Smartphone-based Parkinson’s disease symptom assessment

Somayeh Aghanavesi
Smartphone-based Parkinson’s disease symptom assessment

Somayeh Aghanavesi

Microdata Analysis
School of Technology and Business Studies
Dalarna University
Sweden
2017

Abstract

This thesis consists of four research papers presenting a microdata analysis approach to assess and evaluate the Parkinson’s disease (PD) motor symptoms using smartphone-based systems. PD is a progressive neurological disorder that is characterized by motor symptoms. It is a complex disease that requires continuous monitoring and multidimensional symptom analysis. Both patients’ perception regarding common symptom and their motor function need to be related to the repeated and time-stamped assessment; with this, the full extent of patient’s condition could be revealed. The smartphone enables and facilitates the remote, long-term and repeated assessment of PD symptoms. Two types of collected data from smartphone were used, one during a three year, and another during one-day clinical study. The data were collected from series of tests consisting of tapping and spiral motor tests. During the second time scale data collection, along smartphone-based measurements patients were video recorded while performing standardized motor tasks according to Unified Parkinson’s disease rating scales (UPDRS).

At first, the objective of this thesis was to elaborate the state of the art, sensor systems, and measures that were used to detect, assess and quantify the four cardinal and dyskinetic motor symptoms. This was done through a review study. The review showed that smartphones as the new generation of sensing devices are preferred since they are considered as part of patients’ daily accessories, they are available and they include high-resolution activity data. Smartphones can capture important measures such as forces, acceleration and radial displacements that are useful for assessing PD motor symptoms.

Through the obtained insights from the review study, the second objective of this thesis was to investigate whether a combination of tapping and spiral drawing tests could be useful to quantify dexterity in PD. More specifically, the aim was to develop data-driven methods to quantify and characterize dexterity in PD. The results from this study showed that tapping and spiral drawing tests that were collected by smartphone can detect movements reasonably well related to under- and over-medication.

The thesis continued by developing an Approximate Entropy (ApEn)-based method, which aimed to measure the amount of temporal irregularity during spiral drawing tests. One of the disabilities associated with PD is the impaired ability to accurately time movements. The increase in timing variability among patients when compared to healthy subjects, suggests that the Basal Ganglia (BG) has a role in interval timing. ApEn method was used to measure temporal irregularity score (TIS) which could significantly differentiate the healthy subjects and patients at different stages of the disease. This method was compared to two other methods which were used to measure the overall drawing impairment and shakiness. TIS had better reliability and responsiveness compared to the other methods. However, in contrast to other methods, the mean scores of the ApEn-based method improved significantly during a 3-year clinical study, indicating a possible impact of pathological BG oscillations in temporal control during spiral drawing tasks. In addition, due to the data collection scheme, the study was limited to have no gold standard for validating the TIS. However, the study continued to further investigate the findings using another screen resolution, new dataset, new patient groups, and for shorter term measurements. The new dataset included the clinical assessments of patients while they performed tests according to UPDRS. The results of this study confirmed the findings in the previous study. Further investigation when assessing the correlation of TIS to clinical ratings showed the amount of temporal irregularity present in the spiral drawing cannot be detected during clinical assessment since TIS is an upper limb high frequency-based measure.

Keywords: Parkinson’s disease; symptom assessment; spiral; tapping; smartphone; temporal irregularity; timing variability; approximate entropy;
Included papers

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


My contributions to the papers were as follows:

Paper I – I planned the literature review in the paper together with Jerker Westin. I conducted the review and wrote the paper and revised it.

Paper II – I have been in a session for data collection and I performed the data processing, development of the methods, and data analysis. I wrote the first version of the manuscript and revised it.

Paper III – I was involved in method development, analysis of the results, and writing and revising parts of the manuscript.

Paper IV – I performed data processing, method development, as well as the data analysis, and I wrote the manuscript and revised it.
Table of Contents

Introduction .................................................................................................................................................. 1
I. Review of methods and measures (Paper I) ........................................................................................... 3
II. Smartphone-based system to quantify Parkinson’s disease dexterity (Paper II) ................................. 4
III. Measuring Parkinson’s disease related temporal irregularity (Paper III) ............................................. 6
IV. Verification of Parkinson’s disease temporal irregularity measure (Paper IV) ....................................... 7
Conclusion and future research ................................................................................................................. 8
Acknowledgements ................................................................................................................................... 9
References ................................................................................................................................................. 9
PAPER I .................................................................................................................................................. 12
PAPER II .................................................................................................................................................. 19
PAPER III .................................................................................................................................................. 37
PAPER IV .................................................................................................................................................. 45
Introduction

This is a licentiate thesis in microdata analysis. In the study of Parkinson’s disease (PD) motor symptom assessment, microdata is information at the level of subjects’ time-based motor movements, and time-based treatment. The time levels of collected data in this thesis refer to the movement and treatment data that were collected and analyzed during two schemes. In one data collection scheme, the data was collected during 36-month whereas during another data collection scheme it was collected during a day at various time points. The variables were recorded separately for each subject who performed the tests. The subjects’ data with their time stamp ranging from months to seconds were analyzed. Large quantities of data were involved and the analyses were separately performed for subjects based on variables. An advantage of performing research in microdata is that the results are aggregated from the data that were analyzed at micro-level. However, summarizing results to an aggregate level may result in information loss. Exploring the relationships between variables from aggregated results are not possible. Therefore having access to microdata allows to investigate the interactions between variables and perform detailed analysis. The aim of microdata analysis in this thesis was to fill the identified gaps between data, information, and knowledge.

PD is the second most common neurodegenerative disorder (Shin et al., 2016) and is characterized by degeneration of dopaminergic neurons in the substantia nigra. It is characterized by motor and non-motor symptoms that can impact on the function to a variable degree. The four cardinal motor symptoms of PD comprise tremor, rigidity, bradykinesia and postural instability. Common treatments for PD are levodopa and dopamine agonist tablets, deep brain stimulation (DBS) and levodopa/carbidopa intestinal gel (LCIG) (Schapira, 2009). The primary goal of therapy is to maintain good motor function.

Motor symptoms of PD are typically assessed using clinical scales such as Unified PD rating scale (UPDRS) which is used to evaluate the presence and severity of PD motor symptoms as well as symptom fluctuations. However, clinical scales-based measurements are not able to capture variations in symptoms continuously and they are insensitive to subtle changes. To reveal the full extent of patients’ condition and prevent recall and reporting bias, the symptoms need to be captured frequently, before and after medication (Isacson, et al., 2008a). In addition, the therapeutic decision making requires accurate, comprehensive and accessible quantification of symptoms. Electronic sensor-based systems can facilitate remote, long-term and repeated symptom assessments. They are able to capture the symptom fluctuations more accurately and also they are effective with patient’s hospitalization costs.

Recent advances in information and communication technologies have enabled remote and continuous monitoring of motor symptoms (Maetzler, et al., 2013). Previous studies have shown that such technologies provide accurate and valid objective assessment of symptoms. It was previously reported that they may assist in identifying motor functions (On, Off and dyskinesia) (Andong Zhan, 2016). From the technological point of view, data from different kinds of sensors during standardized tests and passive monitoring of physical activity have been previously analyzed and processed using signal processing and machine learning methods (Griffiths, et al., 2012; Zeng, et al., 2016). Some studies have focused on analyzing data from upper limbs during standardized tasks like finger tapping (Shima, et al., 2009), digital spiral analysis. As an alternative to wearable sensors-based systems, touch screen devices (Haubenberger, et al., 2011; X. Liu, et al., 2005; Saunders-Pullman, et al., 2008) have been the focus for some research studies for assessing dexterity performance of PD patients by analyzing their collected upper limb motor data. Such smartphone measurements were previously used for assessing different fine motor dysfunctions such as tremor (Haubenberger, et al., 2011), dyskinesia (X. Liu, et al., 2005), drawing impairments (Westin, et al., 2010a).
Quantitative measures during alternating tapping tests and digital spiral analysis have been previously used as measures of bradykinesia (Giovannoni, et al., 1999) and severity of PD motor symptoms (Saunders-Pullman, et al., 2008) respectively.

With a view to technologies used to detect PD dexterities, measuring rhythm and regularity in PD are found important. In research to measure the regularity, variability, complexity, and unpredictability of various movement signals in PD, Approximate Entropy (ApEn) method is identified as an essential method. Regularity was basically measured by regularity statistics which were centered on various entropy measures (Pincus, et al., 1991). In statistics, ApEn is a technique to quantify the regularity, unpredictability, uncertainty of fluctuations over time series data. For this purpose signals from ankle joint while PD patients walked on a treadmill (Kurz, et al., 2010), as well as their tremor signals (Morrison, et al., 2013), and wrist movements signals (Powell, et al., 2014) have been analyzed with this method to measure the amount of regularity. In addition, time series like biking speed signal (Mohammadi-Abdar, et al., 2016) and EEG emotional signal (Yuvaraj, et al., 2016) have been analyzed with this method to measure the amount variability in relation to PD.

As introduced in the previous paragraphs, there is a need to review the state of the art, sensors, and signals that were used to evaluate the PD motor symptoms in order to identify the potential points for further development and evaluation of the symptom assessment methods.

In view of the technologies/methods that were used to evaluate PD dexterity, there was found no study reporting an approach where tapping and spiral drawing test data were combined in a data-driven manner and related to objective measures of motor functions, various clinical ratings, and actual treatment. In addition to what is done to measure the regularities in PD movement signals, the question for temporal irregularity measure of spiral drawings that were performed by PD patients remained uninvestigated except in our last study (Memedi, et al., 2016b).

The main objective of this licentiate thesis is to develop PD symptom assessment methods. With regard to what were mentioned as missing points in the last paragraph, the following research questions arise:

- What is the current state of the art, sensor systems, and measures to detect, assess and quantify the four cardinal and dyskinetic motor symptoms?
- How to combine spiral and tapping tests data in a data-driven manner and relate them to the objective assessment of motor function in PD, various clinical ratings, and the actual treatment?
- How to measure the temporal irregularity in Parkinson’s disease using spiral drawing tests?
  - How to verify the findings of last research question using another screen resolution, new dataset, new patient groups, and during shorter term measurements?

In contemplation of answering last research questions, the work in this thesis has been segmented into four steps. In the first step, a review study of the state of the art and sensor systems that have been used to detect, assess and quantify the four cardinal and dyskinetic motor symptoms, has been conducted (Paper I). The review surveys the manifestation of PD motor symptoms, sensors that were used for their detection, types of signals (measure) as well as their signal processing methods. A summary of this review’s finding is represented in a table including devices, measures, and methods that were used in each reviewed motor symptom assessment study.

The second step aims to investigate whether a smartphone-based system can be used to quantify dexterity in PD (Paper II). More specifically, the aim was to develop data-driven methods to quantify and characterize dexterity in PD. The raw tapping and spiral data were processed and analyzed with time series
analysis techniques to extract 37 spatiotemporal features. For each of the five scales, separate machine learning models were built and tested by using principal components of the features as predictors and mean ratings of the three specialists as target variables.

During the third step, the development and evaluation of clinimetric properties of a method were done for measuring PD-related temporal irregularities using digital spiral analysis (Paper III).

Finally, the fourth step aimed to confirm the reproducibility of the results obtained in previous research using a new dataset, new patient groups, using another screen resolution, and during shorter term measurements. In addition, further investigations for assessing the responsiveness of irregularity score to treatment, and it’s correlation to clinical ratings were done (Paper IV).

The above brief description attempts to present the appended papers that are attached at the end of this thesis. The rest of the thesis is organized with summaries of papers in order. The thesis ends with conclusions and proposes possible future studies.

I. Review of methods and measures (Paper I)

PD is a progressive neurological disorder characterized by a large number of motor symptoms that can impact on the patient’s function to a variable degree. Since electronic sensor-based systems can facilitate remote, long-term and repeated symptom assessments, these technologies attempting to assess the PD motor symptoms objectively are increasing. This paper investigates the state of the art, measures and sensor systems that were used to detect, assess and quantify the four cardinal and dyskinetic motor symptoms. The four cardinal motor symptoms of PD comprise of tremor, rigidity, bradykinesia and postural instability. Parkinson’s tremor is the most apparent well-known symptom. Rigidity symptoms cause stiffness and result in inflexibility. Bradykinesia is generally the slowness of movements. Postural instability symptom is a trend to be unstable and dyskinesia is a difficulty in performing voluntary movements, which often occurs as a side effect of long-term therapy with levodopa.

Among the developed electronic techniques to measure and analyze the PD’s symptoms the common sensors and devices for evaluation are accelerometer, electromyograph, magnetic tracker system, gyroscope, digitizing tablet (X. Liu, et al., 2005; Westin, Ghiamati, et al., 2010), video recording, motion detector, and depth sensor. Methods like Wavelet transform, Principal component analysis, Discrete Fourier Transform, Fast Fourier transform (Westin, Dougherty, et al., 2010), Detrended Fluctuation Analysis (Arora, et al., 2015b), and Spectral Analysis are identified as signal processing methods for PD’s motor symptom assessment. Furthermore, Artificial Intelligence (AI) that is used in frameworks of Visual perception, decision making, image processing, and classification techniques enables the development of computer systems to perform tasks that usually require human intelligence. Machine learning in PD symptom assessment (Sama, et al., 2012) often includes techniques to assess the magnitude of addressed symptom. In addition, linear discriminant analysis (Lopane, et al., 2015) with various alternatives of Non-parametric, generalized and multilayer perceptron analysis was invoked as a classification method in AI.

The results of this review paper are summarized in a table which includes the evaluated symptoms, type of the instruments, calculated measures and employed analytical assessment methods. This table may help in recognizing the common or special types of methods and sensors to be used for different types of symptoms assessments. Based on the obtained results the two symptoms, rigidity and postural instability are mostly evaluated as single symptoms. However, among articles which assessed the combination of symptoms,
bradykinesia (together with tremor, dyskinesia, rigidity, and dyskinesia together) is the most studied symptom.

A common sensor for symptom detection was the accelerometer that was mostly used for detecting the tremor, dyskinesia, and postural instability. The digitizing tablet is used for assessing almost all types of symptoms. Smartphone (Arora, et al., 2015b) and Microsoft Kinect (Yeung, et al., 2014) are the latest devices in the market used for this purpose. Smartphones as a new generation of sensing devices are expanding rapidly with PD motor symptom assessments. Since wearable sensors are small, available, accurate, including high-resolution activity data, and flexible with body locations, they are preferred for PD since it's a progressive chronic disease and symptoms need to be assessed continuously throughout the day. For this, mobile applications and wrist watches are more preferred as they are currently part of almost everyone's daily accessories. However, their analysis methods and their validations are important and a question is whether the devices or clinical ratings will become the gold standard. Machine learning techniques are potentially good solutions in the development of assessment systems to determine the effectiveness of drug dosing. Tools that can effectively characterize the severity of symptoms and can discriminate between bradykinesia and dyskinesia are needed. Some successful products are Parkinson's Kinetigraph, Kinesia devices, and PD Holter monitor from Rempark project.

II. Smartphone-based system to quantify the PD dexterity (Paper II)

The aim of this paper is to investigate whether a smartphone-based system can be used to quantify dexterity in PD. More specifically, the aim was to develop data-driven methods to quantify and characterize dexterity in PD.

The current state of the art for assessing PD motor symptoms in clinical routine and studies is by using clinical rating scales based on observations and judgments of clinicians and medical history. However, clinician-based measurements are not able to capture variations in symptoms on a day-to-day basis since they only reflect one brief point in time. To reveal the full extent of patients’ condition and prevent a recall and reporting bias, motor symptoms need to be captured frequently, before and after medication (Isacson, et al., 2008a). From the clinical point of view, it is challenging to remotely and frequently determine the current motor state of the patient to determine whether the patient is under-medicating (a state in which the PD motor symptoms such as bradykinesia, tremor, rigidity, and others appear) or over-medicating (the appearance of hyperkinetic movements related to excessive levels of medication). Therefore, assessing the current motor state of the patients is essential for deriving an optimal treatment strategy.

From the technological point of view, there are different studies with the focus on quantifying various motor symptoms. Some have focused on assessing motor dysfunctions in upper extremities (Heldman, et al., 2014), and some focused on gross motor symptoms like gait (Mariani, et al., 2010), while others focused on combination of both. As an alternative to wearable sensors-based systems, some research groups have focused on assessing dexterity performance of PD patients by analyzing upper limb motor data collected by means of touch screen devices. Smartphone measurements were previously used for assessing different fine motor dysfunctions like tremor, dyskinesia, drawing impairments and global tapping performance. Quantitative measures during alternating tapping tests and digital spiral analysis have been previously used as measures of bradykinesia and severity of PD motor symptoms. To our knowledge, there is no study
reporting an approach where tapping and spiral drawing test data were combined in a data-driven manner and related to clinical ratings and actual treatment.

During the data collection, nineteen fluctuating PD patients and 22 healthy controls were recruited in a single center, open label, single dose clinical trial. Both patients and healthy controls were asked to perform dexterity tests (tapping and spiral drawing) using a smartphone at specific time intervals. For the patients, the dose administered was 150% of their individual levodopa equivalent morning dose to follow transitions between Off, On, and On with dyskinesia motor states. The healthy controls were asked to perform the tests, 8 times each. Along with smartphone-based measurements, patients were video recorded while performing standardized motor tasks according to Unified Parkinson’s disease rating scales (UPDRS) (Fahn, et al., 1987). The recorded videos were presented in a randomized order to three movement disorder specialists so that the ratings were blinded with respect to time from dose administration. The specialists rated three UPDRS-part III (motor examination) items including UPDRS item #23 (finger tapping), UPDRS #25 (rapid alternating movements of hands), and UPDRS #31 (bradykinesia), according to the definitions of the motor examination part of the UPDRS.

The raw dexterity data were processed with time series analysis methods to calculate 37 spatiotemporal features, which represent the severity of symptoms. Different kinematic quantities, including time, distance, speed, and velocity were used as primary signals to be processed and analyzed using time- and wavelet-domain methods. To reduce the dimensions of the features but keep the most important and related information into a smaller set, principal component analysis (PCA) using the correlation matrix method was applied on the 37 features. Seven principal component scores (PCs) were retained and used in subsequent analysis. The PCs were used as predictors to supervised machine learning methods used to map to the mean ratings of the three movement disorder specialists on the clinical rating scales used in the clinical trial. Four machine learning methods were evaluated and a stratified 10-fold cross-validation was applied to test the performance of the machine learning methods. For each of the five scales, separate models were built and tested. The performance of the machine learning methods, the agreement between the three specialists’ ratings were assessed. The mean PCs differences between the groups of patients and healthy controls were examined. The relative ability to detect the change from baseline (no medication) to follow up time points when patients were on medication was assessed by effect sizes. Test-retest reliability (or consistency) of mean specialist and smartphone-based scores was calculated.

The obtained results indicate that the methods could capture motor symptoms reasonably well as compared to the mean assessments of three movement disorder specialists on three items of UPDRS-III, TRS and dyskinesia scales. The correlations were weak to moderate between the scores generated by the methods and the mean clinical ratings, indicating that tapping and spiral drawing tests capture relevant symptom information corresponding to the clinical rating scales. In addition, the PC1 could follow transitions between motor states across the levodopa test cycle since it had similar trends as the TRS and dyskinesia scale. These results suggest that tapping and spiral drawing tests with the smartphone can detect movements reasonably well related to under- and over-medication. In contrast to the clinical rating scales, another advantage with the current system is that PD-related outcomes can be captured and assessed more frequently. As a limitation of this study, there was a considerable amount of inter-rater variability.

As a conclusion, the results presented in this paper indicate that tapping and spiral drawing tests of the smartphone contain relevant symptom information for detecting and assessing PD dexterity. The results suggest that the tests can be useful in detecting changes in motor symptoms related to treatment.
III. Measuring Parkinson’s disease related temporal irregularity in spiral drawings (Paper III)

The motivation of this paper is the research question about measuring the temporal irregularity in PD using spiral drawing tests. One of the disabilities associated with PD is the impaired ability to accurately time movements (Coull, et al., 2011). Many studies have shown an increase in timing variability among patients when compared to healthy subjects, suggesting the Basal Ganglia (BG) has a role in interval timing (Buhusi, et al., 2005). Relating objective measures obtained by instrumented tests to pathological BG fluctuations would be beneficial for facilitating the assessment of high-frequency motor irregularities that could be difficult to be assessed visually (Raz, et al., 2001).

In this study, the upper limb temporal irregularity of 98 patients in different stages of PD and 10 healthy elderly subjects was investigated. The study was carried out when the subjects performed repeated spiral drawing tasks in their home environments using a touch screen device. On inclusion, 35 of these patients were treated with levodopa-carbidopa intestinal gel infusion (LCIG) and 30 patients were candidates for switching from conventional oral PD treatment to LCIG. Approximate entropy method was used to quantify the amount of temporal irregularity during spiral drawing tasks. In addition to this method, two more spiral scoring methods of a standard deviation of drawing velocity (SD-DV) (X. Liu, et al., 2005), and wavelet spiral test score (WSTS) (Westin, et al., 2010b) were adopted and their scores were compared to the ApEn-based method.

Differences in mean scores of the three methods across the four subject groups (healthy, early PD, intermediate PD, advanced PD) were assessed. Three machine learning methods of logistic regression, C4.5 decision tree, and random Forests with stratified 10-fold cross validation were tested to classify the two groups of healthy and patients. The relative ability of the methods to detect the change from baseline (oral treatment) to month 0 (LCIG) was determined. The long-term progressions of the three scores were assessed. Correlations between the three scores as well as the Test-retest reliability of the three scores were evaluated.

As result, the mean temporal irregularity score (TIS) differed significantly between healthy elderly subjects and advanced PD patients (P<0.005). The best performing classifier was the logistic regression that correctly classified the group with an accuracy of 85% and weighted AUC of 0.89. The ApEn-based method had a better responsiveness and test-retest reliability when compared to the other two methods (SD-DV, WSTS). In contrast to the other methods, the mean scores of the ApEn-based method improved significantly during a 3-year clinical study, indicating a possible impact of pathological BG oscillations in temporal control during spiral drawing tasks.

A limitation of the study was that there was no gold standard to validate the TIS. This limitation was related to the data collection scheme where subjects repeatedly used the touch screen device in their home environments without clinical supervision. That is, there was no test occasion level clinical ratings to assess the validity of TIS. This motivates to continue the study with the further investigation of the relation between TIS and clinical ratings as well as investigating the clinimetric properties of TIS on a new data set. Alternatively, since the BG oscillations are affected by deep brain stimulation (DBS) treatment frequencies another interesting research question that could be investigated in the future would be to calculate the TIS on spiral data of patients on DBS treatment. This would allow investigating the relationship between TIS and different DBS frequencies as well as during Off and On stimulation phases and whether TIS could be useful for adjusting DBS treatment.
IV. Verification of Parkinson’s disease temporal irregularity measure during spiral drawings (Paper IV)

There are different studies with the focus on quantifying PD regularity, variability, complexity, and unpredictability of various signals using Approximate Entropy (ApEn). Some studies measured regularity of the ankle joint movement signal (Kurz, et al., 2010), tremor signal (Morrison, et al., 2013), and wrist movement signal (Powell, et al., 2014) in PD. In measuring the variability of signals using ApEn, studies were found that measured the amount of variability in biking speed signal (Mohammadi-Abdar, et al., 2016), and EEG emotional signal (Yuvaraj, et al., 2016). To our knowledge, ApEn has not been investigated for PD where patients performed spiral tests, except in our last study (ST1) (Memedi, et al., 2016b).

As the continuation from ST1, The purpose of this paper is to verify and further investigate the results from ST1 using the calculated upper limb temporal irregularity of PD patients with same methodology (ApEn) but using a new dataset, new patient groups, using another screen resolution, and during shorter term measurements. The paper reports properties of the temporal irregularity measure including the differences in TIS between patients in different stages of PD and healthy subjects and the test-retest reliability of TIS. In addition, the responsiveness of TIS to levodopa treatment during single dose experiments, and the correlations between TIS and clinical ratings are reported.

In this study nineteen PD patients and 22 healthy controls performed repeated spiral drawing tasks on a smartphone. Patients performed the tests before a single levodopa dose and at specific time intervals after the dose was given. Healthy controls were asked to perform a fine motor test (spiral drawing) using a smartphone at specific time intervals. Up to 15 samples per PD patient, and up to 8 samples per healthy control were collected. Three movement disorder specialists rated videos of the patients while performing standardized motor tasks. The specialists rated six UPDRS-part III (motor examination) items including UPDRS item #23 (finger tapping), UPDRS #25 (rapid alternating movements of hands), UPDRS #26 (leg agility), UPDRS #27 (Arising from Chair) and, UPDRS #29 (gait), UPDRS #31 (bradykinesia), according to the definitions of the motor examination part of the UPDRS. They also rated dyskinesia and overall mobility according to treatment response scale (TRS).

A temporal irregularity score (TIS) was developed using an approximate entropy method which was used to measure the repeatability of patterns within the drawing speed signals of the spiral drawings. To be able to do a comparable analysis with groups of patients in ST1, three groups of patients were defined based on the Hoehn and Yahr scale (Goetz, et al., 2004).

The significance of the mean TIS across the four subject groups of healthy, early, intermediate and advanced patients was investigated. To evaluate the consistency of the TIS, test-retest reliability of the TIS was measured. The ability to detect the change from baseline (before medication) to follow up time points when patients were on medication, was assessed. To investigate the relation between TIS and specialists’ visual assessments the correlations between TIS and clinical rating scores were assessed.

The results in the paper indicate that the mean TIS was significantly different between healthy subjects, and all other groups of early, intermediate and advanced patients. There was a good test-retest reliability (ICC = 0.81) indicating the consistency of the three extracted TIS. It can be observed in the results that TIS was worsening with increased disease severity. In addition, in this study, the sensitivity of TIS to the treatment response during single dose experiments from before to 300 minutes after dose intake, were calculated. The sensitivity results could not be comparable to the sensitivity results in ST1 due to the differences in the study designs. However, the effect sizes were related to changes in patients’ symptoms throughout the continuous tests which were performed during the day. The scores were small but could
capture some effects of medication that were presented in patients’ performances. For healthy subjects, the effect sizes were smaller than the effect sizes of patients. As an additional investigation, the correlations between TIS and mean clinical rating scores (UPDRS, TRS, Dys) were weak. A reason could be that TIS was related to temporal fluctuations during spiral drawing. This could be related to the fact that ApEn technique provided a high resolution of a signal by partitioning it into a smaller set of windows and sliding the windows throughout the signal. This enables the TIS to measure high-frequency irregularities in spiral drawing speed in the order of milliseconds where visual assessment of these irregularities cannot be detected.

As a conclusion, the results of this study as well as in ST1 showed that TIS was more related to disease severity than the clinical ratings. It indicates it might be more useful in long-term diagnostic tools than for detecting the treatment response. It also introduces a possibility that TIS might be related to Basal ganglia oscillations. However, to specifically be able to demonstrate the possible correlations between TIS and pathological oscillations there is a need for further investigations. DaT SPECT (Scherfler, et al., 2007) as a tool to determine the degeneration of the presynaptic neurons in the striatum of patients with PD could be used to investigate whether there is a relation between the number of dopamine receptors and TIS in future studies. In addition, the investigation of TIS properties on spiral data of patients with DBS therapy would also be of interest since the BG oscillations are affected by DBS treatment frequencies (Da Cunha, et al., 2015).

Conclusions and future work

In this licentiate thesis, several steps of microdata analysis (from data collection to reporting the results) were taken. Each step individually was considered as important in order to achieve correct results. However, the thesis focuses on data analysis. Microdata analysis as a multidisciplinary area comprise fields such as decision support systems, AI, data modeling, simulation, statistical inference, and optimization techniques that were considered as important in achieving the research goals.

Through reviewing the state of the art and measures in the first paper the potentials about motor symptoms and their assessments were observed. Best sensor systems, more suitable methods and possible measures in order to assess the symptoms were explored. Various combinations of sensor systems, as well as methodologies, could be examined to expand the research in the field of PD motor symptom assessment. Machine learning techniques were found as suitable solutions in the development of assessment systems. Therefore it motivated the data-driven machine learning approach in the next study. In the view of recent technologies, the tapping and spiral tests were combined to examine whether their combination could be used to quantify dexterity in PD. Results from combined tapping and spiral drawing tests of smartphone suggested that this combination can be useful in detecting changes in motor symptoms related to treatments. Since the development of new treatment strategies mainly depends on the clinical information derived from rating scales, there was a limitation with regard to inter-rater variability in this thesis. Even though the specialists used scoring guidelines, the analysis from rating agreements showed that there was a weak agreement between raters. Meanwhile, the temporal irregularity score calculated for the drawing speed during spiral tests found to be useful in long term diagnosis. Even though it would be interesting to investigate the TIS properties after including more patients’ data from different stages of medication (and even DBS therapy) and data from patients from various years of treatment.
The research is ongoing to further discover the properties of methods. The aim is to expand PD symptom assessment with including more sensor data i.e. gait data, hand movements data and etc. As has been mentioned, the overall goal of the project is to come up with a sensor index platform where all the sensor indexes are integrated and linked to PD motor symptoms and treatment.

Acknowledgments

The studies are performed in the framework of the FLOAT and MUSYQ projects, funded by Swedish knowledge foundation and Swedish innovation agency (Vinnova), in collaboration with Cenvigo AB, Sensidose AB, Swedish ICT Acreo, Uppsala University, Sahlgrenska University, and Dalarnan University. I would like to thank all my supervisors and colleagues who helped and supported me in learning and perusing this research.

References


Paper I
A review of Parkinson’s disease cardinal and dyskinetic motor symptoms assessment methods using sensor systems

Somayeh Aghanavesi, Jerker Westin
Dalarna University, Computer Science Department, Rodavagen 3, S78188 Borlänge, SWEDEN saa,jwe@du.se

Abstract. This paper is reviewing objective assessments of Parkinson’s disease (PD) motor symptoms, cardinal, and dyskinesia, using sensor systems. It surveys the manifestation of PD symptoms, sensors that were used for their detection, types of signals (measures) as well as their signal processing (data analysis) methods. A summary of this review’s finding is represented in a table including devices (sensors), measures and methods that were used in each reviewed motor symptom assessment study. In the gathered studies among sensors, accelerometers and touch screen devices are the most widely used to detect PD symptoms and among symptoms, bradykinesia and tremor were found to be mostly evaluated. In general, machine learning methods are potentially promising for this. PD is a complex disease that requires continuous monitoring and multidimensional symptom analysis. Combining existing technologies to develop new sensor platforms may assist in assessing the overall symptom profile more accurately to develop useful tools towards supporting better treatment process.

Keywords: Parkinson's disease; sensors; objective assessment; motor symptoms; machine learning; dyskinesia; bradykinesia; Rigidity; tremor.

1 Introduction

The number of studies using electronic healthcare technologies and sensor systems assessing the Parkinson’s disease (PD) motor symptoms objectively are increasing. PD is a progressive neurological disorder characterized by a large number of motor symptoms that can impact on the function to a variable degree. The four cardinal motor symptoms of PD comprise of tremor, rigidity, bradykinesia and postural instability. The primary goal of therapy is to maintain good motor function. Therefore therapeutic decision making requires accurate, comprehensive and accessible quantification of symptoms. Electronic sensor-based systems can facilitate remote, long-term and repeated symptom assessments. They are able to capture the symptom fluctuations more accurately and also they are effective with patient’s hospitalization costs. This paper reviews methods and sensor systems to detect, assess and quantify the four cardinal and dyskinetic motor symptoms. The method for identifying and accessing resources involved the online databases, Google Scholar, IEEE computer society, Springer link (Springer Netherlands) and PubMed central. The evaluation of resources was based on their relevance to the topic and the year of publication (not older than 2005). Selection of articles is done to have one reference per instrument that was used to detect all our addressed symptoms. The structure of this article is formed into sections of PD symptoms, followed by corresponding sensors and instruments, and computer and statistical methods that were employed for assessments.
2 Parkinson’s disease cardinal and dyskinetic symptoms

Parkinson’s tremor consists of oscillating movements and appears when a person’s muscles are relaxed and disappears when the person starts an action. It’s the most apparent well-known symptom. Rigidity symptoms cause stiffness of the limbs, neck or trunk and result in inflexibility. Bradykinesia (slow movement) describes the general reduction of spontaneous movement (abnormal stillness and a decrease in facial expressivity) and causes difficulties with repetitive movements. It can cause walking with short and shuffling steps and can also affect the speech. Postural instability symptom is a trend to be unstable when standing upright, rising from a chair or turning. And dyskinesia is a difficulty in performing voluntary movements, which often occurs as a side effect of long-term therapy with levodopa. Dyskinetic movements look like smooth tics (uncoordinated periodic moves).

3 Sensors, signals and measures

Among the developed electronic techniques to measure and analyze the PD’s symptoms the common sensors and devices for evaluation are accelerometer, electromyograph (EMG), magnetic tracker system, gyroscope, digitizing tablet, video recording, motion detector, and depth sensor. In accelerometry, an electromechanical sensor device is used to measure acceleration forces and capture the movements by converting it into electrical signals that are proportional to the muscular force producing motion. Gyroscope is a sensor device used to measure angular velocity (angular rate) which senses rotational motion and changes in orientation (Salarian, et al., 2007). Accelerometer and gyroscope are joint in many motion sensing instruments. Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by neurologically activated muscles using Electromyograph that records how fast nerves can send electrical signals. Digitizing tablet in PD symptom detection is a computer input device used to digitize patient’s drawing when he/she traces a pre-drawn shape (X. Liu, et al., 2005), (Westin, Ghiamati, et al., 2010) or freely writes or draws a shape. The position of the tip of the pen (x, y) and the time (milliseconds) are collected for analysis (Westin, Ghiamati, et al., 2010). Electromagnetic tracker system captures the object’s movement displacement (x, y, and z) and orientation (pitch, roll, and yaw). Active optical marker systems are used to capture and record object’s motion. Wired position markers can be placed on different locations of patient’s body to obtain object’s posture and movements.

4 Signal processing and analysis

Wavelet transform as a multi-resolution transformation method uses a variable window size at each level to obtain more information about the sensor signal in the time-frequency (time-scale) domain. Principal component analysis (PCA) is theoretically the best linear dimension reduction technique that uses rectangular transformation to convert the set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables [3]. It’s the direction to where the most variance exists. Wavelet transform is usually used with PCA to reduce the number of features to most important and related ones [3]. Discrete Fourier Transform converts samples of a function (a signal that varies over time) into the list of coefficients of a finite combination of complex sinusoids (ordered frequency that has sample value). Fast Fourier transform converts time (signal) to frequency by decomposing an N point time domain signal into N signals (Westin, Dougherty, et al., 2010) and Detrended Fluctuation Analysis is a method to determine self-
affection of a signal (Arora, et al., 2015b). Often Spectral Analysis (SA) is used in signal processing for PD’s motor symptom assessment. The magnitude of an input signal versus a certain frequency within the full range of the frequency is measured using a spectrum analyzer. Artificial Intelligence (Visual perception, decision making, image processing, and classification techniques) enables the development of computer systems to perform tasks that usually require human intelligence. For image processing, computer vision is a method to acquire, process and analyze a patient body’s images (like face and body posture). Machine learning in PD symptom assessment (Sama, et al., 2012) often includes techniques to assess the magnitude of addressed symptom. Linear discriminant analysis (LDA) classification method is used to optimally separate populations and reduce the dimensionality (Lopane, et al., 2015), (Adkin, et al., 2005). Non-parametric, generalized and multilayer perceptron analysis are different alternatives of LDA.

5 Discussion and conclusion

Table 1 summarizes the research studies that have evaluated the four cardinal PD motor symptoms. From left to right, it lists the evaluated symptoms, type of the instruments, calculated measures and employed analytical assessment methods.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Instrument</th>
<th>Measure</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tremor</td>
<td>Smartphone (3-D accelerometer, timer, finger tapping sensor)</td>
<td>X and Y coordinates, Time duration, 3-D Acceleration</td>
<td>Random forest machine learning technique, Detrended analysis</td>
<td>(Arora, et al., 2015b), 2015</td>
</tr>
<tr>
<td></td>
<td>Pen stylus</td>
<td>Acceleration</td>
<td>Non parametric</td>
<td>(Scanlon, et al., 2013), 2013</td>
</tr>
<tr>
<td></td>
<td>Real time wearable sensor</td>
<td>Acceleration</td>
<td>Shank, Ankle, Knee signal SA</td>
<td>(Bachlin, et al., 2010), 2010</td>
</tr>
<tr>
<td></td>
<td>Custom made goniometer</td>
<td>Angular velocity</td>
<td>SA of vertical leg acceleration</td>
<td>(Moreau, et al., 2010), 2010</td>
</tr>
<tr>
<td></td>
<td>Stride monitor system</td>
<td>Acceleration</td>
<td>Extension-flexion-component analysis</td>
<td>(Moore, et al., 2008), 2008</td>
</tr>
<tr>
<td></td>
<td>Isokinetic dynamometer Biodex System 3</td>
<td>3-D angular velocity, Anatomical zero</td>
<td>Spearman correlation</td>
<td>(Cano-de-la-Cuerda, et al., 2014), 2014</td>
</tr>
<tr>
<td>Dyskinesia</td>
<td>Digitized tablet (Spirography)</td>
<td>Velocity of drawing movements</td>
<td>Standard deviation analysis of drawing velocity</td>
<td>[2], 2005</td>
</tr>
<tr>
<td></td>
<td>Wrist accelerometer</td>
<td>Trunk acceleration, Shank velocity</td>
<td>Support vector Machine learning</td>
<td>(Sama, et al., 2012), 2012</td>
</tr>
<tr>
<td></td>
<td>Wrist-worn inertial sensor</td>
<td>Median angular velocity of trunk Rotation</td>
<td>Linear discriminant analysis</td>
<td>(Lopane, et al., 2015), 2015</td>
</tr>
<tr>
<td>Symptom Set</td>
<td>Device/Technique</td>
<td>Parameters Analyzed</td>
<td>Methodology/Technique</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Postural Instability</td>
<td>MTX Xsens sensor with 3-D accelerometer and 3-D gyroscope</td>
<td>Acceleration, Direction and Distance</td>
<td>Antero-posterior (AP), Medio-lateral (ML), and Vertical directions analysis</td>
<td>(Mancini, et al., 2011), 2011</td>
</tr>
<tr>
<td></td>
<td>Digital angular-velocity transducer</td>
<td>Velocity (pitch, roll, angle), Time</td>
<td>Linear discriminant analysis, Anova</td>
<td>(Adkin, et al., 2005), 2005</td>
</tr>
<tr>
<td></td>
<td>Accelerometer</td>
<td>Acceleration</td>
<td>Posture contextualization algorithm</td>
<td>(Ahlrichs, et al., 2016), 2014</td>
</tr>
<tr>
<td>Tremor and Dyskinesia</td>
<td>Accelerometer, Gyroscope, Infrared camera</td>
<td>Acceleration, Angular velocity and time</td>
<td>Genetic Algorithm spectral classification</td>
<td>[20], 2014</td>
</tr>
<tr>
<td>Tremor and Bradykinesia</td>
<td>Miniature uni-axial gyroscope</td>
<td>Angular velocity in roll, yaw and pitch direction</td>
<td>Biomedical signal processing (Spectrum Analysis)</td>
<td>(Salarian, et al., 2007), 2007</td>
</tr>
<tr>
<td>Tremor and Postural instability</td>
<td>Accelerometer</td>
<td>Mean velocity, Acceleration range, Mean acceleration</td>
<td>Hilbert–Huang transformation of postural parameters</td>
<td>(Mellone, et al., 2011), 2011</td>
</tr>
<tr>
<td>Bradykinesia and Dyskinesia</td>
<td>Digitized tablet (spiral and tapping) Pocket PC device</td>
<td>Radius, Time, Mean speed of correct proportion of taps</td>
<td>Wavelet transform and principal component analysis</td>
<td>(Westin, Ghiamati, et al., 2010), 2010</td>
</tr>
<tr>
<td></td>
<td>Ambulatory Multichannel accelerometer, Video recorder</td>
<td>Acceleration, Body position, Time, Gravitational force, Body segment angle</td>
<td>Direct current component, Discriminant, variance (Anova), regression analysis</td>
<td>(Dunnewold, et al., 1998), 2005</td>
</tr>
<tr>
<td></td>
<td>Kinetigraph(3-D accelerometer)</td>
<td>Time period, Wrist acceleration</td>
<td>Expert system approach</td>
<td>(Griffiths, et al., 2012), 2012</td>
</tr>
<tr>
<td>Rigidity, Bradykinesia and Dyskinesia</td>
<td>Digitized tablet with finger tapping and Spirography</td>
<td>Speed, Accuracy, Standard deviation of radial drawing velocity</td>
<td>Principal component analysis</td>
<td>(Westin, Dougherty, et al., 2010), 2010</td>
</tr>
</tbody>
</table>

According to table 1, rigidity and postural instability are mostly evaluated as single symptoms. However, among articles which research on combined symptoms assessments, bradykinesia (with tremor, dyskinesia, rigidity, and dyskinesia together) is mostly studied. Tremor is assessed in some studies as a single symptom, and also together with each of bradykinesia, dyskinesia, and postural instability symptoms.
A common sensor for symptom detection was the accelerometer that was mostly used for detecting the tremor, dyskinesia, and postural instability. Digitizing tablet is used almost for all types of symptoms. Smartphone (Arora, et al., 2015b) and Microsoft Kinect (motion detector, and depth sensor) (Yeung, et al., 2014) are the latest devices in the market used for this. Smartphones (new generation of sensing devices) could expand rapidly with PD motor symptom assessments. Angular sensor detectors are used to detect rigidity and postural instability as single symptoms, and they are also used to detect bradykinesia and dyskinesia together with tremor. Video recording is often required for clinicians' observational analysis. Wearable sensors (small, available, accurate, including high-resolution activity data, and flexible with body locations) are preferred for PD since it's a progressive chronic disease and symptoms need to be assessed continuously throughout the day. For this, the mobile applications and wrist watches are more preferred as they are currently part of almost everyone's daily accessories. However, their analysis methods and their validations are important and a question is whether the devices or clinical ratings will become the gold standard. Machine learning techniques are potentially good solutions in the development of assessment systems to determine the effectiveness of drug dosing. Tools that can effectively characterize the severity of symptoms and can discriminate between bradykinesia and dyskinesia are needed. Some successful products are Parkinson's Kinetigraph (Dunnewold, et al., 1998) , Kinesia devices (Mera, et al., 2012), and Rempark (Sama, et al., 2012).

References

Paper II
A smartphone-based system to quantify dexterity in Parkinson’s disease patients

Somayeh Aghanavesi1, Dag Nyholm2, Marina Senek2, Filip Bergquist3, Mevludin Memedi1,4,*
1Computer Engineering, School of Technology and Business Studies, Dalarna University
2Dept. of Neuroscience, Neurology, Uppsala University
3Dept. of Pharmacology, University of Gothenburg
4Informatics, Business School, Örebro University
*Author to whom correspondence should be addressed: Email: mevludin.memedi@oru.se; Tel.: +46-19 303 478;

Abstract
Objectives: The aim of this paper is to investigate whether a smartphone-based system can be used to quantify dexterity in Parkinson’s disease (PD). More specifically, the aim was to develop data-driven methods to quantify and characterize dexterity in PD.
Methods: Nineteen advanced PD patients and 22 healthy controls participated in a clinical trial in Uppsala, Sweden. The subjects were asked to perform tapping and spiral drawing tests using a smartphone. Patients performed the tests before, and at pre-specified time points after they received 150% of their usual levodopa morning dose. Patients were video recorded and their motor symptoms were assessed by three movement disorder specialists using three Unified PD Rating Scale (UPDRS) motor items from part III, the dyskinesia scoring and the treatment response scale (TRS). The raw tapping and spiral data were processed and analyzed with time series analysis techniques to extract 37 spatiotemporal features. For each of the five scales, separate machine learning models were built and tested by using principal components of the features as predictors and mean ratings of the three specialists as target variables.
Results: There were weak to moderate correlations between smartphone-based scores and mean ratings of UPDRS item #23 (0.52; finger tapping), UPDRS #25 (0.47; rapid alternating movements of hands), UPDRS #31 (0.57; body bradykinesia and hypokinesia), sum of the three UPDRS items (0.46), dyskinesia (0.64), and TRS (0.59). When assessing the test-retest reliability of the scores it was found that, in general, the clinical scores had better test-retest reliability than the smartphone-based scores. Only the smartphone-based predicted scores on the TRS and dyskinesia scales had good repeatability with intra-class correlation coefficients of 0.51 and 0.84, respectively. Clinician-based scores had higher effect sizes than smartphone-based scores indicating a better responsiveness in detecting changes in relation to treatment interventions. However, the first principal component of the 37 features was able to capture changes throughout the levodopa cycle and had trends similar to the clinical TRS and dyskinesia scales. Smartphone-based scores differed significantly between patients and healthy controls.
Conclusions: Quantifying PD motor symptoms via instrumented, dexterity tests employed in a smartphone is feasible and data from such tests can also be used for measuring treatment-related changes in patients.

Keywords: Parkinson’s disease; motor assessment; spiral tests; tapping tests; smartphone; dyskinesia; bradykinesia; objective measures; telemedicine
I. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder (Shin et al., 2016) and is characterized by degeneration of dopaminergic neurons in the substantia nigra. A common treatment for PD is levodopa. Over the course of the disease, levodopa dose and timing of intake have to be adjusted to optimize the therapeutic effect (Schapira, 2009). PD is a multidimensional, progressive disease and patients have different symptom profiles, which makes it difficult for healthcare professionals and patients themselves to assess and manage PD symptoms. From the clinical point of view, it is challenging to remotely and frequently determine the current motor state of the patient to determine whether the patient is under-medicated (a state in which the PD motor symptoms such as bradykinesia, tremor, rigidity, and others appear) or over-medicated (the appearance of hyper-kinetic movements related to excessive levels of medication). Therefore, assessing the current motor state of the patient is essential for deriving an optimal treatment strategy.

The current state of the art for assessing PD symptoms in clinical routine and studies is by using clinical rating scales based on observations and judgments of clinicians and medical history. The most commonly used clinical rating scale is the Unified PD Rating Scale (UPDRS) (Martínez-Martín et al., 1994), which is used to evaluate the presence, severity and progression of PD symptoms as well as symptom fluctuations. However, clinician-based measurements are not able to capture variations in symptoms on a day-to-day basis since they only reflect one brief point in time. To reveal the full extent of patients' condition and prevent a recall and reporting bias, the motor symptoms need to be captured frequently, before and after medication (Isacson et al., 2008). Combining the elements of common rating scales with frequent self-assessments and objective tests can also help with covering more aspects of the disease than what can actually be obtained by clinical ratings alone.

Recent advances in information and communication technologies have enabled remote and continuous monitoring of motor symptoms (Maetzler et al., 2013). Previous studies have shown that such technologies provide accurate and valid objective assessment of symptoms. It was previously reported that they may assist in identifying motor functions (On, Off and dyskinesia) (Andong Zhan et al., 2016; Galli et al., 2014). The technology-based measures not only generate more valid endpoints for clinical studies but also can be useful in routine clinical care. There is a growing interest in investigating how useful the measures are when providing feedback to patients to increase their symptom and treatment outcome awareness (Bot et al., 2016).

From the technological point of view, data from different kinds of sensors during standardized tests and passive monitoring of physical activity have been previously analyzed and processed using signal processing and machine learning methods (Griffiths et al., 2012; Zeng et al., 2016). There are different studies with the focus on quantifying various motor symptoms. Some have focused on assessing motor dysfunctions in upper extremities (Heldman et al., 2014; Westin, et al., 2010), some on gross motor symptoms like gait (Mariani et al., 2010), while others on combination of both. For instance, Tsipouras et al. (2012) analyzed data from accelerometers and gyroscopes, which were placed on different parts of patients’ bodies with the aim of quantifying drug-induced involuntary movements or dyskinesia, using Fourier transform. A similar approach was employed by Salarian et al. (2007) to quantify bradykinesia and tremor. Other studies have focused on analyzing data from upper limbs during standardized tasks like
finger tapping (Heldman et al., 2014; Shima et al., 2009), digital spiral analysis (Saunders-Pullman et al., 2008) and quantitative digitography (Taylor Tavares et al., 2005; Giovannoni et al., 1999).

As an alternative to wearable sensors-based systems, some research groups have focused on assessing dexterity performance of PD patients by analyzing upper limb motor data collected by means of touch screen devices (Saunders-Pullman et al., 2008; Liu et al., 2005; Haubenberger et al., 2011). The touch screens of the smartphones record physical properties of movements that can be produced either by a pen tip or finger with great spatial and temporal precision. Such smartphone measurements were previously used for assessing different fine motor dysfunctions like tremor (Haubenberger et al., 2011), dyskinesia (Liu et al., 2005), drawing impairments (Westin et al., 2010) and global tapping performance (Memedi et al., 2013). Quantitative measures during alternating tapping tests and digital spiral analysis have been previously used as measures of bradykinesia (Giovannoni et al., 1999) and severity of PD symptoms (Saunders-Pullman et al., 2008). To our knowledge, there is no study reporting an approach where tapping and spiral drawing test data were combined in data-driven manner and related to objective measures such as various clinical ratings and actual treatment.

The purpose of this paper was to investigate whether a smartphone-based system, which consists of tapping and spiral drawing tests, can be used for quantifying dexterity in advanced PD. The paper reports clinimetric properties of smartphone-based measures of dexterity including correlations to clinical rating scales, test-retest reliability, sensitivity to treatment interventions, and ability to differentiate between tests performed by patients and healthy controls.

2. Materials and Methods

2.1 Participants

Nineteen PD patients and 22 healthy controls were recruited in a single center, open label, single dose clinical trial in Uppsala, Sweden (Table 1, Senek et al., 2017). Written informed consent was given after approval by the regional ethical review board (in Uppsala, Sweden).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Mean age (years)</th>
<th>Mean height (m)</th>
<th>Mean weight (kg)</th>
<th>Years with the disease</th>
<th>Years on levodopa</th>
<th>Hoehn &amp; Yahr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>14 males 5 females</td>
<td>71.4 (6.3)</td>
<td>1.75 (0.09)</td>
<td>75.4 (11)</td>
<td>9.7 (6.8)</td>
<td>9.5 (6.5)</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>16 males 6 females</td>
<td>64.2 (7.4)</td>
<td>1.75 (0.1)</td>
<td>83.6 (13.8)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. Characteristics, mean (standard deviation) of patients and healthy subjects.
2.2 Data collection

The trial included a single levodopa-carbidopa dose experiment for the PD patients, where both patients and healthy controls were asked to perform dexterity tests (tapping and spiral drawing) using a smartphone before and at specific time intervals after a dose was given (Senek et al., 2017; Westin et al., 2010). For the patients, the dose administered was 150% of their individual levodopa equivalent morning dose to follow transitions between Off, On, and On with dyskinesia motor states. Up to 15 samples per PD patient were collected, one measurement at baseline (20 minutes prior to dosing), one at the time of dose administration (0 minutes) and thereafter follow-up measurements at 20, 40, 80, 110, 140, 170, 200, 230, 260, 290, 320, and 360 minutes after dose administration. The healthy controls were asked to perform the tests, 8 times each, at time point 0 (first test) and then at 20, 40, 60, 80, 110, 140, and 170 minutes, without receiving any medication.

On each test occasion, subjects performed upper limb motor tests (tapping and spiral drawings), using a smartphone (Figure 1). The smartphone had a 4" (86 x 53 mm) touch screen with a 480 x 800 pixels and recorded both position (x and y coordinates) and time-stamps (in milliseconds) of the pen tip. The subjects were instructed to be seated on a chair and perform the tests using an ergonomic pen stylus with the device that was placed on a table and supporting neither hand nor arm. During tapping tests, they were asked to alternately tap two fields, as shown on the screen of the device, as fast and accurate as possible, using first right hand and then left hand. The time to complete a tapping test was 20 seconds. During the spiral tests, the subjects were instructed to trace a pre-drawn Archimedes spiral as fast (within 10 seconds) and accurately as possible, from the center out, using the dominant hand. The test was repeated three times per test occasion. The total number of measurements with the smartphone for PD patients was 285, and for healthy controls was 176.

2.3 Clinical assessments of motor symptoms

Along with smartphone-based measurements, patients were video recorded while performing standardized motor tasks according to UPDRS at the above-mentioned time points. The recorded videos were presented in a randomized order to three movement disorder specialists, so that the ratings were blinded with respect to time from dose administration. The specialists rated three UPDRS-part III (motor examination) items including UPDRS item #23 (finger tapping), UPDRS #25 (rapid alternating movements of hands), and UPDRS #31 (bradykinesia), according to the definitions of the motor examination part of the UPDRS (Fahn et al., 1987). For items #23 and #25 the specialists were asked to assign a single score per time point without reference to any hand. The specialists also rated dyskinesia on a severity scale from 0 to 4 (Goetz et al., 2008) and overall mobility according to Treatment Response Scale (TRS) (Nyholm et al., 2005), ranging from -3 (very Off) to 0 (On) to +3 (very dyskinetic). For
every scale, mean scores per time point for the three specialists were calculated and used in subsequent analysis.

Figure 1. Implementation of dexterity tests (tapping and spiral drawing) on the smartphone.

2.4 Data processing and Analysis

2.4.1 Feature extraction

The raw dexterity data were processed with time series analysis methods to calculate 37 spatiotemporal features, which represent the severity of symptoms. Different kinematic quantities, including time, distance, speed, and velocity were used as primary signals to be processed and analyzed using time- and wavelet-domain methods. First, 20 tapping and 10 spiral features were calculated based on previous publications. The first 30 features are listed in Table 2.

<table>
<thead>
<tr>
<th>Tapping features (Reference to previous works)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Total Number of taps (Macleod, et al., 2010)</td>
</tr>
<tr>
<td>(2) Mean tapping time difference between two fields (Yahalom, et al., 2004)</td>
</tr>
<tr>
<td>(3) Mean tapping speed from right to left (Memedi, et al., 2013)</td>
</tr>
<tr>
<td>(4) Mean tapping speed from left to right (Memedi, et al., 2013)</td>
</tr>
<tr>
<td>(5) Coefficient of variation of tapping speed from right to left,</td>
</tr>
<tr>
<td>(6) Coefficient of variation of tapping speed from left to right (Taylor Tavares, et al., 2005)</td>
</tr>
<tr>
<td>(7) Mean distance from the centers of the fields (Arora, et al., 2015a)</td>
</tr>
<tr>
<td>(8) Coefficient of variation of distances from the center fields (Arora, et al., 2015a)</td>
</tr>
</tbody>
</table>
(9) Overall distribution of the taps (Memedi, et al., 2013)

(10) Mean distance from center (Arora, et al., 2015a)

(11) Mean tapping speed per cycle (Memedi, et al., 2013)

(12) The absolute mean difference between the first and second part of the time series signal (Memedi, et al., 2013)

(13) Approximate entropy (ApEn) measure of the two parts of the time series signal (Memedi, et al., 2013)

(14) Measure of tapping speed reduction (Arora, et al., 2015a)

(15) Measure of tapping reaction time (Yahalom, et al., 2004)

(16) Overall trend of tapping reaction time during the test trial (Yahalom, et al., 2004)

(17) ApEn measure of mean tapping speed (Memedi, et al., 2013)

(18) The amount of irregularity in vertical tap distance (Memedi, et al., 2013)

(19) Measure of variation in distance between the two fields (Arora, et al., 2015a; Memedi, et al., 2013)

(20) Measure of irregularity caused by time variations during tapping (Memedi, et al., 2015)

Spiral features (Reference to previous works)

(21) Mean drawing speed (Stanley, et al., 2010)

(22) Coefficient of variation of speed (Memedi, et al., 2015)

(23) Skewness of the speed (Memedi, et al., 2015)

(24) Radial velocity (Hess, et al., 2014)

(25) Mean time difference (Banaszkiewicz, et al., 2009)

(26) Global minima of the drawing speed,

(27) Global maxima of the drawing speed (Memedi, et al., 2015)

(28) ApEn measure of drawing speed,

(29) ApEn measure of radial velocity (Memedi, et al., 2016a)

(30) Coefficient of variation of high frequency wavelet coefficients by discrete wavelet transform (DWT) (Memedi, et al., 2015)

Table 2. Spatiotemporal features calculated from tapping and spiral data

In addition to the aforementioned features, 7 new spiral features were calculated and used in the feature set. The rationale behind including more features was to cover more symptom information from the dexterity tests.

(31) Kurtosis (fourth standardized moment) of the drawing speed signal was calculated as following:

\[
K = \frac{E(s - \mu)^4}{\sigma^4}
\]  

where \( s \) is the distribution of the drawing speed per second, \( \mu \) is the mean of \( s \), \( \sigma \) is the standard deviation of \( s \), and \( E(s - \mu) \) is the expected value of the \( s - \mu \) quantity. Kurtosis computes a sample version of this population value and measures how
outlier-prone the distribution of the speed is. Computing the kurtosis of drawing speed was to quantify the amount of delays, abruption and continuation of movements.

(32) The $x$ points of spiral drawing were retrieved and mapped over time. The measure of kurtosis for the series of $x$ coordinates were calculated. This measure quantified the amount of horizontal deviations from the original spiral.

(33) Similarly, the $y$ points of spiral drawing were retrieved and mapped over time. The measure of kurtosis for the series of $y$ coordinates were calculated. This measure quantified the amount of vertical deviations from the original spiral.

(34) Length of the spiral drawing was measured using the parametric Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) approximation and numerical integration over the segments of the spiral drawing (Hoogendam, et al., 2014); (Miralles, et al., 2006). The spiral drawing curve length is associated with the deviations from the template (original spiral) and was used as a measure to quantify the impaired drawing.

(35) The area of the spiral drawing was calculated using the trapezoidal method to extract the region of the curve drawn by the subjects. This is done by breaking the whole area down into trapezoids with easily computable areas. The integration over an interval of every two consecutive points from spiral drawing was calculated and accumulated together to obtain the total area.

Equation (2) shows the formula to calculate the integration between the two points.

$$\int_{x_n}^{x_{n+1}} f(x)dx = \frac{b-a}{2N} \sum_{n=1}^{N} (f(x_n) - f(x_{n+1})) \quad (2)$$

Where $N$ is the total number of points, and the spacing between each point is equal to the scalar value $\frac{b-a}{2}$. There is a relation between the size of the spiral and the speed of the drawing movements. According to (Longstaff, et al., 2003), it is more likely that the larger spirals will be drawn faster than the smaller spirals. The increasing size of the spiral drawing increases the coordination requirements, it is therefore concluded that larger spirals are drawn with greater degree of variability than smaller spirals.

(36) Spiral drawing total time is defined as the time that was required to draw the spiral on the smartphone. It is the time difference between first and last captured points from the smartphone.

(37) The $x$ and $y$ points of the spiral drawing were retrieved from the smartphone with their respective time stamps. An ideal Archimedean spiral has a constant speed during the execution, which means the time-stamp at each point increases constantly. Slower the movements greater the time difference between points. The time differences between consecutive points were mapped over time. Using the alternating nature of the derivatives, the magnitudes of the identified peak points were calculated from a series of time differences. In addition, the sum of the magnitudes was calculated to represent the amount of delays in the spiral drawing execution.

Since there were two trials that were performed during tapping tests (first right hand and then left hand), individual features of both trials were averaged and used in the following analysis. Similarly, for spiral tests the average of the features were calculated for the three trials.
2.4.2 Principal component analysis

To reduce the dimensions of the features but keep the most important and related information into a smaller set, principal component analysis (PCA) using the correlation matrix method was applied on the 37 features. Theoretically, PCA is a linear dimension reduction technique that uses a rectangular transformation to convert the set of correlated features into a set of values of linearly uncorrelated variables. Seven principal component scores (PCs) having eigenvalues higher than 1 were retained and used in subsequent analysis. Applying this threshold resulted in retaining 71% of the variation in data.

2.4.3 Machine learning

The PCs were used as predictors to supervised machine learning methods used to map to the mean ratings of the three movement disorder specialists on the clinical rating scales used in the clinical trial. Four machine learning methods were evaluated, using the Weka datamining toolkit (Eibe Frank, et al., 2016): support vector machines (SVM), linear regression (LR), regression trees (RT), and multilayer perceptron artificial neural networks (MLP). A stratified 10-fold cross validation was applied to test the performance of the machine learning methods. For each of the five scales, separate models were built and tested.

2.4.4 Statistical analysis

The performance of the machine learning methods was assessed by correlation coefficients between the predicted and mean clinical ratings. One-way consistency intra-class correlation coefficients (ICC) were calculated to assess the agreements between the three specialists’ ratings and test-retest reliability of mean specialist and smartphone-based scores between the first two baseline measurements. To test the relevance of the tapping and spiral features when used as predictors in the machine learning methods bidirectional stepwise regression approach was employed using sumUPDRS (the sum of UPDRS #23, UPDRS #25 and UPDRS #31) ratings of the three individual raters as response variables. To investigate differences in mean PCs between the groups patients and healthy controls, linear mixed effects models based on a restricted maximum likelihood estimation method were employed. Group was considered as a fixed effect and subject ID as a random effect. The relative ability to detect change from baseline (no medication) to follow up time points when patients were on medication was assessed by effect sizes. To calculate effect sizes, ANOVA models were fitted for each time point after the baseline test; first test and second test; first test and third test, and so on. A high effect size indicates that a scale is sensitive to treatment response (Goetz et al., 2013). The statistical analyses were performed in R and Minitab statistical software.
3. Results

3.1 Feature evaluation

The 17 features that were the most relevant as predictors of sumUPDRS when using individual ratings of the three movement disorder specialists as response variables are listed in Table 3. Six (3 tapping and 3 spiral) of the 17 features were selected as significant predictors of sumUPDRS by the three separate regression models. The remaining features were either selected by two or one regression model.

<table>
<thead>
<tr>
<th>Feature#</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>13</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>19</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>23</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>31</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>36</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>37</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 3. The most relevant tapping and spiral features using bidirectional elimination by stepwise regression

3.2 Inter-rater agreements

ICCs between the three specialists were moderate to strong: 0.61 for UPDRS #23, 0.52 for UPDRS #25, 0.58 for UPDRS #31, 0.65 for sum of the UPDRS #23, UPDRS #25 and UPDRS #31 (sumUPDRS), 0.8 for TRS, and 0.67 for dyskinesia. These results indicate that for all scales there is an inter-rater variability to some degree. A mean rating per time point and item was calculated and used as a dependent variable when training and evaluating the machine learning methods.
3.3 Correlations between Predicted and Clinical Scores

The absolute correlation coefficients between mean clinical ratings and predicted scores ranged from weak to moderate (Table 4). The best performing method was SVM and had correlations coefficients as follows: 0.52 for UPDRS #23, 0.47 for UPDRS #25, and 0.57 for UPDRS #31, 0.46 for sumUPDRS, 0.59 for TRS, and 0.64 for dyskinesia.

<table>
<thead>
<tr>
<th></th>
<th>SVM</th>
<th>LR</th>
<th>RT</th>
<th>MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPDRS #23</td>
<td>0.52</td>
<td>0.13</td>
<td>0.33</td>
<td>0.13</td>
</tr>
<tr>
<td>UPDRS #25</td>
<td>0.47</td>
<td>0.13</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>UPDRS #31</td>
<td>0.57</td>
<td>0.27</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>sumUPDRS</td>
<td>0.46</td>
<td>0.15</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>TRS</td>
<td>0.59</td>
<td>0.39</td>
<td>0.39</td>
<td>0.30</td>
</tr>
<tr>
<td>Dyskinesia</td>
<td>0.64</td>
<td>0.56</td>
<td>0.57</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 4. Absolute correlation coefficients between mean ratings of the three specialists and predicted scores which were derived from the support vector machines (SVM), linear regression (LR), regression trees (RT), and multilayer perceptron (MLP). UPDRS #23 is Finger Taps, UPDRS #25 is Rapid Alternating Movements of Hands, UPDRS #31 is Body Bradykinesia and Hypokinesia, sumUPDRS is the sum of UPDRS #23, UPDRS #25 and UPDRS #31, and TRS is the Treatment Response Scale.

3.4 Test-Retest Reliability

The ICCs between the first two baseline measurements were calculated. The data for this analysis included measurements at test occasions before patients received the dose and at the moment the dose was administered. The results showed that the mean clinician ratings had better test-retest reliability than the scores derived by the SVM model (Table 5). The SVM scores had good repeatability when assessing TRS and dyskinesia but not for the UPDRS items.

<table>
<thead>
<tr>
<th>Clinical scores</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPDRS #23</td>
<td>0.74</td>
</tr>
<tr>
<td>UPDRS #25</td>
<td>0.62</td>
</tr>
<tr>
<td>UPDRS #31</td>
<td>0.87</td>
</tr>
<tr>
<td>sumUPDRS</td>
<td>0.91</td>
</tr>
<tr>
<td>TRS</td>
<td>0.94</td>
</tr>
<tr>
<td>Dyskinesia</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5. ICCs between the first two baseline measurements for smartphone-based scores and mean ratings of the three clinicians.
3.5 Sensitivity to treatment changes

The most sensitive scales were the clinician-based TRS and dyskinesia. The PC1 had lower sensitivity but, in general, was capable of capturing changes in symptom severity in response to levodopa medication. It could also capture improvements/deteriorations in symptoms throughout the levodopa test cycle i.e. during transitions between different motor states of patients, from Off to On (normal mobility) and/or On with dyskinesia and the wearing Off effects (Figure 2).

![Figure 2. Sensitivity assessment of PC1 and mean ratings of the three movement disorder specialists on the three UPDRS items, TRS and dyskinesia across the levodopa test cycle for all patients. The first data point in the X axis represents the change in scores between the first two baseline (without medication) measurements. The second data point represents the change in scores between first baseline and third measurement, and so on. Number of tests per time slot: 0 (n=19), 20 (19), 40 (n=19), 60 (n=19), 80 (n=18), 110 (n=17), 140 (n=17), 170 (n=17), 200 (n=17), 230 (n=17), 260 (n=14), 290 (n=14), 320 (n=11), and 360 (n=11).](image)

3.6 Separation between patients and healthy subjects

When assessing the ability of the PCs to differentiate between tests performed by patients and healthy controls, the mean scores of 3 (PC1, PC2 and PC4) out of the 7 PCs were significantly different between the two groups (p<0.005). Summary statistics of the 7 PCs for both the groups are shown in Table 6.
<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.9</td>
<td>&lt;0.001</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>Mean</td>
<td>1.27</td>
<td>-0.78</td>
<td>-0.00</td>
<td>0.62</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.84</td>
<td>1.97</td>
<td>1.97</td>
<td>1.55</td>
<td>1.53</td>
</tr>
<tr>
<td>Patients</td>
<td>Mean</td>
<td>-0.98</td>
<td>0.60</td>
<td>0.00</td>
<td>-0.48</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.73</td>
<td>2.45</td>
<td>2.02</td>
<td>1.95</td>
<td>1.62</td>
</tr>
</tbody>
</table>

Table 6. Summary statistics of the first 7 PCs for patients and healthy controls.

4. Discussion and conclusions

In this study, smartphone-generated dexterity measurements were used to quantify the motor performance of PD patients during repeated tapping and spiral drawings tasks. The methods developed in this study were evaluated using measurements from 19 PD patients during a single levodopa dose experiment and 22 healthy controls. The obtained results indicate that the methods could capture motor symptoms reasonably well as compared to the mean assessments of three movement disorder specialists on three items of UPDRS-III, TRS and dyskinesia scales. The correlations were weak to moderate between the scores derived by the methods and the mean clinical ratings, indicating that tapping and spiral drawing tests capture relevant symptom information corresponding to the clinical rating scales. In contrast to the clinical rating scales, another advantage with the current system is that PD-related outcomes can be captured and assessed more frequently.

During clinical assessments, the movement disorder specialists observed the patients while performing standardized motor tasks as defined in the UPDRS scale where the highest weight was given to the symptoms that were prominent during gross motor performance e.g., walking ability. During tapping and spiral drawing tasks, only fine motor movements could be recorded by the smartphone touch screen. This may explain the moderate agreements between SVM and mean clinician ratings, which in turn suggests further work for complementing and fusing dexterity measurements with data from wearable sensors or inertial measurement units of smartphones that are collected during gross motor tasks. Furthermore, based on the correlation coefficients (Table 4) we can notice that the tapping and spiral drawing tests contained relevant information about motor function of patients. In addition, the results from the feature selection (Table 3) indicate that not all of the tapping and spiral features were equally represented in the regression models when using individual ratings on sumUPDRS as response variable. These results may reflect the moderate agreements on the clinical ratings by the three raters.

The clinimetric properties of the motor tests were previously assessed in a longitudinal 36 months clinical trial in Sweden and a two weeks trial in Italy (Memedi et al., 2013; Memedi et al., 2015). In those studies, it was found that data from such tests can be used to measure PD progression over time and to separate patients in different disease stages. Spiral drawing tests were also shown to be useful in automating the process of scoring the Off symptoms and dyskinesia in PD patients (Sadikov et al., 2017). The effect of handedness in the right-handed
patients (84% of the patients) was investigated and the results indicated that the effect of
handedness was more prominent than the effect of the side in which PD symptoms started since
they had better tapping results with the right hand than with the left hand. Smartphones have
previously been tested in detecting and assessing the severity of PD symptoms. (Arora et al.,
2015) combined smartphone data after the PD patients performed a battery of tests including
voice, posture, gait, finger tapping, and reaction time. The results indicated that multimodal
smartphone data can be useful for diagnosis purposes as well as monitoring progression of PD
symptoms.

In our study, dexterity measurements of the smartphone were related to corresponding
clinical ratings and changes in symptoms during single dose experiments. The results show that
combining spatiotemporal features extracted from tapping and spiral drawing data can be used
to detect treatment-related changes in advanced PD. Although the PC1 had a lower sensitivity
when compared to mean clinical ratings on TRS and dyskinesia, we can conclude that PC1 alone
could significantly detect changes in symptoms to the first test on medication (20 minutes
post-dose, Figure 2). In addition, the PC1 could follow transitions between motor states across
the levodopa test cycle since it had similar trends as the TRS and dyskinesia scale. These results
suggest that tapping and spiral drawing tests with the smartphone can detect movements
reasonably well related to under- and over-medication. This could be due to the fact that raw
tapping and spiral data were processed with ApEn and DWT methods. The ApEn in general
measures the amount of irregularity in a signal and could be useful in capturing different
irregular movement patterns during the test trial, which could be related to dyskinesia. The
DWT employs a multiresolution analysis of a signal by separating low-frequency components
from high-frequency components. In our work, the level and variation in frequency components
was derived by calculating mean and standard deviation of the wavelet coefficients. These
features could be useful in quantifying movements related to under- and over-medication.

As a limitation of this study, there was a considerable amount of inter-rater variability. This
is a natural problem when dealing with subjective ratings. For instance, in the study performed
by (Heldman et al., 2014) the raters differently weighted speed, amplitude and rhythm while
observing video recordings of PD patients during finger tapping tasks. The discrepancies in
assessments could be related due to the fact that in our study there was no training of the raters
and/or due to the natural within- and between-rater variability when using scales (Post et al.,
2005). One possible step to reduce the inter-rater variability from the mean rating would be to
include more raters. Future research will focus on improving the performance of the methods
by including spatiotemporal features from wearable sensors (e.g. during gait) into a feature set
that can be used during data-driven modelling. In addition, it would be interesting to
investigate correlations between standard dexterity tests like Purdue Pegboard test and the
measures derived from the tapping and spiral drawing tests of the smartphone.

In conclusion, the results presented in this paper indicate that tapping and spiral drawing
tests of the smartphone contain relevant symptom information for detecting and assessing PD
dexterity. The results suggest that the tests can be useful in detecting changes in motor
symptoms related to treatment.
Acknowledgments
This work is performed in the framework of the FLOAT and MUSYQ projects, funded by Swedish knowledge foundation and Swedish innovation agency (Vinnova), in collaboration with Cenvigo AB, Sensidose AB, Swedish ICT Acreo, Uppsala University and Dalarna University.

Conflict of interest
None.

References


control from medication and deep brain stimulation. *Mov Disord*, 20(10), 1286-1298. doi: 10.1002/mds.20556


Paper III
A method for measuring Parkinson’s disease related temporal irregularity in spiral drawings

Mevludin Memedi, Somayeh Aghanavesi, and Jerker Westin

Abstract—The objective of this paper was to develop and evaluate clinimetric properties of a method for measuring Parkinson’s disease (PD)-related temporal irregularities using digital spiral analysis. In total, 108 (98 patients in different stages of PD and 10 healthy elderly subjects) performed repeated spiral drawing tasks in their home environments using a touch screen device. A score was developed for representing the amount of temporal irregularity during spiral drawing tasks, using Approximate Entropy (ApEn) technique. In addition, two previously published spiral scoring methods were adapted and their scores were analyzed. The mean temporal irregularity score differed significantly between healthy elderly subjects and advanced PD patients (P<0.005). The ApEn-based method had a better responsiveness and test-retest reliability when compared to the other two methods. In contrast to the other methods, the mean scores of the ApEn-based method improved significantly during a 3 year clinical study, indicating a possible impact of pathological basal ganglia oscillations in temporal control during spiral drawing tasks. In conclusion, the ApEn-based method could be used for differentiating between patients in different stages of PD and healthy subjects. The responsiveness and test-retest reliability were good for the ApEn-based method indicating that this method is useful for measuring upper limb temporal irregularity at a micro-level.

I. INTRODUCTION

Parkinson’s disease (PD) is a neurological disorder associated with motor (e.g. bradykinesia, rigidity, tremor) and non-motor (e.g. constipation, depression, eye movement disorders) symptoms. One of the disabilities associated with PD is the impaired ability to accurately time movements [1]. The research suggests that the basal ganglia (BG) play a significant role in temporal processing at milliseconds to seconds range, also known as “interval timing” [2]. The speed of a hypothetical “internal clock” is controlled by basal ganglia and is related to the brain dopaminergic levels [1]. The deterioration of dopaminergic neurons among PD patients is associated with motor symptoms but also with disruption in repetitive movements. Many studies have shown an increase in timing variability among patients when compared to healthy subjects, suggesting the BG has a role in interval timing. For instance, in one study it was shown that treated PD patients had poorer timing control than untreated patients when modulating gait timing during externally-cued conditions [3]. Therefore, relating objective measures obtained by instrumented tests to pathological BG fluctuations would be beneficial for

*Research supported by Swedish Knowledge Foundation and Vinnova.
Authors are with the School of Technology and Business Studies, Computer Engineering, Dalarna University, Falun 791 88, Sweden (e-mail: mmi@du.se; saa@du.se; jwe@du.se).

978-1-5090-2455-1/16/$31.00 ©2016 IEEE
facilitating the assessment of high frequency motor irregularities that could be difficult to be assessed visually [4] [5].

In the present study we investigated the upper limb temporal irregularity of patients in different stages of PD and healthy elderly subjects during spiral drawing tasks. The amount of temporal irregularity during spiral drawing tasks was quantified using an Approximate Entropy (ApEn)-based method. In addition to this method, two more spiral scoring methods [6] and [7] were adopted and their scores were compared to the ApEn-based method. Specific objectives of the study are to: i) investigate mean temporal irregularities between patients and healthy elderly subjects, ii) assess the responsiveness of the three methods, iii) assess the test-retest reliability of the three methods, and iv) investigate whether long-term trend of the methods can be used as indicators of a possible involvement of BG in temporal control during spiral drawing tasks.

II. METHODS

A. Subjects

The results presented in this paper are based on data from two clinical studies, both of which were approved by the relevant agencies and informed consent was given. In total, 98 PD patients in different stages of PD and 10 healthy elderly (HE) subjects participated. Of the 98 patients, 65 patients with advanced idiopathic PD were recruited in an open longitudinal 36-month study at nine clinics in Sweden [8]. On inclusion, 35 of these patients were treated with levodopa-carbidopa intestinal gel infusion (LCIG) and 30 patients were candidates for switching from conventional oral PD treatment to LCIG. In the latter group, the patients were LCIG treatment-naive at study start. In the second study, 38 patients with a clinical diagnosis of idiopathic PD in Milan, Italy participated [9]. The Italian patients were divided into two groups: intermediate stage patients experiencing motor fluctuations (n = 16) and clinically stable, early PD patients (n = 17). Characteristics of the patients and HE subjects are shown in Table I.

B. Experimental setup

Both patients and HE subjects repeatedly used a touch screen telemetry device in their home environments [10]. Assessments with the device were performed four times per day during week-long test periods. The Swedish LCIG-naive patients used the device at baseline (before LCIG), month 0 (first visit; at least 3 months after permanent intraduodenal LCIG), and thereafter quarterly for the first year and biannually for the second and third years. The LCIG-non-naive patients used the device from the first visit that is month 0. In 23 LCIG-naive patients, assessments with the device were available during baseline period and at least one test period after having started LCIG. Hence, n = 23 in the LCIG-naive group. The Italian patients used the device for two week-long test periods with a washout week in between. The HE subjects used the device for one week-long test period.
Table I. CHARACTERISTICS OF PD PATIENTS AND OF HEALTHY ELDERLY PARTICIPANTS, PRESENTED AS MEDIAN ± INTERQUARTILE RANGE.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Swedish study (advanced patients)</th>
<th>Italian study (Intermediate patients)</th>
<th>Italian study (Early patients)</th>
<th>HE (Healthy elderly)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (n, gender)</td>
<td>65 (43m; 22f)</td>
<td>17 (13m; 2f)</td>
<td>16 (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>
</tbody>
</table>

* Assessments performed in the afternoon. ** Assessments performed in the On state.

Abbreviation: HE, healthy elderly; NA, not applicable.

On each test occasion, the subjects were asked to trace a pre-drawn Archimedes spiral that was shown on the screen of the device, using the dominant hand. The subjects were instructed to perform the spiral tests, using an ergonomic pen stylus. In addition, they were instructed to trace the pre-drawn spiral from the center out, as accurately and fast as possible, supporting neither hand nor arm, with the device placed on a table and to be seated in a chair. The spiral test was repeated three times per test occasion and the subjects were instructed to complete it within 10 seconds. The device had a 3.5" touch screen, 240 X 320 pixel resolution. Position and time-stamps (in ms) of the pen tip were recorded and stored for offline processing.

III. FEATURE EXTRACTION

The digitized spiral signals were processed using the following three methods. A new method based on Approximate Entropy (ApEn) was developed to generate a temporal irregularity score (TIS). In addition, two more methods developed by Liu et al. [6] and Westin et al. [7] were adopted and applied on the spiral data and their corresponding scores were compared to TIS.

A. The method for quantification of temporal irregularity during spiral drawing tasks

Initially, drawing speed (DS) was calculated as a rate of spatial change with respect to time, using the following equation:

\[
DS = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i}
\]

(1)

where \(x\) is the horizontal coordinate of pixels on the screen, \(y\) is the vertical coordinate and \(t\) is the time in seconds. Next, the ApEn technique was applied on DS signals to generate the score.
The ApEn is a statistical method for measuring the repeatability of patterns within a signal [11]. A signal containing a single frequency component is associated with a relatively small ApEn value whereas more complex signals containing multiple frequency components are associated with high ApEn values, indicating a high level of irregularity. ApEn reflects the similarity between a chosen window of a given duration and the next set of windows of the same duration. ApEn requires determination of two parameters: \( m \) (length of the window being compared) and \( r \) (measure of similarity), which must remain fixed during all calculations. In the current work and after experimentation, \( m \) was set to 4 and \( r \) to 0.2 (20% of the standard deviation of the signal). The derived ApEn value was then corrected for total drawing completion time. The resulting score is hence on denoted ‘temporal irregularity score’ (TIS).

**B. Other methods for measuring PD motor symptoms based on digital spiral analysis**

Two previously published methods by Liu et al. [6] and Westin et al. [7] were adopted and applied on the spiral data and their scores were used in subsequent analysis. The method developed by Liu et al. [6] was designed to quantify the severity of drug-induced dyskinesias in the upper limbs, using digital spiral analysis. Drawing velocity signals in horizontal, vertical, radial and tangential directions were bandwidth filtered in the range of 1-5 Hz to extract frequency components representing dyskinetic movements. The standard deviation of the frequency transformed signals in each direction was then calculated and the mean standard deviation was used to represent the extent of dyskinesia. Since our device had a lower sampling rate than the one used by Liu et al. [6], their method was modified and its score was compared to TIS. Initially, the radial velocity (RV) was calculated using the following equation:

\[
RV = \frac{r_{i+1}-r_i}{t_{i+1}-t_i}
\]  

(2)

where \( r \) is the radius defined as the square root of the sum of the squares of \( x \) and \( y \) coordinates. Next, the RV was interpolated and bandwidth filtered using Chebyshev Type I filter in the frequency range of 1-5 Hz. Finally, the standard deviation of the filtered RV signal was calculated to form a dyskinesia score, denoted SD-DV (standard deviation of drawing velocity). In the work performed by Westin et al. [7], the digitized spiral data were processed to yield a score for representing spiral drawing impairment, using discrete wavelet transform and principal component analysis (PCA). A 3-level decomposition using a Daubechies (db10) wavelet function family was performed on radius signal to obtain low- and high-frequency components represented by wavelet coefficients. In order to reduce the dimensionality of these coefficients, the PCA was applied and the first principal component was calibrated and linearly transformed into a scale between 0 and 10 where 0 represented no impairment and 10 represented extremely severe impairment. The resulting score is hence on denoted the ‘wavelet spiral test score’ (WSTS) and used in subsequent analysis.

**C. Statistical analysis**

Differences in mean scores of the three methods across the four subject groups were assessed using linear-mixed effects (LME) models with subject ID as a random effect and group as fixed effect of interest. For group (HE vs. advanced PD) classification, three machine learning methods: Logistic Regression (LR), C4.5 decision tree, and random Forests (RF). The machine
learning methods were tested with stratified 10-fold cross validation. The relative ability of the methods to detect change from baseline (oral treatment) to month 0 (LCIG) was determined by effect size correlation coefficient, representing the magnitude of treatment effect [cf. e.g. 12]. The score that had the highest coefficient value was defined as the most sensitive to treatment response. The long-term progressions of the three scores were assessed using LME models in combined data from the two groups of Swedish advanced patients starting from the first test period (month 0) with LCIG treatment. Correlations between the three scores were assessed using Pearson correlation coefficients. Test-retest reliability of the three scores was assessed after taking mean of the three possible correlations between the three spiral test trials.

IV. RESULTS

When comparing mean scores between the four subject groups, the mean TIS score was significantly different between HE subjects and advanced patients (P<0.005, Figure 1). The mean TIS score did not differ between HE and the other two groups i.e. early (P=0.62) and intermediate (P=0.09) patients. The mean TIS score was significantly different only between HE and advanced groups (P<0.05). In contrast to TIS and WSTS, there were no significant differences in mean SD-DV scores across the groups.

WSTS and SD-DV scores were correlated to each other with a coefficient of 0.69 (P<0.001). However, TIS was uncorrelated to neither WSTS nor SD-DV with coefficients of 0.11 and 0.12, respectively, indicating that TIS measures a different aspect of upper limb motor performance in PD patients than WSTS and SD-DV. Table II shows the performance of the three classifiers.
when classifying between HE subjects and advanced PD patients, using first WSTS and TIS individually and then both of them as inputs to the classifiers. The rationale for not including SD-DV in the classification analysis was that WSTS and SD-DV were highly correlated to each other and by adding it did not improve the classification accuracy. The three classifiers performed better when using both WSTS and TIS scores combined. The best performing classifier (LR) correctly classified the group with an accuracy of 85% and weighted AUC of 0.89.

**Table II. Classification results of the three classifiers when using WSTS and TIS separately and combined. The results were obtained with a stratified 10-fold cross validation.**

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (%)</th>
<th>Weighted Kappa</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input: WSTS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4.5</td>
<td>77</td>
<td>0.54</td>
<td>0.75</td>
</tr>
<tr>
<td>LR</td>
<td>74</td>
<td>0.47</td>
<td>0.83</td>
</tr>
<tr>
<td>RF</td>
<td>77</td>
<td>0.54</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Inputs: TIS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4.5</td>
<td>71</td>
<td>0.42</td>
<td>0.7</td>
</tr>
<tr>
<td>LR</td>
<td>67</td>
<td>0.34</td>
<td>0.4</td>
</tr>
<tr>
<td>RF</td>
<td>70</td>
<td>0.4</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Inputs: WSTS and TIS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4.5</td>
<td>84</td>
<td>0.68</td>
<td>0.88</td>
</tr>
<tr>
<td>LR</td>
<td>85</td>
<td>0.71</td>
<td>0.89</td>
</tr>
<tr>
<td>RF</td>
<td>82</td>
<td>0.63</td>
<td>0.89</td>
</tr>
</tbody>
</table>

The effect size correlation coefficients were as follows: TIS (0.078), SD-DV (0.011) and WSTS (0.0009), indicating that TIS has greater responsiveness than the other two scores. In Swedish LCIG-naïve patients, the mean TIS improved to the second test period on LCIG treatment and this improvement remained significant throughout the study period (Figure 2). However, the mean scores of WSTS and SD-DV deteriorated over the study period.
Figure 2. LME fixed effect coefficients of the three scores of Swedish patients over the 36 months study period. All scores for follow-up test periods i.e. from 0 to 36 were significantly different from baseline (-3) test period (each P<0.001) and those periods annotated with symbol * (P<0.005). Abbreviation: n.s. (not significant). Number of observations (test periods, months): 0 (n=4467), 3 (n=4043), 6 (n=3781), 9 (n=3391), 12 (n=3028), 18 (n=2828), 24 (n=2666), 30 (n=1833), 36 (n=1696).

Test-retest reliability coefficients of the three scores were as follows: TIS (0.9), WSTS (0.83) and SD-DV (0.55). These results indicate that the TIS score is more stable and consistent over time than WSTS and SD-DV scores.

V. DISCUSSION AND CONCLUSIONS

In conclusion, TIS could reasonably discriminate well between spiral drawings drawn by patients in different stages of PD and HE subjects, as compared to the scores produced by two previous methods i.e. SD-DV [6] and WSTS [7]. The TIS measure quantifies a different aspect of upper limb motor severity as compared to the previous scores since it was weakly correlated to both of them. In addition, the TIS was weakly correlated to simultaneous tapping speed measurements collected by the touch screen device indicating that the TIS does not measure bradykinesia. The WSTS and SD-DV are related to spatial components of the spiral and thus measure overall drawing impairment or shakiness whereas TIS is related to temporal fluctuations during spiral drawing. This difference was noted in the improvement of classification accuracy of the three classifiers when TIS was used as an input along with WSTS. The TIS measure had a better responsiveness and test-reliability than WSTS and SD-DV. These results could be related to the fact that the ApEn technique provides a high resolution of a signal by partitioning it into smaller set of windows (in our case 4 data points) and sliding the windows throughout the signal. This enables the TIS to measure high-frequency irregularities during spiral drawing in the order of milliseconds. Visual assessment of these irregularities is impossible.
In advanced Swedish patients, the mean TIS improved significantly throughout the study period (P<0.001), except for the second test period i.e. month 3 (Figure 2). This result was surprising to us since TIS was worsening with increased disease severity as seen in Figure 1. If as we assume, the TIS relates to pathological BG oscillations, this makes sense and we had expected a similar profile in Figure 2 as for WSTS where drawing impairment seem to increase over time. What we see is a gradual improvement in TIS over the first two years and then a constant level. One possible explanation for this could be that the pathological BG oscillations correlate to the number of remaining dopamine receptors. This number is related to the number of dopaminergic cells but also the number of receptors on each cell is variable. During levodopa tablet treatment there is a high variation in striatal dopamine levels that causes the number of receptors per cell to decrease in response to high peak levels. This variation in dopamine levels is reduced when starting LCIG pump treatment and over time the number of receptors per cell may begin to increase.

A limitation of the study is that there is no gold standard to validate the TIS. This limitation is related to the data collection scheme where subjects repeatedly used the touch screen device in their home environments without clinical supervision. Therefore, there were no test occasion level clinical ratings to assess the validity of TIS. An interesting research question to be investigated in the future would be to assess clinimetric properties of TIS on a new data set, which consists of sensor and clinical measurements on a test occasion level. The plan is to investigate whether mean TIS is different between PD patients and healthy controls and across different levodopa levels. Since the BG oscillations are affected by deep brain stimulation (DBS) treatment frequencies [4] another interesting research question that could be investigated in the future would be to calculate the TIS on spiral data of patients on DBS treatment. This would allow investigating the relationship between TIS and different DBS frequencies as well as during Off and On stimulation phases and whether TIS could be useful for adjusting DBS treatment.

REFERENCES


Paper IV
Verification of a method for measuring Parkinson's disease related temporal irregularity in spiral drawings

Somayeh Aghanavesi1,*, Mevludin Memedi2, Mark Dougherty1, Dag Nyholm1, Jerker Westin1

1 Computer Engineering, School of Technology and Business Studies, Dalarna University
2 Informatics, Business School, Örebro University
3 Dept. of Neuroscience, Neurology, Uppsala University
*Author to whom correspondence should be addressed: Email: saa@du.se; Tel.: +46-23 778592;

Abstract

The aim of this work is to verify and further investigate the properties of an entropy based method for measuring Parkinson's disease (PD) related upper limb temporal irregularities during spiral drawing tasks.

Nineteen PD patients and 22 healthy controls performed repeated spiral drawing tasks on a smartphone. Patients performed the tests before a single levodopa dose and at specific time intervals after the dose was given. Three movement disorder specialists rated videos of the patients while performing standardized motor tasks. A temporal irregularity score (TIS) was developed using an approximate entropy method which was used to measure the repeatability of drawing speed patterns within the drawing speed signals of the spiral drawings. Differences in mean TIS between the groups of patients and healthy subjects were assessed. Test-retest reliability of the TIS was measured. The ability of TIS to detect changes from baseline (before medication) to later time points was investigated. Correlations between TIS and clinical rating scores were assessed.

The mean TIS was significantly different (P≤0.001) between healthy subjects and patients in early, intermediate and advanced groups that were defined according to Hoehn and Yahr scale. Test-retest reliability of TIS was good (ICC=0.81). When assessing changes in relation to treatment, TIS contained some information to capture changes from Off to On, and wearing off effects from 20 to up to 300 minutes after a single dose intake. However, the correlations between TIS and clinical rating scores (UPDRS, TRS and Dyskinesia) were weak.

The results in this study confirm the reproducibility of the results obtained in the previous study since TIS was able to differentiate spiral drawings drawn by patients in different stages from those drawn by healthy subjects and TIS had good test-retest reliability. Further investigation showed that TIS was somewhat responsive to single dose levodopa treatment. Since TIS is an upper limb high frequency-based measure, the amount of temporal irregularity in spiral drawings cannot be detected during clinical assessment.

Keywords:
Parkinson's disease; motor assessment; spiral tests; temporal irregularity; timing variability; approximate entropy; smartphone; complexity; Basal Ganglia;
1. Introduction

Parkinson’s disease (PD) is a progressive, neurodegenerative disorder that is caused by degeneration of dopaminergic neurons in the substantia nigra and characterized by motor symptoms. Common treatments for PD are levodopa and dopamine agonist tablets, deep brain stimulation (DBS) and levodopa/carbidopa intestinal gel (LCIG) (Schapira, 2009). Over the course of the disease, levodopa dose and timing of intake have to be adjusted to optimize the therapeutic effect (Nyholm, et al., 2014). DBS therapy’s effectiveness is enhanced by selection of stimulation parameters (Davidson, et al., 2016).

Motor symptoms of PD are typically assessed using clinical scales such as Unified PD rating scale (UPDRS) which is used to evaluate the presence and severity of PD symptoms as well as symptom fluctuations. However, clinical scales-based measurements are not able to capture variations in symptoms continuously and they are insensitive to subtle changes. To reveal the full extent of patients’ condition and prevent recall and reporting bias, the symptoms need to be captured frequently, before and after medication (Isacson, et al., 2008b). Regularity of patterns was basically measured by regularity statistics which were centred on various entropy measures (Pincus, et al., 1991). The entropy measures, in a different context, has been an essential component of quantitative development of thermodynamics, statistical mechanics, and information theory. It’s quantifications in physics address the randomness and regularity, and the formulas themselves involve some known integrals and derivatives of functions, such as work, temperature, and energy (Feynman, 1966). In probability theory entropy is defined (Claus, 1968) with regard to the amount of uncertainty that could be calculated for two finite and independent schemes (set of events) of A and B, knowing only the probabilities of possible outcomes. This definition corresponds with intuition in that systems having more random probability distributions have greater entropy. Entropy has also been considered as a critical 'summary' statistic in nonlinear dynamical system analysis and chaos (Crutchfield, et al., 1982). However, large amount of data is required to accurately measure the entropy, and the results can be greatly influenced by system noise. In statistics, Approximate Entropy (ApEn) is a technique to quantify the regularity, unpredictability, uncertainty of fluctuations over time series data. A formula, based on the Kolmogorov and Sinai (K-S) (Latora, et al., 1999), to measure the entropy of a time series has been developed (Grassberger, et al., 1983); which have become the ‘standard’ entropy and regularity measure for use with time-series data. However, since for a stochastic process K-S entropy is often infinite while ApEn is finite, ApEn can provide useful system information to distinguish differing stochastic processes. Also the theoretic properties of ApEn has been considered to be computationally efficient and robust to noise (Pincus, et al., 1991).

There are different studies with the focus on quantifying PD regularity, variability, complexity, and unpredictability of various signals using ApEn. In a study that measured regularity, it’s found that the ankle joint movement signal of PD patients, when walked on a treadmill after levodopa medication, was more regular than before medication (Kurz, et al., 2010). The regularity measure of tremor signal was greater in groups of PD patients compared to healthy controls tremor (voluntarily generated tremor) (Morrison, et al., 2013). In the same way, wrist movement signal in PD patients had higher regularity measure than healthy controls (Powell, et al., 2014).
In a like manner, there were studies which focused on measuring variability. In one study the ApEn-based variability score was higher for power of the speed signal, for dynamic (speed referenced) PD group, compared to static (inertia load) PD group while they experimented a biking exercise (Mohammadi-Abdar, et al., 2016). Likewise, in another study that measured variability of EEG emotional signal of right-affected PD patients, the alpha and Beta emotional phases had higher variability measure than Delta, Theta, and Gamma phases (Yuvaraj, et al., 2016). To our knowledge ApEn has not been investigated for PD where patients performed spiral tests, except in our last study (Memedi, et al., 2016b).

Our previous study (Memedi, et al., 2016b) (denoted as ST1 in this paper) developed and evaluated clinimetric properties of an ApEn score for measuring a PD related temporal irregularity (TIS) using digital spiral analysis. The upper limb temporal irregularities of 98 PD patients in different stages of PD and 10 healthy subjects in an open longitudinal 36-month study were evaluated. The test retest reliability of the TIS was assessed. The long term trend of the TIS after changing the treatment to LCIG pump was researched, and the responsiveness of the TIS before and after LCIG pump treatment was investigated. The results were compared to results from two other methods (X. G. Liu, et al., 2005; Westin, et al., 2010b), both was related to spatial components of spiral drawing and used to measure overall drawing impairments. However, the correlations between TIS and clinical ratings, the sensitivity of TIS over a course of test trials, and the reproducibility of the previously reported TIS properties have not been studied.

The purpose of this study is to verify and further investigate the results from ST1 using the calculated upper limb temporal irregularity of PD patients with same methodology (ApEn) but using a new dataset, new patient groups, using another screen resolution, and during shorter term measurements. The paper reports properties of the temporal irregularity measure including the differences in TIS between patients in different stages of PD and healthy subjects and the test-retest reliability of TIS. In addition, the responsiveness of TIS to levodopa treatment during single dose experiments, and the correlations between TIS and clinical ratings are reported.

2. Methods

2.1 Participants

Nineteen PD patients and 22 healthy controls were recruited in a single center, open label, single dose clinical trial in Uppsala, Sweden (Table 1) (Senek, et al., 2017). Written informed consent was given after approval by the regional ethical review board (in Uppsala, Sweden).

<table>
<thead>
<tr>
<th>Patients</th>
<th>Gender</th>
<th>Mean age (years)</th>
<th>Hoehn &amp; Yahr stage</th>
<th>Years with the disease</th>
<th>Years on levodopa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14 males 5 females</td>
<td>71.4 (6.3)</td>
<td>3.1 (0.8)</td>
<td>10 (6.8)</td>
<td>10 (6.8)</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>16 males 6 females</td>
<td>64.2 (7.4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. Characteristics, mean (standard deviation) of patients and healthy subjects.
2.2 Data Collection

The trial included a single levodopa-carbidopa dose experiment for the PD patients, where patients were asked to perform a fine motor test (spiral drawing) using a smartphone before, and at specific time intervals. Healthy controls were asked to perform a fine motor test (spiral drawing) using a smartphone at specific time intervals. For the patients, the administered dose was 150% of their individual levodopa equivalent morning dose to follow transitions between Off, On, On with dyskinesia, and back to off motor states. Up to 15 samples per PD patient were collected, one measurement at baseline (20 minutes prior to dosing), one at the time of dose administration (at time 0) and thereafter follow-up measurements at 15, 30, 45, 60, 80, 100, 120, 150, 180, 210, 240, 300 and 360 minutes after dose administration. The healthy controls were asked to perform the tests, up to 8 times each, at time point 0 (first test) and then at 20, 40, 60, 80, 110, 140, and 170 minutes, without receiving any medication.

On each test occasion, subjects performed spiral drawings tests using a smartphone (Figure 1). The smartphone had a 4” (86 x 53 mm) touch screen with a screen resolution of 480 x 800 pixels (~233 ppi pixel density) and recorded both positions (x and y coordinates) and time-stamps (in milliseconds) of the pen tip. The device sampling was event-based which means it wasn’t possible to choose the sampling frequency and read directly from the inertial sensors. Instead a sensor event was generated every time the sensor values x and y were changed. The subjects were instructed to be seated on a chair and perform the tests using an ergonomic pen stylus with the device placed on a table and supporting neither hand nor arm. During the spiral tests, the subjects were instructed to trace a pre-drawn Archimedes spiral as fast (within 10 seconds) and accurately as possible, from the center out, using the dominant hand. The test was repeated three times per test occasion. The total number of measurements with the smartphone for PD patients was 240, and for healthy controls it was 176.

Along with smartphone-based measurements, patients were video recorded while performing standardized motor tasks according to UPDRS at corresponding time points.

![Figure 2. Implementation of a fine motor test (spiral drawing) on the smartphone.](image)
2.3 Clinical assessments of motor symptoms

The recorded videos were presented in a randomized order to three movement disorder specialists, so that the ratings were blinded with respect to time from dose administration. The specialists rated six UPDRS-part III (motor examination) items including UPDRS item #23 (finger tapping), UPDRS #25 (rapid alternating movements of hands), UPDRS #26 (leg agility), UPDRS #27 (Arising from Chair) and, UPDRS #29 (gait), UPDRS #31 (bradykinesia), according to the definitions of the motor examination part of the UPDRS (Fahn, et al., 1987). The specialists also rated dyskinesia on a severity scale from 0 to 4 (Goetz, et al., 2008) and overall mobility according to Treatment Response Scale (TRS) (Nyholm, et al., 2005), ranging from -3 (very Off) to 0 (On) to +3 (very dyskinetic). For every scale, mean scores per time point for the three specialists were calculated and used in subsequent analysis.

2.4 Feature extraction

Since the sampling rate was event based and at a higher resolution compared to the sampling rate obtained before, the data from the new study was down sampled by extracting the every second data points. Extracting every second data points yielded an effective average sampling rate comparable to ST1. The digitized spiral signals were processed using a new method based on ApEn to generate a TIS. Initially, drawing speed was calculated as a rate of spatial change with respect to time, using the following formula:

\[ DS = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i} \]  

Where \( x \) is the horizontal coordinate of pixels on the screen, \( y \) is the vertical coordinate and \( t \) is the time in seconds. The ApEn technique was then applied on DS signals to generate the TIS. Theoretically ApEn is a statistical method that measures the complexity and repeatability of patterns within a signal (Pincus, 1991). Signals which contain a single frequency component are associated with relatively small ApEn value whereas more complex signals which contain multiple frequency components are associated with high ApEn values, indicating a high level of irregularity. ApEn reflects the similarity between a chosen window of a given duration and the next set of windows of the same duration. ApEn formula is given by (Subha, et al., 2010).

\[ \text{ApEn}(m, r, N) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log_C^m(r) - \frac{1}{N-m} \sum_{i=1}^{N-m} \log_C^{m+1}(r) \]  

Where \( m \) is an integer that specifies length of the window being compared, \( r \) is a positive real number that specifies a filtering level, \( N \) is the total number of data samples and \( C(r) \) is the correlation integral. \( M \) and \( r \) must remain fixed during all calculations. Searching for the best \( m \) and \( r \) configurations across a set of commonly used values (\( m = 2, 4 \) and \( 0.1, 0.2 \)) ApEn value was computed for each possible resulting combinations. The criteria to select best \( m \) and \( r \) were based on the results for higher test-retest reliability, higher sensitivity, higher correlations between TIS and clinical ratings, and better separation of the patients groups with healthy
subjects. After experimentation, \( m \) was set to 2 and \( r \) was set to 0.2 (20% of the standard deviation of the signal). The extracted ApEn value was then corrected for total drawing completion time, hence ApEn values were inversely proportional to the degree of regularity of a time series. The resulting score was hence denoted 'temporal irregularity score' (TIS). Since there were three trials that were performed during each spiral test, the average of the three TIS was calculated and used in the following analysis.

2.5 Statistical analysis

To be able to do a comparable analysis with groups of patients in ST1, three groups of patients were defined based on the Hoehn and Yahr scale (Goetz, et al., 2004) and presented as median ± interquartile range with following criteria: Early (≤2±1), Intermediate (3±0), and Advanced (≥4±1).

Significance of the mean TIS across the four subject groups of healthy, early, intermediate and advanced patients was investigated using Linear Mixed Effect (LME) models based on a restricted maximum likelihood estimation method. To evaluate the consistency of the TIS, test-retest reliability of the TIS was assessed using the Intra correlation coefficient (ICC). The ability to detect change from baseline (before medication) to follow up time points when patients were on medication, was assessed by effect sizes. To calculate effect sizes, ANOVA models were fitted for each time point after the baseline test; first test and second test; first test and third test, and so on. A high effect size indicates that a scale is sensitive in detecting the treatment response (Goetz et al., 2013). To investigate the relation between TIS and specialists’ visual assessments the correlations between TIS and clinical rating scores were assessed by Pearson correlation coefficients.

3. Results

When assessing the differences of TIS between tests performed by patients at different PD stages and healthy controls, the mean TIS was significantly different between healthy subjects, and all other groups of Early, intermediate and advanced patients (Figure 2).
Figure 3. LME fixed effects coefficients of the TIS across the four subjects groups. Data points are mean TIS for groups, with 95% confidence intervals. P-values (groups: early, intermediate and advanced) with respect to healthy subjects were: 0.001, <0.001, <0.001. Number of participants: Healthy (n = 22), Early (n = 7), Intermediate (n = 8), Advanced (n = 4). Number of observations: Healthy (n = 176), Early (n = 93), Intermediate (n = 90), Advanced (n = 57). Test-retest reliability of TIS was 0.81 (ICC) which was similar to the results in ST1 (R = 0.74) and indicating the consistency of the three extracted TIS.

The ability of TIS to detect the change related to treatment was assessed. Previously in ST1 it was calculated for week long periods with 4 spiral drawings per day before switching to LCIG treatment, and after. TIS had a greater effect size (0.078) compared to the effect sizes of the other two methods (X. G. Liu, et al., 2005; Westin, et al., 2010b). In the current study the TIS effect sizes (Figure 3) were calculated for test periods up to 300 minutes after levodopa medication for patients and up to 170 minutes of test trials for healthy subjects. The design of the study was different to ST1 and patients performed the tests for up to 15 times during only one day. When assessing the sensitivity of TIS to the treatment response for patients, the results indicate that the patients scores are small but can capture some effects of medication that were presented in patients’ performances. Effect sizes were increased from 15 minutes after dose intake to 60 minutes, indicating the changes in symptoms from off to On/dyskinesia. Likewise from 100 minutes to 300 minutes the effect sizes were decreased, indicating wearing off effects. At time 80 the effect size was smaller than could be expected from the medication trend. For healthy subjects the effect sizes were smaller than the effect sizes of patients. At the first tests there are some variations in the effect sizes but they tend to converge around a fixed point during later tests which could be an indication of learning effect after repeated spiral drawings.
Figure 3. Sensitivity assessment of TIS across the test occasions. Lower X axis represents the minutes after taking the levodopa dose for patients and upper X axis represents the tests time points for healthy subjects. The first data point in the X axis represents the difference in scores between first (baseline) and second measurements; the second data point represents the difference in scores between first and third measurements, and so on. Number of patients for periods: 0 (n=19), 15 (n=19), 30 (n=19), 45 (n=19), 60 (n=18), 80 (n=18), 100 (n=18), 120 (n=18), 150 (n=18), 180 (n=17), 210 (n=15), 240 (n=13), 300 (n=9). Test 15 at 360 minutes contained only 1 patient that wasn’t enough to be included in this analysis. Number of healthy subjects for periods: 20 (n=22), 40 (n=22), 60 (n=22), 80 (n=22), 110 (n=22), 140 (n=22), 170 (n=22).

In STI the correlations between TIS and clinical rating scales were not available. As an additional investigation, the correlations between TIS and mean clinical rating scores (UPDRS, TRS, Dys) were weak: -0.18 for UPDRS item 23 (finger tapping), -0.11 UPDRS item 25 (rapid alternating movements of hands), -0.29 for UPDRS item 26 (leg agility), -0.24 UPDRS item 27 (arising from chair), -0.20 for UPDRS item 29 (gait), -0.10 for UPDRS item 31 (bradykinesia), -0.22 for Sum UPDRS (summation of UPDRS items), -0.06 for TRS (Treatment rating scale), and -0.31 for Dys (Dyskinesia).

4. Discussion and Conclusion

In this study we verified some clinimetric properties of TIS which was proposed and evaluated in a study (Memedi, et al., 2016b). We used an ApEn method applied on spiral drawings of 19 PD patients and 22 healthy controls. As it has been achieved in previous study, the extracted TIS score as a temporal irregularity measure in the present study was significantly different between the spiral drawings drawn by patients in early, intermediate and advanced stages and those drawn by healthy subjects. In the same way, the test-retest reliability of TIS was good. Referring to the results in the previous study, TIS was worsening with increased disease severity. This can also be observed in the results from this study where the TIS mean value was significantly smaller in advanced patients from healthy subjects (Figure 2). In addition in this study the sensitivity of TIS to the treatment response during single dose experiments from before to 300 minutes after dose intake, were calculated. The sensitivity results could not be comparable to the sensitivity results in STI due to the differences in the study designs. However, the effect sizes were related to changes in patients’ symptoms throughout the continuous tests which were performed during the day.
Moreover, in ST1 the clinical evaluation was not possible since the limitations was related to the data collection scheme where subjects repeatedly used the touch screen device in their home environments without clinical supervision and therefore there were no test occasion with clinical ratings. However in the current study the collected data were timestamped and the videos of the patients were rated by movement disorder specialists which made it possible to investigate the possible correlations between TIS and clinical scores. We found that correlations between TIS and clinical rating scores in this study were weak. A reason could be that during clinical ratings, the movement disorder specialists observed the patients while performing standardized motor tasks as defined in the UPDRS scale. The highest concern was given to the symptoms that were prominent during gross motor performance e. g., walking ability. On the other hand, during spiral drawing tasks, only fine motor movements of the dominant hands could be recorded by the smartphone touch screen. Another reason could be that TIS was related to temporal fluctuations during spiral drawing. This could be related to the fact that ApEn technique provide a high resolution of a signal by partitioning it into smaller set of windows and sliding the windows throughout the signal. This enables the TIS to measure high-frequency irregularities in spiral drawing speed in the order of milliseconds where visual assessment of these irregularities can not be detected.

TIS represents measures like timing variability, temporal irregularity and complexity of the signals. In some studies the timing variability was increased among patients when compared to healthy subjects. For instance, in (Almeida, et al., 2007) it was shown that medicated PD patients had poorer timing control than those who were withdrawn from medication when modulating timing to an external stimulus. (Jones, et al., 2011) has found that measures of motor timing accuracy can discriminate PD patients from healthy controls. On the other hand, the Basal Ganglia (BG) area play an important role in temporal processing at milliseconds to seconds range and neuronal firing rates in BG vary as a function of duration, suggesting a neurophysiological mechanism for the representation of time in the brain (Coull, et al., 2011). The speed of a hypothetical “internal clock” is controlled by BG and is related to the brain dopaminergic levels (Coull, et al., 2011). Correlation has been established between some of this pathological activity and the motor symptoms of PD (Davidson, et al., 2016). The observations from BG and its relation to motor movements have encouraged investigations of the functional role of abnormal oscillatory activity in BG circuits in the symptomology of movement disorders (Walters., 2017). The quantitative measures of upper limb temporal irregularities indicating these oscillations would be beneficial in facilitating the assessment of high frequency-based motor irregularities that could be difficult to be assessed visually.

The results in this study as well as in ST1 showed that TIS was more related to disease severity than the clinical ratings which indicates it might be more useful in long term diagnostic tools than for detecting the treatment response. It also introduces a possibility that TIS might be related to Basal ganglia oscillations. However, to specifically be able to demonstrate the possible correlations between TIS and pathological oscillations there is a need for further investigations. DaT SPECT (Scherfler, et al., 2007) as a tool to determine the degeneration of the presynaptic neurons in striatum of patients with PD could be used to investigate whether there is a relation between number of dopamine receptors and TIS in future studies. In addition the results in this study suggest applying the current methodology to a larger dataset including
more subjects at various stages of medication, and subjects at various years of treatment. Optimization of the ApEn algorithm parameters may help to attain results more efficiently. The investigation of TIS properties on spiral data of patients with DBS therapy would also be of interest, since the BG oscillations are affected by DBS treatment frequencies (Da Cunha, et al., 2015).

Acknowledgments

This work is performed in the framework of the FLOAT and MUSYQ projects, funded by Swedish knowledge foundation and Swedish innovation agency (Vinnova), in collaboration with Cenvigo AB, Sensidose AB, Swedish ICT Acreo, Uppsala University, Sahlgrenska University, and Dalarna University.

Conflict of interest

None.

References


