, bradykinesia or dyskinesia) for the given impairment was marked. The general impression of the shape of the spirals determined the level of severity; homogeneous 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.

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. Both raters then observed these preliminary ratings and used them as templates for rating spirals in other tracks. In the “standardised rating” track, the application displayed all three spiral drawings from three randomly selected test occasions per patient. A ‘standardised manual rating’ (SMR) score 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 from those in the “standardised rating” track) per patient were randomly displayed.

- Computerized assessment of drawing impairment

The development of the method consisted of first transforming Cartesian coordinates to polar coordinates followed by feature extraction using DWT, dimension reduction using PCA and finally scaling of the score. The development of the method was based on the data from the “preliminary rating” track which was considered as a training set whereas evaluation of the method was performed on the SMR ratings dataset which was considered to be the testing set.

The spiral data consisted of x (horizontal position) and y (vertical position) coordinates. In order to perform quantitative evaluation, the coordinates were transformed into polar coordinates, more specifically into radius (i.e. square root of the sum of squares of the coordinates) to represent the degree to which the spiral drawing spatially deviated from the pre-drawn spiral. In order to avoid onset effects, validation of input data was performed so that only those spiral drawings that contained more than 50 data points were considered and processed.

A 3-level decomposition using a Daubechies (db10) (Daubechies, 1988) wavelet function family was performed on the radius time series signal to
obtain A and D coefficients. A feature vector, containing 265 coefficients was obtained after D coefficients ($D_n, n = 1, 2, 3$) at the first, second and third levels (128+64+32) and $A_3$ coefficients at the third level (32) were appended to each other in order of descending levels, i.e. $A_3, D_3, D_2$ and $D_1$ (Figure 2).

In order to reduce the dimensionality of these coefficients, the PCA was applied. Initially, the PCA was applied on a subset of data, preselected on the basis of the 10% worst and 10% best tapping results (speed and accuracy). This was done with the aim of better representing the severity of PD symptoms by finding a desired direction in multidimensional feature space by selecting the extreme cases. Secondly, the PC1 was then defined as a linear combination of the weights derived from the subset of data and the DWT coefficients of the full dataset.

The PC1 was then finally 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. In order to reduce systematic errors, such as over-prediction of low impairments or under-prediction of high impairments, a DDM approach was considered by employing a linear regression method to map manually rated spirals from the “preliminary rating” track and scaled logarithmic values of the PC1. The resulting spiral score is hence denoted on the ‘wavelet spiral test score’ (WSTS).

4.2 Paper II – Automatic and objective assessment of alternating tapping performance in Parkinson’s disease

In addition to spirography, the other approach to assess upper limb motor function of PD patients relates to finger tapping. In clinical settings, the most common strategy is to employ items 23-25 (finger tapping, hand movements, rapid alternating of movements, respectively) of the UPDRS-part III for quantifying the severity of upper limb motor symptoms (Fahn et al., 1987). This part of UPDRS consists of different dimensions each of which are scored on a five-point scale ranging from 0 (normal) to 4 (severe). Since the in-clinic UPDRS provides only a snapshot of symptom severity and fluctuations and is highly subjective, there is a need for objective methods that provide a means for long-term and repeated assessments and have a better resolution. These methods are sensitive to subtle symptom changes and can be used to track kinematic and dynamic performance during tapping tests.

The objective of Paper II was to develop and evaluate a method for enabling quantitative and automatic scoring of alternating tapping performance (ATP) of PD patients. The aim was to derive objective measures of
ATP by first deriving four tapping dimensions: ‘speed’, ‘accuracy’, ‘fatigue’ and ‘arrhythmia’, and a global tapping severity (GTS) score. The paper reported on different metrics to evaluate the quality of assessments of the method including correlations to visually assessed scores of ATP and UPDRS motor ratings, reliability, and sensitivity to treatment interventions and natural PD progression over time. In addition, the ability of the method to discriminate between healthy elderly subjects and patients in different stages of the disease was reported.

The four dimensions were considered specific for the type of movement disorder found in PD patients, as defined by the items 23-25 of the UPDRS (Fahn et al., 1987). The ‘speed’ dimension measured the ability to tap rapidly during the test. The subject’s ability to correctly tap the fields in the touch screen was measured by the ‘accuracy’ dimension. The amount of tapping irregularity and the progressive reduction of movements across the test trial were measured by ‘arrhythmia’ and ‘fatigue’ dimensions, respectively. The GTS was assumed to be a composite score of the four dimensions providing a holistic representation of the patient’s ATP during the tapping test. To avoid onset and offset effects, data points collected during the first and last 2 seconds of the test time were discarded. Hence, the time series of interest were in the range between 2-18 seconds.

- **Visual assessment of ATP**

A web application was developed to visualize the performance of patients during tapping tests and allow users (PD specialists) to rate different tapping impairments (Memedi et al., 2013). The system was designed as a three-tier web application using JavaServer Pages and MySQL Server as a back-end database. The system retrieved time series of raw data from the database tables and visually depicted them in different types of graphs (Figure 3). Information presented included i) distribution of taps over the two fields, ii) horizontal tap distance vs. time, iii) vertical tap distance vs. time, and iv) tapping reaction time over the test length. A neurologist was instructed first to visually interpret the tapping variation, patterns and trends within the graphs and then to assess the observed impairments on a categorical scale ranging from 0 (normal) to 4 (extremely severe). First ratings of the four dimensions were done followed by rating the GTS. The ratings included at least 20 test occasions per each GTS level. The visually assessed scores are hereafter denoted as V-SPEED, V-ACCURACY, V-FATIGUE, V-ARRHYTHMIA and V-GTS.

- **Computerized assessment of ATP**

In total, 24 quantitative measures were extracted from time series data to represent the severity of the patients’ symptoms while performing the tapping tests, using time series analysis and machine learning methods. The
data were summarized into scores for the four tapping dimensions and the GTS.

To quantify the speed performance during the tapping tests, the following measures were calculated and used subsequently during analysis: i) total number of taps, calculated as the total sum of taps in a test occasion for the mid 16s, ii) mean tapping speed (MTS), defined as the mean rate of change of tap distance with time, iii) MTS from left to right, iv) coefficient of variation (CV) of tapping speed from left to right, v) MTS from right to left and vi) CV of tapping speed from right to left. PCA using correlation matrix method was then applied to these 6 measures to reduce their dimensions and obtain a single measure represented by PC1. The PC1 of the measures accounted for 69% of the total variance in the data and was used to represent the automated speed score (A-SPEED).

To quantify the accuracy during tapping, the following measures were calculated: i) mean distance from the center fields (for taps that were tapped within the area of the fields, the distance was preset to zero), ii) CV of distances from the center fields, iii) the overall distribution of taps, calculated as mean variation (ratio between summed distance and total number of taps) of the two fields and iv) overall tapping precision, calculated as the mean distance from center fields, irrespective of whether the taps were inside or outside the field areas, divided by the total number of taps. After applying PCA to these 4 measures, the PC1 accounted for 65% of the variance in the data and was used to represent the automated accuracy score (A-ACCURACY).

The following measures were defined to quantify the fatigue dimension: i) MTS per cycle (MTSPC); a cycle was defined as the movement from one field to the other and backwards, ii) the absolute mean difference of the change in time ($\Delta t$) of the first (2-10s) and the second (10-18s) part of the time series signal, iii) the absolute mean difference of MTSPC of the first and second parts of the time series signal, iv) the absolute mean difference of the ApEn values of $\Delta t$ of the first and second parts of the time series signal, v) the absolute mean difference of the DTW values of the tapping speed of the first and second parts of the time series signal, vi) the absolute mean difference of the DTW values of $\Delta t$ of the first and second parts of the time series signal and vii) the mean correlation coefficient of the jackknifed samples of $\Delta t$ and the corresponding timestamp sequences. After PCA on these 7 measures, the PC1 accounted for 35% of the variance in the data and was used to represent the automated fatigue score (A-FATIGUE).
In order to quantitatively measure arrhythmia during alternating tapping, the following measures were calculated: i) the ApEn value of the time series of the tapping speed (the rate of change of the tap distance with time), ii) the ApEn value of the y-coordinate time series, iii) the standard deviation of the distance variations of the taps, iv) the mean of both the distance and time variations of the taps, v) the standard deviation of both the distance and time variations of the taps, vi) the mean cross-correlation coefficient of
an artificial perfectly-periodic and original tapping time series and vii) the Cross-ApEn value of an artificial perfectly-periodic and original tapping time series. The PCA was applied to these 7 measures and the PC1 accounted for 35% of the variance in the data and was used to represent the automated arrhythmia score (A-ARRHYTHMIA).

In order to classify the ATP tests based on the 5 GTS levels, a LR model was used to map the extracted quantitative measures to the corresponding V-GTS scores. Initially, the PCA was applied to all of the 24 measures in order to reduce their dimensions and to obtain a smaller set of uncorrelated measures (i.e. PCs) which can be used as independent variables. The appropriate number of “significant” PCs was decided by selecting a cumulative percentage of total variation for which it was desired that the selected PCs should account for more than 70% of the total variance in the data. Applying this criterion resulted in retention of the first 5 PCs. The PC1 of the 24 quantitative measures is hence denoted on First PC. A stratified 10-fold cross-validation was applied to assess the generalization ability of the LR model to future independent datasets. The output of the LR classifier was used to represent the automated GTS score (A-GTS).

4.3 Paper III – Spiral drawing during self-rated dyskinesia is more impaired than during self-rated off

In clinical settings, rating scales of impairment and disability are the most commonly used outcome measures. However, they are considered mainly as physician-oriented since they partially represent the patient’s QoL which is affected by the complex nature of the disease including different motor and non-motor symptoms. Therefore, in addition to physicians’ perspectives there is a need to take into account the patients’ perspectives towards their symptoms in order to capture a broader picture of general patient’s health. PD patients may have difficulties in self-assessing their disability in relation to assessments of daily function (Shulman et al., 2006) and executive functions (Koerts et al., 2011), as well as have difficulties in recognizing their treatment-related motor complications (Vitale et al., 2001). However, advanced PD patients treated with LCIG are often well aware of their motor states in relation to physicians’ assessments (Nyholm et al., 2012b). In addition to patient- and physician-oriented outcomes, objective measures of a patient’s motor functions can offer complementary information.

The objective of Paper III was to examine repeated measures of fine motor function in relation to self-assessed motor conditions in PD. More specifically, the aim was to investigate the severity of objective assessments of
spiral drawing and tapping performance relative to self-assessed motor conditions (On, Off and dyskinesia) in advanced PD patients.
5 Methods and systems for remote and long-term assessment of symptoms

This chapter summarizes the three papers (Paper IV-Paper VI), illuminating the second research theme in terms of motivation, objectives and methodology.

5.1 Paper IV – Combined fine-motor tests and self-assessments for remote detection of motor fluctuations

The PD is progressive and its nature is complex and multidimensional that includes a wide variety of symptoms ranging from those related to cognition to gross motor function. For all these reasons, there is a need for constant assessing and following-up of symptoms and treatments over time. Any treatment would be considered to have a positive effect if it simultaneously contributes to improvement of both motor and non-motor symptoms. The quantitative assessment tools should consist of repeated measures of both subjective and objective information in order to provide a more in-depth assessment of the patient’s general health, motor and non-motor symptom fluctuations and treatment effects. Combining subjective and observer-independent objective measures into composite scores provides a means for cross-evaluation of patients’ perceptions of their symptoms and their actual physiological functioning.

The main objective of Paper IV was to investigate the approaches to data collection and processing of previously patented IT-based systems designed for telemedicine applications. A secondary objective of the paper was to summarize the development of the presented IT-based system for remote monitoring of PD symptoms.

5.2 Paper V – A web application for follow-up of results from a mobile device test battery for Parkinson’s disease patients

Computer systems provide a means for data acquisition and transmission, data storage and retrieval, and data processing and presentation (van Bemmelen and Musen, 1997). According to Shim et al. (2002), CDSS should comprise components for i) database management capabilities with access to internal and external data, information, and knowledge, ii) modelling functions accessed by a model management system, and iii) user interface designs that enable interactive queries, reporting and graphing functions. In addition to data processing which is essential for deriving semantic information from the raw data, adequate presentation and visualization of information to the end-users is essential for revealing new insights and
knowledge about the overall fluctuation of the patient, symptom states and the progress of the disease itself. During visualization of clinical data, the simplest approach is to present textual and numerical data by using lists and tables, for example by presenting quantitative raw data as averages over time. However, in order to facilitate a rapid and easy interpretation of subtle changes and overall trends of symptoms, a graphical representation of data should be considered for example by using more advanced graphing functions such as plots and charts.

The objective of Paper V was to describe a web-based system for enabling objective, visual and longitudinal representation of symptoms to support clinicians in monitoring and treating their PD patients. The aim was to develop separate software modules for processing of time series data, collected with the telemetry test battery, and present summaries to the end-users in a graphical manner. The data processing part focused on developing a data-driven approach to combine subjective and objective measures of the telemetry test battery into scores for representing the severity of a patient’s symptoms over week-long test periods. These scores related to six conceptual symptom dimensions (4 subjectively-reported and 2 objectively-measured) and an overall test score (OTS) that reflects the patient’s global health condition and disability over a test period.

- **Calculation of symptom dimensions**

Some of the items of the telemetry test battery were highly correlated indicating that they measured the same concept. Because of this redundancy, it was possible to reduce these items into a smaller number of PCs for representing the conceptual symptom dimensions. According to the information content of the test battery, a test period can be described by the following six dimensions: ‘walking’ (based on q1), ‘satisfaction’ (q6), ‘dyskinesia’ (q2-dyskinetic and q4), ‘off’ (q2-Off and q3), ‘tapping’ (q8-q11), and ‘spiral’ (q12-q14).

Initially, time series of the test battery were summarized by calculating test period statistical features, such as the level (MEAN), fluctuation (standard deviation, SD) and the mean squared deviation (MSD) from the best answer alternative of q1, q3, q4 and q6. In total, 28 features were derived which were then used in subsequent analysis. The rational to select these features was to define scores that take into account the intensity, frequency and importance of occurring symptoms. Mean values are the obvious choices to represent levels in time series whereas standard deviations are obvious for representing overall variation. To have at least three variables in a dimension reduction data analysis, MSD was also used. For each dimension, PCA was then applied to the features for the items that the dimensions were based on by retaining the PC1s. Each dimension was rep-
resented as a linear combination of the three features (after normalized to zero mean and one standard deviation) and the weights derived from the PCA. Linear transformations were then finally used to scale dimensions to a scale from 0 (worst) to 1 (best), based on their minimum and maximum values.

- **Calculation of OTS**

  Using the six derived dimensions, the OTS can be calculated as an unweighted score by placing them in a regular hexagon where each one of them becomes a corner. According to domain experts the geometrical placement of the dimensions should resemble Figure 4 (lower left corner). A problem with this approach is that all six dimensions have the same weight in the assessment of the overall condition of the patient which is in contradiction to the common 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 OTS was to weight the dimensions using the UPDRS ratings. The standard least-squares MLR was used to examine the relationships between PC1s of the six dimensions and the patient’s UPDRS scores given by the neurologists. The OTS was then defined as a linear combination of the PC1s of the dimensions and the weights estimated by MLR method in which total UPDRS was considered as a dependent variable. The OTS was then scaled to a scale from 0 (worst) to 1 (best), using linear transformations.

- **System description**

  The data processing is handled by a software module called the data processing sub system (DPSS) and the data presentation is handled by a web application (WA).

  o **Data processing sub system**

    The DPSS is a stand-alone application, which incorporates knowledge to analyse and interpret the raw test battery data. The interpretation is done based on the equations derived for calculating the dimensions and the OTS. This module parses, processes and stores the raw and summarized data into relational database tables to be accessed later by the WA. A connection with the RDM software is first established followed by receiving and parsing of XML data from files. The DPSS was written in the VB.NET programming language and designed to run with a personal computer-compatible machine. For computing the WSTS scores, a single piece of M-code was written in Matlab® (MathWorks Inc.) which was then encrypted and wrapped into a C# interface by using the Matlab Builder for .NET. This piece of code is accessed by the DPSS.

  o **Web application**
The WA is a feedback system comprising a secure web server and a database with web-based access for medical staff (Figure 4). The main role of the WA in the overall system was to present graphical summaries of test results to the end-user clients. To enable rapid patient status assessment, the information in the WA was displayed and ordered using a top-down approach where the general overview of the patient’s performance is highlighted. Raw data and other detailed information may be accessed in more “advanced” displays.

5.3 Paper VI – Self-assessments and motor tests via telemetry in a 36-month levodopa-carbidopa intestinal gel infusion trial

Understanding the relationships between objective measures of physiological processes and clinical outcomes of severity, e.g. UPDRS, is essential in understanding the nature and progression of the disease. As in the case of outcomes derived by clinical rating scales, outcome measures derived by computer assessment tools should be scientifically sound in terms of validity (i.e. the extent whether they measure what they are supposed to measure), reliability (i.e. internal consistency of sub-items and test-retest reliability of results) and sensitivity to change (i.e. sensitivity to treatment changes and natural disease progression) (Haubenberger et al., 2011; Maetzler et al., 2013; Horak and Mancini, 2013).

The objective of Paper VI was to investigate whether computed scores, which were derived in Paper V, can be used to measure effects of PD treatment intervention and disease progression. In particular, the aims of the paper were i) to determine if PD progression over time can be followed in computed scores, ii) to assess sensitivity of the computed scores in relation to treatment interventions and iii) to assess correlations between computed and clinical scores.

In 20 LCIG-naïve patients, telemetry assessments with the test battery were available during oral treatment and at least one test period after having started LCIG treatment. Three LCIG-naïve patients did not use the test battery at baseline but had at least one test period of assessments thereafter. Hence, n = 23 in the LCIG-naïve group and n = 35 in the LCIG-non-naïve group.
Figure 4. Patient status report in WA with graphical visualization of time series data. The “Overall Test Score” graph presents a longitudinal view of OTS over the test periods. The hexagon graphs represent the severity of symptoms in the six dimensions for two test periods, i.e. for the first and last periods. Another graph shows the trend of statistical summaries (mean ± SD) of test responses throughout the day (8-, 12-, 16- and 20-o’clock) for a selected test period. These graphs can be updated interchangeably.
6 Results

6.1 Paper I

Test-retest reliability was assessed by Spearman rank correlations by taking the mean of all three possible correlations of the three spiral drawings per test occasion. Between-rater agreements were assessed by Spearman rank correlations, using ratings from the “rater agreements” track. In order to avoid the problem of multiple test occasions per patient, 200 random samples of single test occasions per patient were drawn and mean correlations in these samples were calculated. The Bland-Altman analysis of difference (Bland and Altman, 1986) was used to estimate the prediction error of the WSTS versus SMR. Correlations between WSTS, SMR and the tapping test results were assessed on test occasion level by taking mean values of the three spiral drawings. Correlations with UPDRS were assessed after taking mean values of all (approximately 3×28) of the spirals drawn during each weeklong test period.

Test-retest reliability coefficients were as follows: 0.77 for WSTS, 0.71 for the first rater and 0.7 for the second rater. Correlations between WSTS and SMR were strong (0.89) whereas correlations between WSTS and objective measures of tapping performance were low to moderate (Table 2). The 95% confidence interval (CI) for the prediction error of the WSTS was ±1.5 units with a mean value of 0.39. Rater agreements were good with a 0.87 correlation coefficient. Correlations between WSTS and UPDRS scores were as follows: 0.41 to total UPDRS, 0.51 to UPDRS-part II, and 0.38 to UPDRS-part III.

<table>
<thead>
<tr>
<th></th>
<th>WSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMR</td>
<td>0.89</td>
</tr>
<tr>
<td>Total UPDRS</td>
<td>0.41</td>
</tr>
<tr>
<td>Average tapping speed (q8 and q9)</td>
<td>-0.40</td>
</tr>
<tr>
<td>Average tapping accuracy (q8 and q9)</td>
<td>-0.45</td>
</tr>
<tr>
<td>Tapping Speed (q10)</td>
<td>-0.56</td>
</tr>
<tr>
<td>Tapping Speed (q11)</td>
<td>-0.52</td>
</tr>
<tr>
<td>Tapping Accuracy (q11)</td>
<td>-0.50</td>
</tr>
</tbody>
</table>

Table 2. Mean Spearman correlations after repeated random sampling of one occasion per patient.

6.2 Paper II

Agreements between V-GTS and A-GTS were evaluated using the AUC and weighted Kappa statistics as major performance evaluation metrics. Spearman’s rank correlation coefficients were used for assessing linear relationships
between computed and visual scores. Reliability i.e. internal consistency of the four tapping dimensions was assessed using Cronbach’s \( \alpha \) test. Sensitivity to treatment interventions and disease progression over time was assessed by evaluating changes in mean automated dimension scores of LCIG-naïve patients over time i.e. at baseline and follow-up test periods, with LME models with a patient ID as a random effect and test period as a fixed effect of interest. The LME models were also used to i) assess the ability to discriminate between healthy elderly subjects and the two patient groups, with subject ID as a random effect and group as a fixed effect of interest and ii) assess differences in mean scores of the First PC relative to categories of the items 23-25 of the UPDRS, with the patient ID as a random effect and category as a fixed effect of interest. Tukey post-hoc comparison tests were performed to determine differences between subject groups. Inter-subject variability of the automated dimension scores was assessed using ICCs.

The agreements between V-GTS and A-GTS were very good with a Kappa coefficient of 0.87 (\( p<0.001 \)) and weighted AUC value of 0.86. The best agreements were seen at extreme classes i.e. class 0 (normal, AUC=0.93) and class 4 (extremely severe, AUC=0.95). Correlations between computed and visual scores were adequate. The mean scores of the First PC for each category of items 23-25 of the UPDRS are displayed in Figure 5. The internal consistency among the four dimensions was acceptable (Cronbach’s \( \alpha=0.75 \)), indicating that they measure the same underlying construct of the ATP. Mean A-SPEED, A-FATIGUE and A-GTS scores improved to the first test period on LCIG treatment and this improvement remained statistically significant until month 24 (\( p<0.001 \)). The mean scores of A-ARRHYTHMIA dimension deteriorated but only at different test periods. In contrast, mean A-ACCURACY scores deteriorated throughout the study period along with the natural disease progression. Figure 6 summarizes the mean scores of the four automated dimensions for each subject group. Significant differences were found between HE subjects and patient groups in all four dimensions. In general, HE subjects had better scores in the tapping dimensions than advanced patients. In general, the variability of tapping dimensions in advanced Swedish patients was generally greater than that observed in HE subjects.

6.3 Paper III

Three self-assessed (SA) motor states were derived after the self-assessment answers of q7 were pooled as follows: SA:Off (including -3; very Off and -2; moderately Off), SA:On (including -1; slightly Off, 0; On and +1; slightly On), and SA:Dys (including +2; moderately dyskinetic and +3; very dys-
kinetic). For the spiral tests, a ‘test occasion spiral score’ (TOSS) was defined as the mean of the three WSTS scores per test occasion. For the tapping tests, the mean speed (number of taps per 20s) and accuracy (percentage of correct taps) for the right hand were calculated and used in the subsequent analysis. Differences in the mean values of the motor test results among motor self-assessments were analysed using the LME models. The patient ID variable was regarded as a random effect and SA states as fixed effects. The dependent variables were the motor test results: mean tapping speed and accuracy, and TOSS.

Figure 5. Mean scores of First PC for each category of items #23 (Finger Tapping), #24 (Hand Movements) and #25 (Rapid Alternating Movements of Hands) of the UPDRS scale, corrected for individual subject variation using linear-mixed effects models. Y-axis: a high score means good function. UPDRS items: 0 (Normal), 1 (Mild slowing and/or reduction in amplitude), 2 (Moderately impaired), 3 (Severely impaired) and 4 (Can barely perform the task). P-values and % changes are shown with respect to category 0. ** = p < 0.01 and *** = p < 0.001. Abbreviation: n.s., not significant. UPDRS item #23: 0 (n=2323), 1 (n=3649), 2 (n=2793), 3 (n=687), 4 (n=25). UPDRS item #24: 0 (n=2532), 1 (n=3964), 2 (n=2441), 3 (n=491), 4 (n=49). UPDRS item #25: 0 (n=2201), 1 (n=4085), 2 (n=2511), 3 (n=598), 4 (n=82). Note that the category 4 of the UPDRS items had small number of observations that were assessed in very few patients.
For the Swedish study, the mean TOSS score differed significantly across the three SA motor states (p<0.001, ANOVA) with a mean value (for all states combined) of 4.46 [Wald-type approximate 95% CI; 4.14, 4.77]. The worst mean TOSS score was found in SA:Dys with 0.43 [0.35, 0.51] units higher than the SA:Off (p<0.001). However, the mean TOSS score did not differ significantly between SA:On and SA:Off (p=0.16). Both tapping speed and accuracy scores were found to be the highest in the SA:On. The tapping speed was 2.24 [1.3, 3.18] units (#taps/20s) higher in the SA:Dys than in the SA:Off (p<0.001). In contrast, tapping accuracy was better in the SA:Off than in the SA:Dys with 1.79 [0.39, 3.2] units (percentage accurate taps) difference (p<0.005). In the Italian study, patients in the F group had trends similar to those of the Swedish patients regarding motor test results over SA motor states. However, there were no significant differences between the three SA motor states for the three types of motor test results. Figure 7 shows the fixed effects coefficients representing the mean motor scores for each individual self-assessment category for the Swedish and Italian studies.

6.4 Paper IV
The patent search included the use of a number of keywords such as “subjective”, “objective”, “patient”, “remote”, “monitoring”, “system”, “movement disorders”, “Parkinson’s” and combinations of these terms in patent search engines of WIPO and Google. Only those patents that were filed after year 2000 were included in the review. The findings from the review included that there were a small number of systems which collected subjective and objective measures in telemedicine settings. In contrast to the presented IT-based system, there was a lack of systems that collected and combined these two types of information into scores which can provide a more general overview of the patients’ health status.
6.5 Paper V

To evaluate whether the six dimensions measure the same construct of a patient’s health status the Cronbach’s $\alpha$ test was used. To estimate the amount of contributions of dimensions to the OTS, the MLR regression coefficients were analysed. The evaluation of the system was done in two stages: first presenting and demonstrating its functionalities to an advisory board consisting of neurologists and to nurses who had experience using the telemetry device during the clinical study period. A second evaluation determined the level to which nurses were satisfied with the usability of the system by administering a standard questionnaire.

The internal consistency of the six dimensions was good with a Cronbach’s $\alpha$ coefficient of 0.81. The objectively-measured dimensions i.e. tapping and spiral were given a high weight, just as the motor section (part III) has the highest weight in the UPDRS scale.

The advisory board consisted of 14 neurologists 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. To a question about their overall impression of the system, eleven
had a positive impression, one had a neutral impression and two had a negative impression.

When demonstrating the functionalities of the system to the nurses who took part in the clinical study at least data for one patient per clinic was shown. The responses were summarized in a qualitative manner as follows: i) the WA is very useful, ii) the results during the test periods showed agreement with the qualitative observations of the patient during that test period, e.g. “one patient was in a bad condition at baseline, he improved after starting LCIG, then he became worse again, 24-h infusion started and the patient became better again; we can clearly follow this change in the WA”, iii) comparisons between patients are possible (one patient is in better/worse condition than another).

One year later, the IBM Computer System Usability Questionnaire (CSUQ) was administered to evaluate the nurses’ satisfaction with the WA. Initially, they were asked to perform a series of tasks using the WA, such as i) login, ii) select a patient, iii) check the patient’s performance by looking at the graphs of summarized scores and compare these results with own clinical observations, iv) select and check the other patients’ results by repeating steps ii) and iii), v) complete the survey and vi) 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 consisted of 19 items graded on a seven-point Likert rating scale ranging from 1, “strongly agree” to 7, “strongly disagree” (Likert, 1932). Four usability scores were derived by averaging responses to the CSUQ items: Overall Satisfaction, System Usefulness, Information Quality and Interface Quality (Lewis, 1995). Data about gender, age and previous experience with computer applications were noted. 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 aged between 38 and 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.
Figure 7. LME fixed effects coefficients of the motor test results over the SA categories for the two studies (solid line: Swedish study; dashed line: Italian F; dotted line: Italian S). Self-assessments: -3=‘very Off’, 0=‘On’, 3=‘very dyskinetic’. TOSS is the mean WSTS score where high score means severe spiral drawing impairment. Tapping speed (#taps/20s) and accuracy (percent correct taps) are shown for the right hand. Swedish study: -3 (n=523), -2 (n=1214), -1 (n=2233), 0 (n=2436), 1 (n=2597), 2 (n=863), 3 (n=213). Italian F: -3 (n=21), -2 (n=44), -1 (n=191), 0 (n=384), 1 (n=100), 2 (n=69), 3 (n=13). Italian S: -2 (n=21), -1 (n=100), 0 (n=670), 1 (n=14). SA categories with less than 10 observations were not included in the analysis. Note that the extreme categories in the Italian groups had a small number of responses that were reported by very few patients.

6.6 Paper VI
Change in the computed scores over time, that is at baseline and follow-up test periods, was assessed using LME models using a restricted maximum likelihood estimation method with patient ID as a random effect and test period as a fixed effect of interest. The dependent variables were the test battery scores; the raw test battery items, the six symptom dimensions and the OTS. The within-patient variability of OTS was assessed using ICC, for the two patient groups. Correlations between the six symptom dimensions and OTS and clinical rating scores were assessed using Spearman’s rank correlation coefficients.

The mean OTS scores at baseline were 0.47 and 0.51 for LCIG-naïve and LCIG-non-naïve patients, respectively (Figure 8). In LCIG-naïve patients, the mean OTS improved to the first test period on LCIG treatment and this improvement remained statistically significant until month 24. The maximum improvement, 0.15 units (32%) higher than at baseline (p<0.001), was seen at month 3. The mean scores of all subjectively-reported dimensions improved significantly throughout the course of the study, except ‘walking’ at month 36 (p=0.41, n=4). However, there were no significant differences in the mean scores of objectively-measured dimensions between baseline and other test periods, except improved ‘tapping’ at month 6 and month 36, and ‘spiral’ at month 3 (p<0.05).
There were no significant changes in the mean scores of the LCIG-non-naïve patients over the course of the study, except for worse mean OTS at month 36 (p<0.01, n=16) (Figure 8). The LCIG-naïve patients had a higher within-subject variability in their OTS scores (ICC=0.67) compared to LCIG-non-naïve patients (ICC=0.71). The mean scores of raw question items (q1-q6) in LCIG-naïve patients improved significantly, except for q5 (perceived painful cramps) at month 24 (p=0.18). All changes were seen at the first LCIG period and were sustained over the study period. Tapping measures of speed improved starting from the first test period on LCIG treatment and the changes remained significant to month 30. In contrast, tapping measures of accuracy were not improved by LCIG treatment. The trend of TOSS scores indicated that the drawing performance of the patients started to get worse from month 6 on LCIG-treatment.

The correlations (absolute values) between OTS and clinical ratings in the full sample were as follows: total UPDRS (0.59), total PDQ-39 (0.59), UPDRS-part II (0.64) and UPDRS-part III (0.48). Among subjectively-reported dimensions, ‘dyskinesia’ had the strongest correlation to both total UPDRS and UPDRS-part III with coefficients of 0.41 and 0.29, respectively. ‘walking’ had the strongest correlations to total PDQ-39 (0.49) and to UPDRS-part II (0.58). Among objectively-measured dimensions, ‘tapping’ had the strongest correlations to total UPDRS (0.53), total PDQ-39 (0.52), UPDRS-part II (0.58) and UPDRS-part III (0.43).
Figure 8. Trends of mean OTS scores of LCIG-naïve and non-naïve patients over the 36-month study period. A high OTS score means good function. P-values are shown with respect to initial test periods. * = p<0.05, ** = p<0.01, *** = p<0.001. Abbreviation: n.s., not significant. Test period (LCIG-naïve): -3, baseline (n=20); 0 (n=19); 3 (n=19); 6 (n=16); 9 (n=16); 12 (n=16); 18 (n=15); 24 (n=13); 30 (n=9); 36 (n=4). Test period (LCIG-non-naïve): 0 (n=35); 3 (n=32); 6 (n=30); 9 (n=26); 12 (n=25); 18 (n=24); 24 (n=20); 30 (n=14); 36 (n=16).
7 Conclusions
This chapter first summarizes and discusses the work in terms of main contributions and answers to the research questions. Secondly, a discussion follows which points out the general implications of the approaches and results of the thesis as well as the limitations for each appended paper. Finally, directions for future research and concluding remarks are given.

7.1 Summary and Discussion
Methods for measuring the severity of symptoms and treatment-related complications are crucial for effective clinical management of PD. A challenge is the large within- and between-patient variability in symptom profiles as well as the emergence of motor complications, i.e. swings between On, Off and dyskinesia, which represent a significant source of disability in patients (Thanvi et al., 2007; Pechevis et al., 2005). The current state of the art in clinical studies is to use the UPDRS scale. However, as it was stressed earlier, the use of this scale is accompanied with a number of limitations including large within- and between-clinician variability in ratings, low resolution assessments, the need for a clinician’s presence, and others (Stocchi et al., 1986; Taylor Tavares et al., 2005; MDSTFRSPD, 2003; Hagel et al., 2003; Haaxma et al., 2008).

In the work reported here, the use of statistical and machine learning methods coupled with mobile computing and Web technology was explored in order to address the above mentioned limitations. The aim was to develop a system architecture, consisting of computer-based assessment methods and the software around this platform, for providing an effective alternative to management of PD. In general, IT relates to the use of hardware and software in an effort to collect, store, transmit, process, manipulate, and present data and information. An important aspect of IT is how systems are designed and developed to provide useful applications and solutions. The main focus of this thesis was on the last three stages i.e. data processing, data manipulation and data presentation.

One part of the thesis focused on investigating and developing methods for quantifying the severity of PD-related impairments during fine motor tests (spiral drawing and alternating tapping). This work included the use of digitized movement data in combination with time series analysis and DDM for automating the process of scoring the severity of symptoms.

The methods presented in Paper I and Paper II were evaluated as being feasible approaches for quantitative and objective assessment of the severity of fine motor impairments of PD patients. The methods were developed using the research approach presented in section 1.3 where the
aim was to derive methods which are designed and tailored for long-term telemetering of symptoms as well as methods which have good level of interpretability. In agreement with several previous publications, the methods had good validity, reliability and sensitivity to treatment changes (Pullman et al., 2008; Goetz et al., 2009; Shima et al., 2009; Yokoe et al., 2009; Haubengerger et al., 2011). In addition to these metrics, the method presented in Paper II had good ability to discriminate between healthy elderly subjects and patients in different stages of the disease. Both methods provided means for deriving objective measures of fine motor severity which can be used as valid outcome measures in clinical trials and for remote evaluation of fine motor performance of PD patients. The findings from both papers answer the RQ 1. Generally, the methods, used for classification or numerical prediction, should not only have good level of validity, reliability and sensitivity but they should also be easily interpretable so that their results match with the clinical domain user expertise. These methods should be able to explain why a particular result was obtained and provide an intuitive explanation in the language of domain expertise. As mentioned above, DDM models lack this property. In order to tackle this problem, in this thesis, multivariate analysis methods were mainly considered and used since they provided results as linear combinations and were easy to visualize. In the field of machine learning, other approaches to interpretable modelling include decision trees and fuzzy inference systems.

A cross-analysis between objective measures of fine motor function and patient-based assessments of motor conditions was studied in Paper III. The findings indicated that the objective measures are useful in differentiating motor conditions of patients. More specifically, dyskinesias caused impaired spiral drawing performance and tapping accuracy whereas Off symptoms i.e. bradykinesia caused impaired alternating tapping speed, answering the RQ 2. PD patients may have difficulties in self-assessing their disability in relation to clinical assessments (Shulman et al., 2006), which may question the reliability of self-assessments. However, in the studies performed by Synnott et al. (2012) and Nyholm et al. (2012b), self-assessments correlated well with objective measures and clinical assessments, respectively. To the best of our knowledge, the study presented in Paper III is the first attempt to systematically analyse objective measures of motor function in relation to patient-based assessments.

The other part of the thesis focused on developing a method for providing comparable information content as the UPDRS scale by combining simultaneous subjective and objective measures into composite scores useful for representing the severity of different symptoms and a global health
condition and disability of the patient. In clinical settings, both physician- and patient-oriented outcomes offer complementary information (Christchilles et al., 1998) where the global disability of the patient is represented by a cumulative score of the four parts of the UPDRS scale, and is widely known as total UPDRS. Different parts of the scale contribute differently depending on their number of items, e.g. the self-rated ADL (part II) consists of 13 items whereas the clinician-rated motor performance (part III) consists of 34 items. The review of patents presented in Paper IV indicated that there were a small number of systems which collected and processed subjective and objective measures, answering the RQ 3. Therefore, a data-driven approach of defining an OTS was developed and presented in Paper V. The main idea was to initially summarize the time series into statistical features for representing symptom levels and fluctuations and then use these features in a DDM in order to form the OTS. By examining the MLR model of the OTS, it was observed that the objectively-measured dimensions i.e. ‘spiral’ and ‘tapping’ contributed more than subjectively-reported dimensions i.e. ‘walking’, ‘satisfied’, ‘off’ and ‘dyskinesia’, to the OTS with a 65% contribution. This indicated that the OTS contained similar information content as the total UPDRS score, answering the RQ 4. In general, using an OTS may facilitate the patient screening process and help in avoiding sub-optimization of treatments. In addition, given the fact that there is a high variability in symptom profiles both within- and between-patients, the OTS can be beneficial for deciding if a treatment change leads to an improvement of a patient’s general condition or not.

In addition to computer-based assessment methods, there is a need to introduce graphical user interfaces to end users, useful for providing visual feedback about patient’s overall functioning, symptom states and progress of the disease itself over time. The system presented in Paper V related to a non-knowledge-based CDSS consisting of a relational database management system with access to telemetry data, DDM functions for interpreting future data based on the DPSS, and the WA for enabling interactive reports and graphing functions (Shim et al., 2002). Generally, web-based systems have shown to provide several advantages when used for telemedicine purposes including mobility, accessibility, compatibility with other complementary technologies and easy installation. However, other issues such as i) architectural and technical, ii) security, privacy and confidentiality, and iii) usability and user acceptance need to be taken into account (Bellazi et al., 2001). The results from the usability evaluation of the WA showed that the information presented was comparable to the qualitative clinical observations of the nurses who had experience using the telemetry device during the Swedish study. In addition, from the demonstration of its functionali-
ties to the neurologists it was concluded that using the system assists in identifying patients who are not doing well and facilitates follow-up optimization of an individual patient’s treatment. Going further, the system was seen as most important for fluctuating patients and for regional patients i.e. patients living in regions far away from a clinic. Their general conclusion was that the system can be considered as a tool that will assist in the management of patients. These findings answer the RQ 5. Another web-based system called MercuryLive was developed by Chen et al. (2011) which enabled clinicians to access processed and analysed wearable sensor data and assess UPDRS motor scores in their patients through two-way video communication. In contrast to the proposed WA, this system was not a subject to a multicenter user evaluation and its results lacked representation of longitudinal trends of symptoms.

The feasibility in terms of validity, reliability and sensitivity to treatment interventions and disease progression of the computed scores i.e. six symptom dimensions and OTS, was evaluated in Paper VI. The problem of interest this paper presented was the mean changes of scores in advanced Swedish patients during the longitudinal 36-months study period. All the metrics were good indicating that that the scores can be considered as valid measures for long-term and remote assessment of PD symptoms. These findings answer the RQ 6. An important finding was that in contrast to subjectively-reported dimensions, objectively-measured dimensions i.e. ‘tapping’ and ‘spiral’ did not improve throughout the course of the study. These results are in line with those of a three-year follow-up study (Zibetti et al., 2013) of 25 advanced PD patients treated with LCIG where significant improvements were seen only in motor complications (UPDRS-part IV) and QoL (PDQ-39). The non-improvement of the two objectively-measured dimensions may be directly linked with the long-term therapy and over-treatment with LCIG which may have caused dyskinesias (abrupt involuntary movements) among patients. When analysing the longitudinal trends of the individual objective measures in Paper II and Paper VI, it was observed that they had different trends i.e. tapping speed improved whereas tapping accuracy and spiral drawing performance deteriorated over time. The deterioration of tapping accuracy and spiral drawing performance was linked with the occurrence of self-rated dyskinesias, as it was shown in Paper III. The relation between spiral drawing impairment and dyskinesias was also explored in Paper I by the strong correlation between WSTS scores and the scores of the method proposed by Liu et al. (2005) which was designed specifically to measure dyskinesias. Therefore, these results suggest that that the derived objective measures could reflect both
motor coordination deficits and fast movements being represented in the different fine motor tasks of the device.

Another part of this thesis investigated how to make use of raw time series data of fine motor tests so that clinicians can observe motor impairments of their patients while they perform the tests from their homes using the telemetry device. When dealing with raw data, visualization tools can be useful for presenting the data and for identifying trends and patterns at the patient’s individual level. These tools provide a visual overview of patient’s performance essential to understanding, analysing and interpreting the severity of symptoms. Paper I and Paper II presented two web-based frameworks for visualizing the performance of patients during spirography and alternating tapping tests, respectively and to allow clinicians to rate different motor impairments. A similar approach was recently presented in the study performed by Synnott et al. (2012) where a fine-grained visual reconstruction of the motor tasks performance of patients during a set of minigames, using a Nintendo Wii Remote device, was done with the aim of facilitating the understanding of the collected data. In their study, the cursor movement path and game objects were displayed to clinicians allowing them to observe motor impairments such as tremor and bradykinesia as well as to assess the reliability of the self-assessed severity of motor symptoms. From computer method development and evaluation point of view, the visualization approaches presented in the two papers were very important. In the two clinical studies i.e. Swedish and Italian, clinical ratings were available but only on a week-long test period level. By using visualizations it was possible to obtain test occasion clinical measures which in turn were used as dependent (or target) measures during the machine learning process where they were mapped to the quantitative measures. In addition, the visually assessed measures were used during correlation/agreement analysis for evaluating the generalization performance of the methods.

Although the in-clinic rating scales have proved to be useful in quantifying the severity of symptoms, their main drawback is related to the low resolution of assessments by providing only a momentary snapshot of the severity of symptoms. On the other hand, IT-based systems, coupled with statistical and machine learning methods, can be useful for deriving quantitative, objective (observer-independent), unobtrusive and valid measures which can improve the clinical management of PD by providing solutions to deliver healthcare by remote monitoring of patients. These kinds of systems have been previously shown to be more sensitive than the clinical measures (Goetz et al., 2009; Haubenberger et al., 2011) as well as to provide a means for higher-resolution assessments and for identifying subtle symptom changes (Nyholm et al., 2004). Yet, despite the efficacy of the
currently available assessment systems (see section 2.3), their effectiveness in routine clinical practice has yet to be established and should be widely adopted (Maetzler et al., 2013). In a recent review study performed by Baig and Gholamhosseini (2013), several issues were discussed for employing these kinds of systems in practical clinical use. They identified three main barriers including collection of high quality data, security and privacy, and usability and acceptability. Going further, they suggested that during each stage of system development domain experts i.e. clinicians should be consulted.

7.2 General implications

The visualization approaches presented in this thesis may be applied in other domains e.g. other medical applications, industrial applications, business intelligence, where there is a need for understanding, analysing and interpreting raw time series data. The visualizations may also be beneficial during supervised machine learning tasks where there is a lack of labelled training data.

The approach of combining multi-resolution analysis and dimension reduction techniques provided means for analysing non-stationary, transient and high-dimensional data. This approach could be useful for transforming real-world sensor data into a reduced and meaningful representative set of measures, a process useful for improving the regression/classification performance of the methods.

The data-driven approach of defining derived variables of different types of measures representing levels, trends and irregularities in the time series using dimension reduction techniques and incorporation of domain knowledge can also be generally applicable.

7.3 Limitations

In Paper I, it was observed that the method could not correctly classify the drawing impairment into exactly the correct category out of 10. This could be further adjusted by using an ordinal scale with fewer classes followed by employment of DDM models, designed for classification problems e.g. the LR. Another limitation of the method was that its outcome was based only on spatial information of the spiral i.e. by computing the spatial deviation from the pre-drawn spiral. Because of this limitation, the method could not detect and measure time-dependent effects such as drawing speed and hesitation. Therefore, it was not feasible to automatically separate different treatment-related motor dysfunctions being represented in the motor states i.e. Off/On/Dyskinesia.
A limitation of the Paper II was that the clinical evaluation of ATP was done by visual inspection of graphs and not by live/video observations of the patient’s tapping performance. However, even these kinds of observations can be biased as a result of the marked within- and between-clinician variability in ratings. For instance, in the study performed by Heldman et al. (2011), it was shown that clinicians differently weighed movement components of speed, amplitude and rhythm during video recorded tasks on items 23-25 (finger tapping, hand movements, rapid alternating of movements, respectively) of the UPDRS scale. In the paper, the rationale for visualizing graphs of ATP to clinicians was to derive a target measure to be used in correlation/agreement analysis against computed measures in order to develop and evaluate the method.

A limitation of the exploratory data analysis study presented in Paper III was the small number of observations in the extreme categories of q7 i.e. -3 (very Off) and +3 (very dyskinetic) in relation to other categories, in the two clinical studies.

One limitation of the usability evaluation of the WA presented in Paper V was the absence of feedback from PD physicians. Additional functionalities that were not yet implemented included functions for producing alerts if symptoms get worse compared to previous test periods, enabling a distributed access control to connect the treating clinical staff with their patients’ hospital records and to decide whether the requested information can be viewed by a particular clinical staff. At the time of publication of the Paper V, the system was in a prototype stage. To date, the whole system, consisting of the telemetry device, the data processing DPSS software module and the WA, is known as PAULA. PAULA is a CE-marked class 1 medical device fulfilling ISO13485 requirements and the EU IVD directive (AP, 2013).

A limitation of the longitudinal data analysis study presented in Paper VI was the data in the Swedish study were unbalanced i.e. sample sizes for the test periods were not equal. However, the employment of LME models enabled using all the data available, account dependencies within- and between-patients, and model mean scores, with patient ID as random effects (Cnaan et al., 1997). Even in clinical studies, it was shown that keeping patients self-reporting in long-term studies is difficult (Fernandez et al., 2013). Possible reasons for the high drop-out rate in the study included: the lack of feedback to patients about their test results, absence of immediate and direct benefit to patients’ own health, usage of the test battery was voluntary, fatigue development during the long-term study (Alves et al., 2004), the lack of feedback to nurses about their patients’ performance.
leading to deterioration of their interest towards the test battery, and technical issues.

7.4 Future prospects

Fine motor tests visualizations presented in Paper I and Paper II may help clinicians in the study of fine motor impairments using reconstructed time-space measurements. The web-based framework (Memedi et al., 2013) presented in Paper II not only visualized tapping performance but also the performance during spirometry. The spiral performance was depicted by animating the three spiral drawings per test occasion, allowing clinicians to observe real-time accelerations or hesitations and sharp spatial changes during the actual drawing process. Clinicians were instructed to observe the animations and to initially rate the drawing speed, hesitation, irregularity and finally identify the patient’s motor state i.e. Off/On/Dyskinesia. It would therefore be interesting to further investigate the feasibility of these visualizations. In order to accomplish this, there is a need for simultaneous collection of video recordings of patients’ performances during fine motor tests along with telemetry measurements of the test battery. This would enable comparison of clinicians’ ratings on videos and visualizations as well as the within-clinician variability in ratings, evaluating whether this assessment approach is feasible or not. In addition, this work will also provide a means for investigating whether animation of spirals provides more information to clinicians as compared to showing static images of spirals to them. The feedback part of the WA presented in Paper V could possibly be improved by adding visualizations of the dynamics of motor tests which would help clinicians during evaluation of the results.

As it was stressed earlier, one limitation of the method presented in Paper I was the inability to detect and measure the severity of time-dependent effects such as drawing speed and hesitation. Investigating the development and evaluation of a computerized approach to automatically separate therapy-related motor dysfunctions being represented in the fine motor tests will be considered in the near future. This will assist clinicians in optimizing the therapy, enabling patients to spend more time in the On (normal) state. Given the fact that there is a large degree of variability within- and between-patients in manifestation of motor complications, infrequent assessments are not enough for capturing the severity and variation of complications. Therefore, continuous and prolonged measurements e.g. using wrist-worn devices such as the Parkinson’s KinetiGraph (Griffiths et al., 2012), may provide a more reliable estimate of the amount of variability of these complications as compared to just using measurements that were taken four times per day using the telemetry device.
Another aspect that will be investigated in the future is the impact of routinely using the system presented in Paper V in clinical practice. The focus will be on analysing the actual benefits and problems when using the system in practice by healthcare professionals.

7.5 Concluding remarks
The clinical management of progressive neurological disorders may be enhanced by application of IT-based systems composed of high-quality data collection schemes, well-designed computer-based methods and user-friendly graphical user interfaces. This thesis analysed the problem of objective quantification and remote monitoring of the severity of PD symptoms by using repeated measures collected by means of a telemetry device.

Methods for automating the process of scoring the severity of symptoms provide a means for deriving precise, accurate and objective measures which can be used as outcome measures in clinical trials, enabling remote and frequent assessments. However, aspects like validity, reliability, sensitivity to treatment interventions, data security issues, user satisfaction and acceptability, and interpretability must be considered before IT-based systems can be applied in routine clinical practice.

The application of data-driven methods in combination with different time series analysis methods not only allows deriving models for interpreting future symptom data but also facilitates the analysis of large data sets for discovering patterns and relationships. This helps in extending our knowledge about mechanisms linking the patients’ health status and their actual performance in motor tests.
References:


wireless technologies. In L. Bos, S. Laxminarayan, and A. Marsh, editors. Medical and Care Compunetics 1. IOSPress, Amsterdam.


Publications in the series
Örebro Studies in Technology


