



<http://www.diva-portal.org>

This is the published version of a paper published in *Informatics in Medicine Unlocked*.

Citation for the original published paper (version of record):

Aghanavesi, S., Nyholm, D., Marina, S., Bergquist, F., Memedi, M. (2017)
A smartphone-based system to quantify dexterity in Parkinson's disease patients.
Informatics in Medicine Unlocked, 9: 11-17
<https://doi.org/10.1016/j.imu.2017.05.005>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:du-25034>



A smartphone-based system to quantify dexterity in Parkinson's disease patients



Somayeh Aghanavesi ^a, Dag Nyholm ^b, Marina Senek ^b, Filip Bergquist ^c, Mevludin Memedi ^{a, d, *}

^a Computer Engineering, School of Technology and Business Studies, Dalarna University, Sweden

^b Dept. of Neuroscience, Neurology, Uppsala University, Sweden

^c Dept. of Pharmacology, University of Gothenburg, Sweden

^d Informatics, Business School, Örebro University, Sweden

ARTICLE INFO

Keywords:

Parkinson's disease
Motor assessment
Spiral tests
Tapping tests
Smartphone
Dyskinesia
Bradykinesia
Objective measures
Telemedicine

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.

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder [36] 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 [33]. 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-

* Corresponding author. Informatics, Business School, Örebro University, Örebro, Sweden.
E-mail address: mevludin.memedi@oru.se (M. Memedi).

<http://dx.doi.org/10.1016/j.imu.2017.05.005>

Received 14 March 2017; Received in revised form 28 April 2017; Accepted 9 May 2017

Available online 13 May 2017

2352-9148/© 2017 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

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) [22], 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 [16]. 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 [20]. 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) [1,7]. 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 [4].

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 [11,43]. There are different studies with the focus on quantifying various motor symptoms. Some have focused on assessing motor dysfunctions in upper extremities [13,40,41], some on gross motor symptoms like gait [21], while others on combination of both. For instance [39], 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 Ref. [31] to quantify bradykinesia and tremor. Other studies have focused on analyzing data from upper limbs during standardized tasks like finger tapping [13,35], digital spiral analysis [32] and quantitative digitography [10,38].

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 [12,17,32]. 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 [12], dyskinesia [17], drawing impairments [40,41] and global tapping performance [24]. Quantitative measures during alternating tapping tests and digital spiral analysis have been previously used as measures of bradykinesia [10] and severity of PD symptoms [32]. 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 advanced PD patients and 22 healthy controls were recruited in a single center, open label, single dose clinical trial in Uppsala, Sweden (Table 1, [34]). Written informed consent was given after approval by the regional ethical review board (in Uppsala, Sweden).

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 [34,40,41]. 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 min prior to dosing), one at the time of dose administration (0 min) and thereafter follow-up measurements at 20, 40, 80, 110, 140, 170, 200, 230, 260, 290, 320, and 360 min 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 min, without receiving any medication.

On each test occasion, subjects performed upper limb motor tests (tapping and spiral drawings), using a smartphone (Fig. 1). The smartphone had a 4" (86 × 53 mm) touch screen with a 480 × 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 s. During the spiral tests, the subjects were instructed to trace a pre-drawn Archimedes spiral as fast (within 10 s) 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 [6]. 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 [8] and overall mobility according to Treatment Response Scale (TRS) [28], 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. 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

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

	Gender	Mean age (years)	Mean height (m)	Mean weight (kg)	Years with PD	Years on levodopa	Hoehn & Yahr
Patients	14 males, 5 females	71.4 (6.3)	1.75 (0.09)	75.4 (11)	9.7 (6.8)	9.5 (6.5)	3.16 (0.9)
Healthy controls	16 males, 6 females	64.2 (7.4)	1.75 (0.1)	83.6 (13.8)	–	–	–

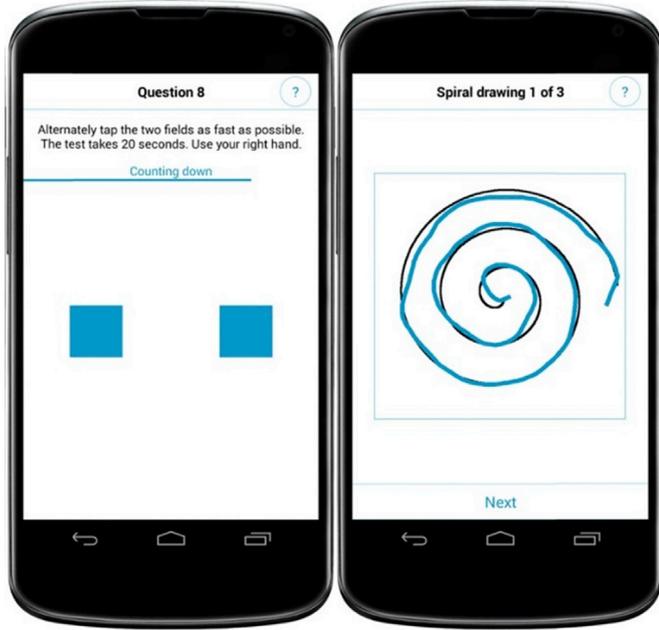


Fig. 1. Implementation of dexterity tests (tapping and spiral drawing) on the smartphone.

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.

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:

$$K = \frac{E(s - \mu)^4}{\sigma^4} \quad (1)$$

where s is the distribution of the drawing speed per second, μ is the mean of s , σ 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, abruptness 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 [15,27]. The spiral drawing curve length is

Table 2
Spatiotemporal features calculated from tapping and spiral data.

Tapping features (Reference to previous works)

- (1) Total Number of taps [19]
- (2) Mean tapping time difference between two fields [42]
- (3) Mean tapping speed from right to left [24]
- (4) Mean tapping speed from left to right [24]
- (5) Coefficient of variation of tapping speed from right to left,
- (6) Coefficient of variation of tapping speed from left to right [38]
- (7) Mean distance from the centers of the fields [2]
- (8) Coefficient of variation of distances from the center fields [2]
- (9) Overall distribution of the taps [24]
- (10) Mean distance from center [2]
- (11) Mean tapping speed per cycle [24]
- (12) The absolute mean difference between the first and second part of the time series signal [24]
- (13) Approximate entropy (ApEn) measure of the two parts of the time series signal [24]
- (14) Measure of tapping speed reduction [2]
- (15) Measure of tapping reaction time [42]
- (16) Overall trend of tapping reaction time during the test trial [42]
- (17) ApEn measure of mean tapping speed [24]
- (18) The amount of irregularity in vertical tap distance [24]
- (19) Measure of variation in distance between the two fields [2,24]
- (20) Measure of irregularity caused by time variations during tapping [25,26]

Spiral features (Reference to previous works)

- (21) Mean drawing speed [37]
- (22) Coefficient of variation of speed [25,26]
- (23) Skewness of the speed [25,26]
- (24) Radial velocity [14]
- (25) Mean time difference [3]
- (26) Global minima of the drawing speed,
- (27) Global maxima of the drawing speed [25,26]
- (28) ApEn measure of drawing speed,
- (29) ApEn measure of radial velocity [23]
- (30) Coefficient of variation of high frequency wavelet coefficients by discrete wavelet transform (DWT) [25,26]

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 $b - \frac{a}{2}$. There is a relation between the size of the spiral and the speed of the drawing movements. According to [18], 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 [5]: 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 [9]. 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

Table 3

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

Feature#	Rater 1	Rater 2	Rater 3
1	X	X	X
3	X	X	X
7	X		
9	X		
10			X
13		X	X
16	X	X	X
19	X		
20	X		X
21		X	
23	X		
24	X	X	X
28		X	X
31	X		
35	X	X	X
36	X		X
37	X	X	X

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.

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 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.

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.

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 (Fig. 2).

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.

	SVM	LR	RT	MLP
UPDRS #23	0.52	0.13	0.33	0.13
UPDRS #25	0.47	0.13	0.13	0.19
UPDRS #31	0.57	0.27	0.24	0.26
sumUPDRS	0.46	0.15	0.11	0.14
TRS	0.59	0.39	0.39	0.30
Dyskinesia	0.64	0.56	0.57	0.53

Table 5

ICCs between the first two baseline measurements for smartphone-based scores and mean ratings of the three clinicians.

	Clinical scores	SVM
UPDRS #23	0.74	0.17
UPDRS #25	0.62	0.16
UPDRS #31	0.87	0.23
sumUPDRS	0.91	0.46
TRS	0.94	0.51
Dyskinesia	1	0.84

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.

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

Table 6

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

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
P-value	<0.001	<0.001	0.9	<0.001	0.1	0.9	0.1
Healthy controls							
Mean	1.27	-0.78	-0.00	0.62	0.12	-0.00	-0.08
SD	2.84	1.97	1.97	1.55	1.53	1.06	1.23
Patients							
Mean	-0.98	0.60	0.00	-0.48	-0.09	0.00	0.06
SD	2.73	2.45	2.02	1.95	1.62	1.40	1.11

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

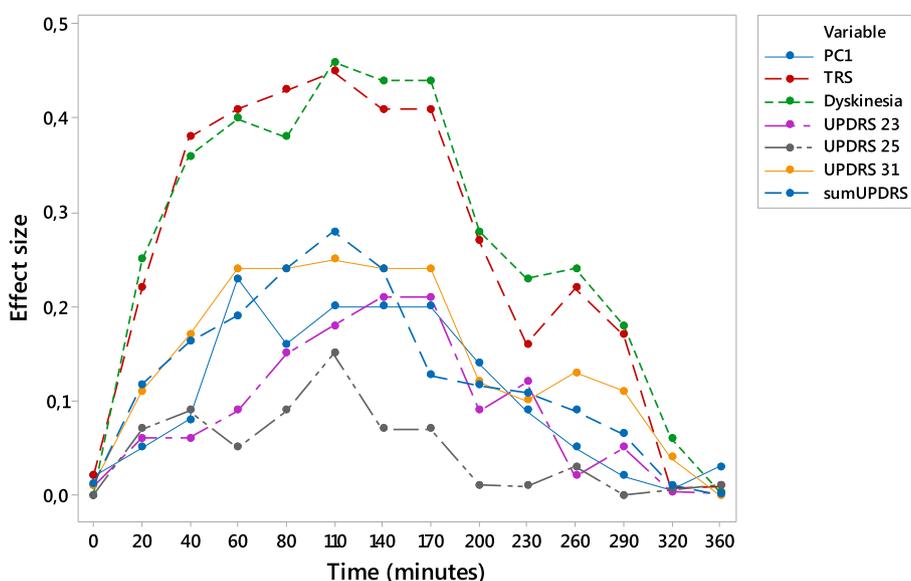


Fig. 2. Sensitivity assessment of PC1 and mean ratings of the three movement disorder specialists on the three UPDRS items, TRS and dyskinesia across the levodopa test cycle for all patients. The first data point in the X axis represents the change in scores between the first two baseline (without medication) 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).

in a longitudinal 36 months clinical trial in Sweden and a two weeks trial in Italy [24,25]. 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 [30]. 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 [2]. 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 min post-dose, Fig. 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 Ref. [13] 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 [29]. 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.

Conflict of interest

None.

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.

References

- [1] Andong Zhan MAL, Harris Denzil A, Abiola Solomon O, Ray Dorsey E, Saria Suchi, Terzis Andreas. High frequency remote monitoring of Parkinson's disease via smartphone: platform overview and medication response detection. arXiv Prepr 2016. arXiv:1601.00960.
- [2] Arora S, Venkataraman V, Zhan A, Donohue S, Biglan KM, Dorsey ER, et al. Detecting and monitoring the symptoms of Parkinson's disease using smartphones: a pilot study. *Park Relat Disord* 2015;21(6):650–3. <http://dx.doi.org/10.1016/j.parkreldis.2015.02.026>.
- [3] Banaszekiewicz K, Rudzinska M, Bukowczan S, Izworski A, Szczudlik A. Spiral drawing time as a measure of bradykinesia. *Neurol Neurochir Pol* 2009;43(1):16–21.
- [4] Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study: Parkinson disease mobile data collected using ResearchKit. *Sci Data* 2016;3:160011. <http://dx.doi.org/10.1038/sdata.2016.11>.
- [5] Eibe Frank A, Hall M, Ian H. The WEKA workbench. Online appendix for "data mining: practical machine learning tools and techniques. fourth ed. Morgan Kaufmann; 2016.
- [6] Fahn S, Elton RL. Unified Parkinson's disease rating scale—recent developments in Parkinson's disease, vol. 2. Florham Park, NJ.: Macmillan Health Care Information; 1987 (15), 3-163, 293–304.
- [7] Galli M, Vimercati SL, Manetti E, Cimolin V, Albertini G, De Pandis MF. Spiral analysis in subjects with Parkinson's disease before and after levodopa treatment: a new protocol with stereophotogrammetric systems. *J Appl Biomater Funct Mater* 2014;12(2):e107–112. <http://dx.doi.org/10.5301/JABFM.2012.9265>.
- [8] Goetz CG, Nutt JG, Stebbins GT. The unified dyskinesia rating scale: presentation and clinimetric profile. *Mov Disord* 2008;23(16):2398–403.
- [9] Goetz CG, Stebbins GT, Chung KA, Hauser RA, Miyasaki JM, Nicholas AP, et al. Which dyskinesia scale best detects treatment response? *Mov Disord* 2013;28(3):341–6.
- [10] Giovannoni G, van Schalkwyk J, Fritz VU, Lees AJ. Bradykinesia akinesia inco-ordination test (BRAIN TEST): an objective computerised assessment of upper limb motor function. *J Neurol Neurosurg Psychiatry* 1999;67(5):624–9.
- [11] Girffiths RI, Kotschet K, Arfon S, Xu ZM, et al. Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *J Parkinson's Dis* 2012;2(1):47–55.
- [12] Haubenberger D, Kalowitz D, Nahab FB, Toro C, Ippolito D, Luckenbaugh DA, et al. Validation of digital spiral analysis as outcome parameter for clinical trials in essential tremor. *Mov Disord* 2011;26(11):2073–80. <http://dx.doi.org/10.1002/mds.23808>.
- [13] Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP. Clinician versus machine: reliability and responsiveness of motor endpoints in Parkinson's disease. *Park Relat Disord* 2014;20(6):590–5.
- [14] Hess CW, Hsu AW, Yu Q, Ortega R, Pullman SL. Increased variability in spiral drawing in patients with functional (psychogenic) tremor. *Hum Mov Sci* 2014;38:15–22. <http://dx.doi.org/10.1016/j.humov.2014.08.007>.
- [15] Hoogendam YY, van der Lijn F, Vernooij MW, Hofman A, Niessen WJ, van der Lugt A, et al. Older age relates to worsening of fine motor skills: a population-based study of middle-aged and elderly persons. *Front Aging Neurosci* 2014;6:259. <http://dx.doi.org/10.3389/fnagi.2014.00259>.
- [16] Isacson D, Bingefors K, Kristiansen IS, Nyholm D. Fluctuating functions related to quality of life in advanced Parkinson disease: effects of duodenal levodopa infusion. *Acta Neurol Scand* 2008;118(6):379–86. <http://dx.doi.org/10.1111/j.1600-0404.2008.01049.x>.
- [17] Liu X, Carroll CB, Wang SY, Zajicek J, Bain PG. Quantifying drug-induced dyskinesias in the arms using digitised spiral-drawing tasks. *J Neurosci Methods* 2005;144(1):47–52. <http://dx.doi.org/10.1016/j.jneumeth.2004.10.005>.
- [18] Longstaff MG, Mahant PR, Stacy MA, Van Gemmert AW, Leis BC, et al. Discrete and dynamic scaling of the size of continuous graphic movements of parkinsonian patients and elderly controls. *J Neurol Neurosurg Psychiatry* 2003;74(3):299–304.
- [19] Macleod AD, Counsell CE. Timed tests of motor function in Parkinson's disease. *Park Relat Disord* 2010;16(7):442–6. <http://dx.doi.org/10.1016/j.parkreldis.2010.04.006>.
- [20] Maetzler W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. *Mov Disord* 2013;28(12):1628–37. <http://dx.doi.org/10.1002/mds.25628>.
- [21] Mariani B, Hoskovec C, Rochat S, Bula C, Penders J, Aminian K. 3D gait assessment in young and elderly subjects using foot-worn inertial sensors. *J Biomech* 2010;43(15):2999–3006.
- [22] Martínez-Martín P, Gil-Nagel A, Gracia LM, Gómez JB, Martínez-Sarriés J, Bermejo F. Unified Parkinson's disease rating scale characteristics and structure. *Mov Disord* 1994;9(1):76–83. <http://dx.doi.org/10.1002/mds.870090112>.
- [23] Medemi, M., Aghanavesi, S., & Westin, J. (vol. 2016, 24–27 Feb. 2016). A method for measuring Parkinson's disease related temporal irregularity in spiral drawings.

- Paper presented at the 2016 IEEE-EMBS international conference on biomedical and health informatics (BHI).
- [24] Memedi M, Khan T, Grenholm P, Nyholm D, Westin J. Automatic and objective assessment of alternating tapping performance in Parkinson's disease. *Sensors (Basel)* 2013;13(12):16965–84. <http://dx.doi.org/10.3390/s131216965>.
- [25] Memedi M, Sadikov A, Groznic V, Zabkar J, Mozina M, Bergquist F, et al. Automatic spiral analysis for objective assessment of motor symptoms in Parkinson's disease. *Sensors (Basel)* 2015a;15(9):23727–44. <http://dx.doi.org/10.3390/s150923727>.
- [26] Memedi M, Nyholm D, Johansson A, Pålhagen S, Willows T, Widner H, et al. Validity and responsiveness of at-home touch screen assessments in advanced Parkinson's disease. *IEEE J Biomed Health Inf* 2015b;19:1829–34.
- [27] Miralles F, Tarongi S, Espino A. Quantification of the drawing of an Archimedes spiral through the analysis of its digitized picture. *J Neurosci Methods* 2006; 152(1–2):18–31. <http://dx.doi.org/10.1016/j.jneumeth.2005.08.007>.
- [28] Nyholm D, Nilsson Remahl AI, Dizdar N, Constantinescu R, Holmberg B, et al. Duodenal levodopa infusion monotherapy vs oral polypharmacy in advanced Parkinson's disease. *Neurology* 2005;64(2):216–23.
- [29] Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Mov Disord* 2005; 12:1577–84.
- [30] Sadikov A, Groznic V, Mozina M, Zabkar J, Nyholm D, Memedi M, et al. Feasibility of spirography features for objective assessment of motor function in Parkinson's disease. *Artif Intell Med* 2017. In Press.
- [31] Salarian A, Russmann H, Wider C, Burkhard PR, Vingerhoets FJ, Aminian K. Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. *IEEE Trans Biomed Eng* 2007;54(2):313–22. <http://dx.doi.org/10.1109/TBME.2006.886670>.
- [32] Saunders-Pullman R, Derby C, Stanley K, Floyd A, Bressman S, Lipton RB, et al. Validity of spiral analysis in early Parkinson's disease. *Mov Disord* 2008;23(4): 531–7. <http://dx.doi.org/10.1002/mds.21874>.
- [33] Schapira AH. Neurobiology and treatment of Parkinson's disease. *Trends Pharmacol Sci* 2009;30(1):41–7. <http://dx.doi.org/10.1016/j.tips.2008.10.005>.
- [34] Senek M, Aquilonius S, Askmark H, Bergquist F, Constantinescu R, Ericsson A, et al. Levodopa/carbidopa microtablets in Parkinson's disease: a study of pharmacokinetics and blinded motor assessment. *Eur J Clin Pharmacol* 2017;73: 563–71.
- [35] Shima K, Tamura Y, Tsuji T, Kandori A, Yokoe M, Sakoda S. Estimation of human finger tapping forces based on a fingerpad-stiffness model. In: Conference proceedings : annual international conference of the IEEE engineering in medicine and biology society IEEE engineering in medicine and biology society annual conference; 2009. p. 2663–7. 2009.
- [36] Shin JY, Habermann B. Medication adherence in people with Parkinson disease. *J Neurosci Nurs* 2016. <http://dx.doi.org/10.1097/JNN.0000000000000198>.
- [37] Stanley K, Hagenah J, Bruggemann N, Reetz K, Severt L, Klein C, et al. Digitized spiral analysis is a promising early motor marker for Parkinson disease. *Park Relat Disord* 2010;16(3):233–4.
- [38] Taylor Tavares AL, Jefferis GS, Koop M, Hill BC, Hastie T, Heit G, et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Mov Disord* 2005;20(10):1286–98. <http://dx.doi.org/10.1002/mds.20556>.
- [39] Tsiouras MG, Tzallas AT, Rigas G, Tsouli S, Fotiadis DI, Konitsiotis S. An automated methodology for levodopa-induced dyskinesia: assessment based on gyroscope and accelerometer signals. *Artif Intell Med* 2012;55(2):127–35.
- [40] Westin J, Ghiamati S, Memedi M, Nyholm D, Johansson A, Dougherty M, et al. A new computer method for assessing drawing impairment in Parkinson's disease. *J Neurosci Methods* 2010a;190(1):143–8. <http://dx.doi.org/10.1016/j.jneumeth.2010.04.027>.
- [41] Westin J, Dougherty M, Nyholm D, Groth T. A home environment test battery for status assessment in patients with advanced Parkinson's disease. *Comput Methods Programs Biomed* 2010b;98:27–35.
- [42] Yahalom G, Simon ES, Thorne R, Peretz C, Giladi N. Hand rhythmic tapping and timing in Parkinson's disease. *Park Relat Disord* 2004;10(3):143–8. <http://dx.doi.org/10.1016/j.parkreldis.2003.10.001>.
- [43] Zeng W, Liu F, Wang Q, Wang Y, Ma L, Zhang Y. Parkinson's disease classification using gait analysis via deterministic learning. *Neurosci Lett* 2016;633:268–78.