Graphical Tasks to Measure Upper Limb Function in Patients with Parkinson's Disease: Validity and Response to Dopaminergic Medication

Esther J. Smits[1], Antti J. Tolonen[2], Luc Cluitmans[2], Mark van Gils[2], Rutger C. Zietsma[3], Robbert W.K. Borgemeester[1], Teus van Laar[1], Natasha M. Maurits[1]

***Abstract*—The most widely used method to assess motor functioning in Parkinson's Disease (PD) patients is the Unified Parkinson's Disease Rating Scale-III (UPDRS-III). The UPDRS-III has limited ability to detect subtle changes in motor symptoms. Alternatively, graphical tasks can be used to provide objective measures of upper limb motor dysfunction. The present study investigated the validity of such graphical tasks to assess upper limb function in PD patients and their ability to detect subtle changes in performance. Fourteen PD patients performed graphical tasks before and after taking dopaminergic medication. Graphical tasks included figure tracing, writing and a modified Fitts' task. The Purdue pegboard test was performed to validate these graphical tasks. Movement time (MT), writing size and the presence of tremor were assessed. Movement time (MT) on the graphical tasks correlated strongly with performance on the Purdue pegboard test (Spearman's rho > 0.65; p < 0.05). MT decreased significantly after the intake of dopaminergic medication. Tremor power decreased after taking dopaminergic medication in most PD patients who suffered from tremor. Writing size did not correlate with performance on the Purdue pegboard test, nor did it change after taking medication. Our set of graphical tasks is valid to assess upper limb function in PD patients. MT proved to be the most useful measure for this purpose. The response on dopaminergic medication was optimally reflected by an improved MT on the graphical tasks in combination with a decreased tremor power, whereas writing size did not respond to dopaminergic treatment.**

***Index Terms*—Parkinson's disease, graphical tasks, handwriting, drawing, upper limb function, validity, bradykinesia, tremor, micrographia.**

Manuscript submitted for review June 11, 2015. This study is part of the DiPAR project for which funding was received from the EC in the FP7-SME-201001 programme (grant agreement 262291). The hardware of the measurement setup was provided by Fraunhofer IPMS (Dresden, Germany) and the data acquisition software by Fraunhofer IPA (Stuttgart, Germany), based on a concept by Manus Neurodynamica Ltd (patent WO/2011/141734). We would like to thank H. Giezen and M. Folkerts, Master students of Human Movement Sciences at the University of Groningen, the Netherlands, who performed most of the measurements. [1]University of Groningen, University Medical Center Groningen, Department of Neurology, AB51 P.O. Box 30001, 9700 RB Groningen, the Netherlands (e.j.smits@umcg.nl; r.w.k.borgemeester@umcg.nl; t.van.laar@umcg.nl; n.m.maurits@umcg.nl)

[2]VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Tampere, Finland (antti.tolonen@vtt.fi; luc.cluitmans@vtt.fi; mark.vangils@vtt.fi)

[3]Manus Neurodynamica Ltd 4a John Street, Cullercoats, North Shields Tyne and Wear, NE30 4PL, United Kingdom (r.c.zietsma@manusneuro.com)

I. Introduction

CORRECT diagnosis of Parkinson's disease (PD), a neurodegenerative movement disorder, is essential for optimal treatment and prognosis. Diagnosis is based upon the cardinal motor symptoms of PD: bradykinesia (slow movement), rigidity (muscle stiffness), resting tremor (trembling of a body part in rest) and postural and gait impairment [1]. The motor part (part III) of the most recent revision of the Unified Parkinson's Disease Rating Scale by the Movement Disorder Society (MDS-UPDRS-III) is the most widely used scale in clinical practice to assess parkinsonian symptoms [1], [2]. However, the MDS-UPDRS-III has several limitations. Firstly, the MDS-UPDRS-III needs to be evaluated by a trained assessor, which makes it less suitable for home-based monitoring. Secondly, the inter-rater reliability is high for movement disorders specialists, but unfortunately not all patients have access to specialized movement disorder centers, which negatively influences the accuracy and reliability of the MDS-UPDRS-III. Thirdly, the MDS-UPDRS-III has limited ability to detect subtle changes in motor function, which are common in early PD. Therefore, an objective, sensitive, reliable and accurate assessment system could potentially overcome these limitations of the existing MDS-UPDRS-III scale.

Objective motor assessment systems have been developed previously, such as Kinesia (Great Lakes NeuroTechnologies, USA) [3], [4] and Parkinson's KinetiGraph (PKG, Global Kinetics Corporation, Australia). These systems involve movement sensors that need to be worn on the index finger (Kinesia) or wrist (PKG) and are used to assess and monitor movement of the upper limb. Another tool which might be useful for monitoring employs a digital tablet and pen and can be used to perform and record graphical tasks, i.e. handwriting and drawing [5]. Similar to the Kinesia and PKG systems, the pen and tablet tool is non-invasive, portable, and can be used easily at home without an examiner. An advantage of Kinesia and a digital tablet compared to PKG is that specific short-term tasks can be performed several times a day and that patients don't have to wear a sensor for multiple days. An advantage of a digital tablet when compared to Kinesia is that a clinician can obtain feedback on how tasks were performed, because pen-tip movements are recorded. On the contrary, if a patient performs a task while wearing the Kinesia-sensor, like

holding the arms in a specific posture, there is no way to verify whether the task was performed correctly.

Graphical tasks employing a digital tablet provide objective measures of important motor symptoms of PD and were used previously to show differences between PD patients and healthy control (HC) participants [6]–[10]. Medication effects have also been investigated in PD patients using such a set-up [11]–[15]. In general, these studies investigated one graphical task. However, to assess and monitor PD, it is important to measure several aspects of motor behavior, because PD patients do not always suffer from the same combination of symptoms, and treatment could have variable effects on different aspects of motor functioning of PD patients. Therefore, the assessment battery of the current study consists of several graphical tasks, providing useful measures for bradykinesia, tremor and micrographia [6]. A newly developed system consisting of a pen and tablet and custom software, based on a concept by Manus Neurodynamica Ltd, was used. The advantage of this system is that it includes an integrated comprehensive sensor and data acquisition system for highly accurate recordings and analysis. The aim of the current study was to investigate the validity of these graphical tasks to assess upper limb function in PD patients. Validity of comparable methods in the literature is mostly determined by their correlation with the total UPDRS-III [11], [16]–[19]. However, our graphical tasks only involve movements of the upper limb, and since the MDS-UPDRS-III score involves more than just upper limb functioning, it is less suitable for validation. Therefore, a validated test for upper limb functioning [17], [20], [21], the Purdue pegboard test (PPT), was used as reference, in addition to MDS-UPDRS-III scores, involving hand items only.

Graphical tasks in this study include simple tracing and writing tasks and a modified Fitts' task, which are easy to perform and cover a large range of upper limb functions. A resting task was included to measure resting tremor, because PD tremor is typically a resting tremor [22]. Furthermore the modified Fitts' task was used to assess the speed-accuracy trade off, which may be impaired in PD patients [23]. This exploratory study investigated which of the tasks could be used most optimally to assess upper limb motor functioning and to detect changes in motor performance after use of dopaminergic medication in PD patients.

II. Methods

A. Participants

Fourteen PD patients performed the tasks with their right hand. PD patients were diagnosed by a movement disorders specialist (according to the UK Parkinson's Disease Society Brain Bank criteria [24]) and were treated at the movement disorders clinic of the University Medical Center Groningen (UMCG). Since the patients had to be able to hold a pen for at least 30 minutes and perform tracing and writing tasks, PD patients in relatively early stages of the disease (Hoehn and Yahr stages 1-2 [25]) were selected. Table 1 provides a summary of the PD patient characteristics. Patients agreed with overnight withdrawal of their usual dose of dopaminergic medication. Exclusion criteria were a neurological or motor disorder other than PD and a low score (< 26) on the Mini Mental State Examination (MMSE) to ensure understanding of task instructions. All patients signed informed consent and the protocol was approved by the Medical Ethical Committee of the UMCG.

B. Experimental design

Patients were seated in front of a table in a comfortable position to write. A tablet computer (ASUS Eee Slate EP121) and a newly developed digital pen with custom software were used. Position of the pen-tip on the tablet and gyroscope signals in three directions (pitch, yaw and roll) was recorded. All recordings had a sampling frequency of 200 Hz. The pen had a wireless connection to the tablet. Patients performed seven tasks (see below). The examiner was seated behind an operator computer to start and stop recordings and determined whether patients executed the tasks correctly. If a task was executed incorrectly, the recording was stopped and restarted after re-instruction. The OFF measurement was performed in the morning, after overnight withdrawal of medication. Thereafter, patients took their medication and had a one-hour break, to allow for an optimal medication effect. After one hour the ON measurement was performed. The hand items of the MDS-UPDRS-III were assessed and videotaped and scored by a clinician from the UMCG (RWKB), who was blinded to the medication status of the patients.

C. Tasks

Patients were instructed to start each task after a signal of the examiner. Firstly, a recording at rest (30 seconds) was performed. Patients were seated with the right elbow resting on the table, the hand resting on the tablet and the pen-tip touching a target (filled circle, 0.7 cm in diameter) in the center of the tablet. Next, the patients subsequently traced circle, spiral and zigzag figures which were displayed on the tablet (see Figure 1). The circle and spiral were traced ten times in a clockwise direction, starting from the 12 o'clock position (circle) or from inside to outside (spiral). The zigzag was traced five times, from left to right and back.

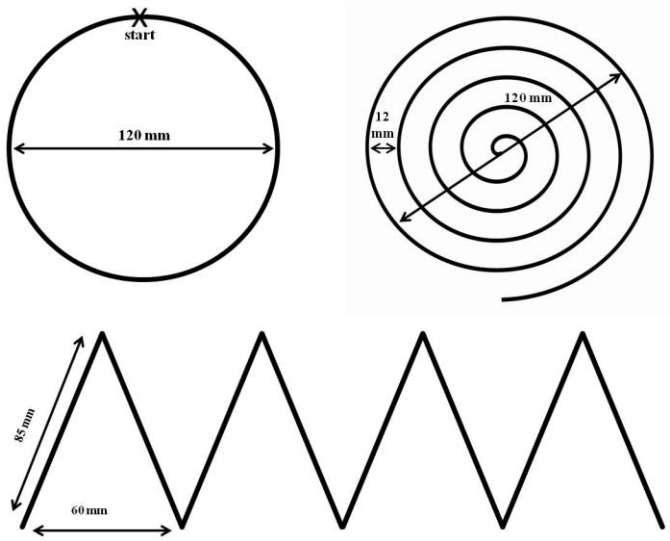


Figure.1. Templates and their dimensions for the tracing tasks: a circle, spiral and zigzag figure.

TABLE I
SUMMARY OF PD PATIENT CHARACTERISTICS

							MDS-UPDRS-III				Purdue pegboard	
					Years since		Bradykinesia subscore		Tremor subscore		Right hand	
ID	Gender	Age	MMSE	D	diagnosis	H&Y	OFF	ON	OFF	ON	OFF	ON
PD001	M	73	26	PD	3	2	1[a]	0[b]	0[c]	0[c]	9	9
PD002	M	78	29	PD	4	1	0[a]	0	0[c]	0	11	11
PD003	M	60	28	PD	2	2	6	0	0[c]	0	7	8
PD004	M	81	29	PD	0	2	0	1	0	0	8	8
PD005	F	70	29	PD	3	2	1	0	0[d]	0	12	13
PD006	M	69	28	PD	10	2	1	1	0	0	8	4
PD007	F	76	30	PD	1	2	0	0	0	0	8	9
PD008	F	68	26	PD	1	2	3	**	0	**	7	7
PD009	M	69	30	PD	10	2	2	0	0	0	7	8
PD010	M	67	29	PD	6	1	0	0	0	0	11	11
PD011	F	50	28	PD	8	2	0	0	2	0	12	13
PD012	M	66	30	tdPD	5	2	0	1	0	0[d]	7	11
PD013	M	73	27	tdPD	7	2	0	2	3	3	8	7
PD014	M	61	26	tdPD	5	2	2	3	0	0	10	11

MMSE = Mini Mental State Examination; D = Diagnosis; PD = Parkinson's disease; tdPD = Tremor dominant Parkinson's Disease; H&Y = Hoehn and Yahr score; OFF = off medication, after overnight withdrawal of anti-parkinsonian medication; ON = on medication, 1 hour after taking medication; Med = Medication
MDS-UPDRS-III = the motor part (part III) of the Movement Disorder Society revision of the Unified Parkinson's Disease Rating Scale
Bradykinesia subscore = sum of right hand items on 'finger tapping', 'hand movements', and 'pronation-supination movement of the hand'
Tremor subscore = sum of right hand tremor-items
Purdue pegboard Right hand score = number of pins inserted with the right hand in 30 seconds.
a. For these patients the 'Hand movements' score of the right hand was missing
b. For this patients the 'Finger tapping' score of the right hand was missing
c. For these patients the 'Kinetic tremor' scores for the right hand were missing
d. For these patients the 'Postural tremor' score of the right hand was missing
** For this patient too many items were missing

The next task consisted of writing 'elelelel' five times with each phrase starting at the left side of the tablet with visual feedback on the screen. An example was provided on paper on the table above the tablet. Thereafter, a modified Fitts' task was performed, which was similar to Fitts' original task [26] but adapted to the dimensions of our system. Patients touched two targets (filled circles, placed on an imaginary horizontal line in the middle of the tablet) alternately with the pen-tip as fast and as accurately as possible (20 seconds). The distance between the targets was 7 cm for subtasks 1 to 4 and 20 cm for subtasks 5 to 8, while the diameter of the targets increased (0.7, 1.3, 1.9, 2.5 cm). Finally, the PPT was performed, which employed a board with a vertically oriented row with 25 holes and metal pins located in a reservoir at the top. We limited the PPT to the right hand task, similar to the other graphical tasks. Patients were instructed to place as many pins as possible in the holes within 30 seconds. Participants were allowed to practice before the test [27].

D. Data analysis

Graphical tasks were analyzed using Matlab (R2014A). Mean movement time (MT) per trial was calculated for the tracing tasks (CircleMT; SpiralMT and ZigzagMT), using the x and y coordinates of the pen-tip.

Elelelel writing task

The pen-tip position data (x and y coordinates) were preprocessed. First, the data were split into separate segments, where each segment represented one line of text. This was done using an 'in range' signal, which indicates whether or not the pen is in detection range of the tablet employing that, after writing one line of text, the patient lifts the pen so that it is outside the detection range of the tablet. Subsequently, the segments corresponding to an 'e' or an 'l' were identified. The shapes in each line were recognized by using a state machine that employs the direction of change of the pen-tip position as input (similar to the method used in [6], see Supplementary Methods 1). For each detected letter 'e' and 'l', movement time (MT) was calculated by counting the samples and dividing it by 200 (since the sample frequency was 200 Hz). MT was averaged over all detected 'e's' and 'l's', resulting in the features E_MT and L_MT. The mean width and height of the letters was also calculated, using the x and y coordinates of the pen-tip (E_Width; E_Height; L_Width; L_Height).

Modified Fitts task analysis

The modified Fitts' task was analyzed according to Fitts' law [26]. The tradeoff between speed and accuracy was modeled by Fitts [26] in the time required for movement (T):

$$T = a + b\left(\log\frac{2A}{D}\right)$$

Here, A is the distance between targets and D the target diameter. The part *log(2A/D)* is known as the index of difficulty (ID). When multiple IDs are available, *a* and *b* can be estimated by linear regression. In our modified Fitts' task eight IDs could be determined, since the task consists of eight subtasks, with varying difficulty. For each patient the mean T for each ID (each subtask) was calculated as the average time needed to move the pen from one target to the other, to allow determination of the relationship between movement time and ID. A linear curve was then fitted to the data points and a least squares calculation was used to determine the goodness of fit (FittsR2). FittsR2 refers to the degree of compliance with Fitts' law and was determined for each patient. The slope of the fitted curve (FittsSlope) describes the extent to which the

performance becomes slower with an increase in ID and was calculated for each patient, as well.

Tremor analysis

For the resting, circle, and spiral tasks the gyroscope signals were analyzed to assess tremor. In detail, the procedure to extract tremor features consists of the following steps:

1. The gyroscope signals were filtered with a 5 samples long running median filter to remove artifactual peaks.
2. To dampen the lowest frequencies (< 3 Hz) to remove the frequency components not related to tremor, the gyroscope signals were filtered by removing the output of a second order Savitzky-Golay filter. The frame size of the filter is 0.33 times the sample rate of the signal for the circle and spiral tasks and 0.5 times the sample rate for the rest task. This filtering process increases the signal to noise ratio for the determination of the dominant tremor frequency.
3. A principal component analysis was performed on the three filtered gyroscope signals (pitch, yaw and roll) only for the periods in which the patient was performing the task. Then, the first principal component was selected.
4. Next, the power spectral density (PSD) of the first principal component was estimated using Welch's method (3 s segments with 2 s overlap, Fourier transform length of 2048 samples) and the band power in a 1 Hz band around each frequency (overlapping bins) was computed.
5. The PSD plots were inspected to determine whether PD patients showed tremor during the tasks. This was done by visually checking whether a clear peak between 4 and 9 Hz (typical tremor band [28]) was present in the PSD plot. Only the PD patients who did show a clear peak were selected for further analysis.
6. The frequency with the highest band power was selected as the tremor frequency for the PD patients who were selected in the previous step. The relative power band was calculated by dividing the power in the 1 Hz band around the peak frequency by the total power.

E. Statistical analysis

Statistical analyses were conducted using SPSS (IBM SPSS Statistics 22). Normality of measures was assessed by the Shapiro-Wilk test. Validity was estimated by analyzing the correlation between performance on the graphical tasks and on the PPT. The PPT yields scores of ordinal level and therefore the correlation between PPTright and the tracing, writing and modified Fitts' task measures was analyzed by Spearman's rank correlation. The correlation between performance on the graphical tasks and the MDS-UPDRS-III-bradykinesia subscore was also assessed. A Spearman's rank correlation coefficient between 0.90 and 1.00 was regarded as a very high correlation, between 0.70 and 0.90 as a high correlation, between 0.50 and 0.70 as a moderate correlation, between 0.30 and 0.50 as a low correlation and below 0.30 as a negligible correlation [29]. Response to dopaminergic medication was determined by a significant improvement on the movement time, writing size, FittsSlope and FittsR2 and tremor measures after taking dopaminergic medication, by a paired t-test for normally distributed measures or a related-samples Wilcoxon signed rank test, otherwise.

III. RESULTS

All PD patients (N=14, mean age 68 years, 10 male; see Table 1) completed the tasks OFF and ON medication. Some patients, who had missing data due to a technical problem, were excluded from statistical analysis (see Table 2). There was no significant difference between the MDS-UPDRS-III-bradykinesia subscore OFF and ON medication.

A. Validity

MT on the circle, spiral and zigzag tasks correlated significantly with the PPTright score: CircleMT (ρ=-0.69, $p<0.001$), SpiralMT (ρ=-0.63, $p<0.001$) and ZigzagMT (ρ=-0.66, $p<0.001$). The MT measures of the 'elelelel' writing task showed low correlations with the PPTright score (E_MT: ρ=-0.33, $p=0.08$ and L_MT: ρ=-0.37, $p=0.05$). None of the writing size measures correlated with the PPTright score. FittsSlope correlated moderately with the PPTright score (ρ=0.64, $p<0.001$) while FittsR2 did not correlate with the PPTright score. Figure 2 shows two examples of the relationship between the PPTright score and graphical task measures. None of the graphical task measures correlated with the MDS-UPDRS-III-bradykinesia subscore.

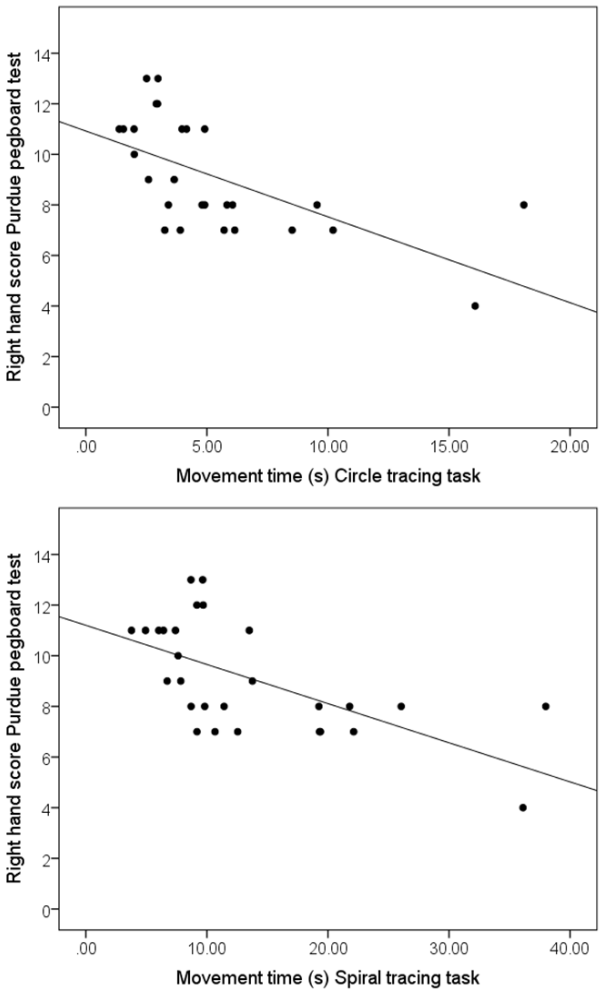


Figure 2. Scatterplots and regression lines for the right hand score on the Purdue pegboard test against movement time on the circle tracing task (Upper plot; Spearman's rho = -0.69) and against movement time on the spiral tracing task (Bottom plot; Spearman's rho =- 0.63).

B. Response to dopaminergic medication

Table 2 provides the test statistics for the response to dopaminergic medication. CircleMT, SpiralMT, E_MT, L_MT and FittsSlope were significantly lower at ON medication as compared to OFF medication (related-samples Wilcoxon signed rank test, $p<0.05$). The other measures did not respond to dopaminergic medication.

TABLE II.
TEST STATISTICS OF THE RESPONSE TO DOPAMINERGIC MEDICATION. MEDIAN AND INTERQUARTILE RANGE (IQR) VALUES ARE PROVIDED FOR BOTH 'ON' (ON) AND 'OFF' MEDICATION (OFF).

	OFF	**ON**	**Z-value[a]**	**p-value**
CircleMT (s) [b]	4.16 (5.56)	4.12 (3.39)	-2.55	0.01*
SpiralMT (s)	11.42 (10.12)	9.74 (10.93)	-2.61	0.01**
ZZMT (s) [c]	11.92 (3.81)	11.55 (8.04)	-1.18	0.24
E_MT (s)	0.39 (0.12)	0.36 (0.14)	-2.17	0.03*
L_MT (s)	0.52 (0.22)	0.48 (0.13)	-2.67	0.01**
E_width (mm)	9.23 (2.83)	7.79 (4.46)	-0.41	0.68
E_height (mm)	14.87 (7.79)	15.26 (11.06)	-0.22	0.83
L_width (mm)	12.84 (7.58)	12.81 (11.05)	-0.03	0.97
L_height (mm)	39.14 (18.86)	41.20 (26.12)	-0.03	0.97
FittsSlope [d]	0.14 (0.09)	0.12 (0.08)	-2.22	0.03*
FittsR2 [d]	0.89 (0.16)	0.93 (0.08)	-1.87	0.09

MT = Movement Time
a. Results of the related-samples Wilcoxon signed rank test
b. Data of PD007 was missing, due to a technical problem
c. Data of PD002 and PD006 was missing, due to a technical problem
d. Data of PD001, PD002 and PD003 was missing, due to a technical problem
* statistically significant at $\alpha = 0.05$; ** statistically significant at $\alpha = 0.01$

C. Tremor assessment

If patients did not suffer from tremor during the graphical tasks, the tremor (peak) frequency could not be determined. Further analysis was therefore only performed on patients exhibiting tremor during at least one of the tasks, indicated by the availability of tremor frequencies. This group was too small to perform statistical analysis, so we investigated each of the patients individually. Table 3 shows the results of the tremor analysis. Two patients exhibited tremor during the resting task at OFF medication, six patients during the circle task at OFF medication, and five patients during the spiral task at OFF medication. In general, RelPower decreased after taking medication and for most of the patients no tremor frequency was seen during the ON medication phase. Patient PD008 showed tremor during the resting task at ON medication, while RelPower clearly decreased. Tremor frequencies for PD013 during the circle and spiral tasks were available at ON medication, while RelPower did not clearly decrease. Only PD011 and PD013 had a MDS-UPDRS-III-tremor score higher than 0 (see Table 3).

IV. DISCUSSION

This study explored whether a set of graphical tasks could be used for monitoring PD by assessing its validity and response to dopaminergic medication. Moderate correlations between performance on graphical tasks and the PPT suggest that this set of graphical tasks is valid to assess and monitor upper limb function in PD patients. The advantage of using these graphical tasks, instead of, for instance, the PPT is that no examiner is needed, which makes the graphical tasks suitable for home-based monitoring. Additionally, this study showed that this set of graphical tasks can be used to detect subtle changes, which are barely visible by observing the patient, in performance after taking medication in PD patients. We investigated the validity and response to dopaminergic medication of different measures of MT on graphical tasks to assess bradykinesia.

Moderate correlations with the PPT score suggest that MT measures for the tracing tasks are valid for assessing and monitoring upper limb function in PD patients. In contrast, MT measures for the writing task showed only weak correlations with the PPT and seem to be less valid. FittsSlope, referring to the extent to which a patient becomes slower with increasing difficulty of Fitts' task, also correlated significantly with the PPT score, which means that FittsSlope can be used as a proper measure of bradykinesia. In agreement with previous studies regarding medication effects on performance of graphical tasks in PD [12], [14], [15], MT on the writing task improved after taking medication (see Table 2). In addition, FittsSlope and MT on the simple circle and spiral tracing tasks also improved after taking medication in PD patients. Simple circle and spiral tracing tasks might be easier to perform correctly than a writing task in home-based settings without an examiner present. We also studied the validity and response to dopaminergic medication of different measures of tremor (presence or absence of tremor, tremor frequency and relative power around the tremor frequency) during the resting task and circle and spiral tracing tasks. The present study showed that tremor power generally decreased after medication intake in PD patients, which was reported previously as well [28]. Since tremor often is a prominent and disabling symptom of PD that may be influenced by treatment, a useful monitoring tool should include a measure to assess tremor. Previous studies investigating medication effects on graphical tasks did not include a measure to assess tremor [11], [12], [14], [15]. Remarkably, MDS-UPDRS-III-tremor scores were 0 for almost all patients, although a tremor was detected during the graphical tasks. This suggests that graphical tasks might be more sensitive to detect a subtle tremor than the UPDRS. The difference in the presence of tremor between the UPDRS and the graphical tasks could be explained by the fact that tremor scoring during the UPDRS involves observing movements of the upper limb different from the movements which are involved in the graphical tasks. For example, a clinician observes whether tremor is present during a simple flexion movement of the upper limb, whereas the graphical tasks require more complex movements of the upper limb. In addition, a tremor with a low amplitude and high frequency could be difficult to observe during a neurological assessment, while graphical tasks performed with a sensor pen could capture such tremors.

Besides bradykinesia and tremor, micrographia is a common symptom in PD patients [5], [22], [30], which refers to a reduction in writing size and could be assessed quantitatively [5]. For screening PD, assessing micrographia has been shown to be useful, because differences were found between mildly affected PD and HC participants [31]. Our data showed that

TABLE III
TREMOR VARIABLES FOR PD PATIENTS WHO EXHIBITED TREMOR ON AT LEAST ONE OF THE TASKS

	Rest				Circle Tracing				Spiral Tracing				UPDRS Tremor Score (Right hand)	
	OFF		ON		OFF		ON		OFF		ON		OFF	ON
ID	Freq	RP	Freq	RP	Freq	RP	Freq	RP	Freq	RP	Freq	RP		
PD002	n/a	n/a	n/a	n/a	6.8	46.8	n/a	n/a	6.9	30.5	n/a	n/a	0	0
PD004	n/a	n/a	n/a	n/a	6.3	70.3	n/a	n/a	6.5	63.9	n/a	n/a	0	0
PD007	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0
PD008	4.9	84.9	n/a	n/a	5.1	66.5	n/a	n/a	5.2	62.5	n/a	n/a	0	0
PD009	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	0	0
PD011	5.7	87.8	n/a	n/a	7.6	24.3	n/a	n/a	5.6	48.2	n/a	n/a	2	0
PD012	n/a	n/a	n/a	n/a	7.4	30.6	n/a	n/a	n/a	n/a	n/a	n/a	0	0
PD013	n/a	n/a	n/a	n/a	7.1	33.8	7.0	37.0	6.5	43.1	6.3	42.0	3	3

n/a = not available. If no value is available for the tremor measures, no tremor was detected during this task. Freq = Tremor (peak) frequency; RP = Relative Power.

writing size measures correlated weakly with the PPT, suggesting that writing size, as assessed in this study, is not a valid measure to assess and monitor upper limb function in PD. Additionally, writing size measures did not change after taking medication in PD patients, in agreement with previous studies that investigated writing size ON and OFF medication [11]–[13]. However, duration and size of writing are related [5], [8], so bradykinesia could compensate for micrographia. Therefore, micrographia assessments could be improved by writing a fixed number of letters or words in a particular time frame.

No significant correlations were found between performance on graphical tasks and the MDS-UPDRS-III-bradykinesia score. This could be explained by the fact that the MDS-UPDRS-III-bradykinesia items entail different movements of the arm and hand than graphical tasks and therefore involve other upper limb functions. Hand function of most PD patients in this study was only mildly affected, according to the MDS-UPDRS-III-bradykinesia subscore. Subtle improvements of hand function after taking medication were therefore hard to detect according to the MDS-UPDRS-III. In contrast, performance of almost all PD patients improved on graphical tasks after taking medication, which suggests that graphical tasks are more suitable to detect subtle changes in upper limb function than the MDS-UPDRS-III.

To use the current system as a home-based monitoring tool, the analysis methods should and could be converted to automatic methods, which generate simple outcome measures for the clinician. In addition, a home-based system should include a clear instruction manual for patients and error detection and feedback to ensure correct task execution and allow use of the system without an examiner present. Once the system is fully automated, a new study would be needed to re-assess validity. To further investigate long-term disease progression and treatment effects, a longitudinal study should be performed in which PD patients will be followed for a longer time-period.

As a limitation, the medication effects in this study may have been influenced by a learning effect, because the measurements ON medication were performed approximately 1.5 hours after the measurements OFF medication. However, since the tasks were easy to perform and not all patients improved at the second measurement, it is assumed that this learning effect is not substantial. Another limitation of this study is that the resting task involved holding the pen-tip at a certain location, which makes it a postural rather than a resting task. Therefore this task is not suited to assess resting tremor. Since resting tremor is typical for PD patients, another resting task should be included in future studies that actually measures resting tremor. Furthermore, it would be interesting to assess the responsiveness of the graphical tasks. This could be done by comparing the change in performance on the graphical tasks with the clinically accepted change between ON and OFF medication. The clinically accepted change for PD patients would be a change in the UPDRS score. Since the patients in our study were only mildly affected and showed low UPDRS scores, a clear change in UPDRS was often not seen between ON and OFF medication. Therefore we did not compare the changes in graphical tasks to the small and often absent changes in UPDRS score, but only compared performance on the graphical tasks ON and OFF medication and pooled measurements ON and OFF medication to allow assessment of the validity of the graphical tasks.

In conclusion, the present study showed that our set of graphical tasks using a digital tablet and sensor-pen is valid to assess and monitor upper limb functioning in PD patients, especially with respect to the circle, spiral and modified Fitts' task and their derived measures of MT and tremor. Our method is non-invasive, portable and can be used easily at home without an examiner, which offers great opportunities as an endpoint for future medication trials.

REFERENCES

[1] J. Massano and K. P. Bhatia, "Clinical approach to Parkinson's disease: features, diagnosis, and principles of management.," *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 6, p. a008870, Jun. 2012.

[2] C. G. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, G. T. Stebbins, M. B. Stern, B. C. Tilley, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A. E. Lang, A. Lees, S. Leurgans, P. a. LeWitt, D. Nyenhuis, C. W. Olanow, O. Rascol, A. Schrag, J. a. Teresi, J. J. Van Hilten, and N. LaPelle, "Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan," *Mov. Disord.*, vol. 22, no. 1, pp. 41–47, 2007.

[3] J. D. Hoffman and J. McNames, "Objective measure of upper extremity motor impairment in Parkinson's disease with inertial sensors," *2011 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 4378–4381, 2011.

[4] D. a. Heldman, A. J. Espay, P. a. LeWitt, and J. P. Giuffrida, "Clinician versus machine: Reliability and responsiveness of motor endpoints in Parkinson's disease," *Park. Relat. Disord.*, vol. 20, no. 6, pp. 590–595, 2014.

[5] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, "From micrographia to Parkinson's disease dysgraphia," *Mov. Disord.*, vol. 29, no. 00, pp. 1467–1475, 2014.

[6] E. J. Smits, A. J. Tolonen, L. Cluitmans, M. Van Gils, B. a. Conway, R. C. Zietsma, K. L. Leenders, and N. M. Maurits,

"Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease," *PLoS One*, vol. 9, no. 5, pp. 1–8, 2014.

[7] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, "Handwriting as an objective tool for Parkinson's disease diagnosis," *J. Neurol.*, vol. 260, pp. 2357–2361, 2013.

[8] M. P. Broderick, A. W. a Van Gemmert, H. a Shill, and G. E. Stelmach, "Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease.," *Exp. Brain Res.*, vol. 197, no. 3, pp. 223–33, Aug. 2009.

[9] A. Ünlü, R. Brause, and K. Krakow, "Handwriting Analysis for Diagnosis and Prognosis of Parkinson ' s Disease," *ISBMDA Lect. Notes Comput. Sci.*, vol. 4345, pp. 441–450, 2006.

[10] a W. Van Gemmert, H. L. Teulings, and G. E. Stelmach, "Parkinsonian patients reduce their stroke size with increased processing demands.," *Brain Cogn.*, vol. 47, pp. 504–512, 2001.

[11] H. Ling, L. a. Massey, A. J. Lees, P. Brown, and B. L. Day, "Hypokinesia without decrement distinguishes progressive supranuclear palsy from Parkinson's disease," *Brain*, vol. 135, pp. 1141–1153, 2012.

[12] C. Bidet-Ildei, P. Pollak, S. Kandel, V. Fraix, and J. P. Orliaguet, "Handwriting in patients with Parkinson disease: Effect of l-dopa and stimulation of the sub-thalamic nucleus on motor anticipation," *Hum. Mov. Sci.*, vol. 30, no. 4, pp. 783–791, 2011.

[13] W. G. Ondo and P. Satija, "Withdrawal of visual feedback improves micrographia in Parkinson's disease," *Mov. Disord.*, vol. 22, no. 14, pp. 2130–2131, 2007.

[14] K. W. Lange, L. Mecklinger, S. Walitza, G. Becker, M. Gerlach, M. Naumann, and O. Tucha, "Brain dopamine and kinematics of graphomotor functions," *Hum. Mov. Sci.*, vol. 25, pp. 492–509, 2006.

[15] O. Tucha, L. Mecklinger, J. Thome, a. Reiter, G. L. Alders, H. Sartor, M. Naumann, and K. W. Lange, "Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease," *J. Neural Transm.*, vol. 113, pp. 609–623, 2006.

[16] J. Westin, S. Ghiamati, M. Memedi, D. Nyholm, A. Johansson, M. Dougherty, and T. Groth, "A new computer method for assessing drawing impairment in Parkinson's disease," *J. Neurosci. Methods*, vol. 190, no. 1, pp. 143–148, 2010.

[17] C. a. Haaxma, B. R. Bloem, G. F. Borm, and M. W. I. M. Horstink, "Comparison of a timed motor test battery to the Unified Parkinson's Disease Rating Scale-III in Parkinson's disease," *Mov. Disord.*, vol. 23, no. 12, pp. 1707–1717, 2008.

[18] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, and S. L. Pullman, "Validity of spiral analysis in early Parkinson's disease.," *Mov. Disord.*, vol. 23, no. 4, pp. 531–7, Mar. 2008.

[19] T. E. Eichhorn and T. Gasser, "Computational Analysis of Open Loop Handwriting Movements in Parkinson ' s Disease□: A Rapid Method to Detect Dopamime tic Effects," vol. 11, no. 3, pp. 289–297, 1996.

[20] E. L. Proud and M. E. Morris, "Skilled Hand Dexterity in Parkinson's Disease: Effects of Adding a Concurrent Task," *Arch. Phys. Med. Rehabil.*, vol. 91, no. 5, pp. 794–799, 2010.

[21] T. Müller and S. Benz, "Quantification of the dopaminergic response in Parkinson's disease," *Park. Relat. Disord.*, vol. 8, pp. 181–186, 2002.

[22] J. Jankovic, "Parkinson's disease: clinical features and diagnosis.," *J. Neurol. Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–76, Apr. 2008.

[23] M. K. Rand, G. E. Stelmach, and J. R. Bloedel, "Movement accuracy constraints in Parkinson's disease patients," *Neuropsychologia*, vol. 38, pp. 203–212, 2000.

[24] a J. Hughes, S. E. Daniel, L. Kilford, and a J. Lees, "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.," *J. Neurol. Neurosurg. Psychiatry*, vol. 55, pp. 181–184, 1992.

[25] C. G. Goetz et al., "Movement Disorder Society Task Force report on Hoehn and Yahr staging scale: Status and recommendations.," *Mov. Disord.*, vol. 19, no. 9, pp. 1020-1028, 2004.

[26] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement. 1954.," *J. Exp. Psychol. Gen.*, vol. 121, no. 3ParP, pp. 262–269, 1992.

[27] J. Desrosiers, R. Hébert, G. Bravo, and E. Dutil, "The Purdue Pegboard Test: normative data for people aged 60 and over.," *Disabil. Rehabil.*, vol. 17, no. 5, pp. 217–224, 1995.

[28] B. S. Connolly and A. E. Lang, "Pharmacological treatment of Parkinson disease: a review.," *JAMA*, vol. 311, no. 16, pp. 1670–83, 2014.

[29] M. M. Mukaka, "A guide to appropriate use of Correlation coefficient in medical research.," Malawi Med J.., vol. 24, no. 3, pp. 69-71, 2012.

[30] J. E. Mclennan, K. Nakano, H. R. Tyler, and R. S. Schwab, "Micrographia in Parkinson ' s Disease .," no. 5, pp. 141–152, 1972.

[31] M. M. Ponsen, A. Daffertshofer, E. C. Wolters, P. J. Beek, and H. W. Berendse, "Impairment of complex upper limb motor function in de novo Parkinson's disease," *Park. Relat. Disord.*, vol. 14, pp. 199–204, 2008.

OPEN ACCESS Freely available online

PLOS ONE

Standardized Handwriting to Assess Bradykinesia, Micrographia and Tremor in Parkinson's Disease

Esther J. Smits[1]*, Antti J. Tolonen[2], Luc Cluitmans[2], Mark van Gils[2], Bernard A. Conway[3], Rutger C. Zietsma[4], Klaus L. Leenders[1], Natasha M. Maurits[1]

1 Department of Neurology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands, **2** VTT Technical Research Centre of Finland, Tampere, Finland, **3** The Department of Biomedical Engineering, University of Strathclyde, Glasgow, United Kingdom, **4** Manus Neurodynamica Ltd, Newcastle, United Kingdom

Abstract

Objective: To assess whether standardized handwriting can provide quantitative measures to distinguish patients diagnosed with Parkinson's disease from age- and gender-matched healthy control participants.

Design: Exploratory study. Pen tip trajectories were recorded during circle, spiral and line drawing and repeated character 'elelelel' and sentence writing, performed by Parkinson patients and healthy control participants. Parkinson patients were tested after overnight withdrawal of anti-Parkinsonian medication.

Setting: University Medical Center Groningen, tertiary care, the Netherlands.

Participants: Patients with Parkinson's disease (n = 10; mean age 69.0 years; 6 male) and healthy controls (n = 10; mean age 68.1 years; 6 male).

Interventions: Not applicable.

Main Outcome Measures: Movement time and velocity to detect bradykinesia and the size of writing to detect micrographia. A rest recording to investigate the presence of a rest-tremor, by frequency analysis.

Results: Mean disease duration in the Parkinson group was 4.4 years and the patients were in modified Hoehn-Yahr stages 1–2.5. In general, Parkinson patients were slower than healthy control participants. Median time per repetition, median velocity and median acceleration of the sentence task and median velocity of the elel task differed significantly between Parkinson patients and healthy control participants (all $p<0.0014$). Parkinson patients also wrote smaller than healthy control participants and the width of the 'e' in the elel task was significantly smaller in Parkinson patients compared to healthy control participants ($p<0.0014$). A rest-tremor was detected in the three patients who were clinically assessed as having rest-tremor.

Conclusions: This study shows that standardized handwriting can provide objective measures for bradykinesia, tremor and micrographia to distinguish Parkinson patients from healthy control participants.

Citation: Smits EJ, Tolonen AJ, Cluitmans L, van Gils M, Conway BA, et al. (2014) Standardized Handwriting to Assess Bradykinesia, Micrographia and Tremor in Parkinson's Disease. PLoS ONE 9(5): e97614. doi:10.1371/journal.pone.0097614

Editor: Josef Priller, Charité-Universitätsmedizin Berlin, Germany

Received November 27, 2013; **Accepted** April 22, 2014; **Published** May 22, 2014

Funding: This study is part of the DiPAR project for which funding was received from the EC in the FP7-SME-201001 programme (grant agreement 262291). VTT and UMCG collaborated as subcontracted RTD performers. Manus Neurodynamica Ltd, Newcastle, United Kingdom provided support in the form of salaries for author RZ, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: One of the authors, NM, is a PLOS ONE Editorial Board member. One of the authors, RZ, is employed by a commercial company, Manus Neurodynamica Ltd, Newcastle, UK. There is a patent associated to this manuscript (patent WO/2011/141734, name: Apparatus for use in diagnosing and/or treating neurological disorder). There are no further patents, products in development or marketed products to declare in relation to this study. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and material.

* E-mail: e.j.smits@umcg.nl

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder which generally results in several motor symptoms. The cardinal signs of the disease are bradykinesia (slowness of movement), rest tremor, rigidity (muscular stiffness throughout the range of passive movement in a limb segment) and postural and gait impairment [1]. Not all PD patients present these classical symptoms and several other motor symptoms can be observed, such as freezing, shuffling gate, hypomimia and micrographia (small handwriting) [2]. Clinical examination can be expressed in rating scales, e.g. the Unified Parkinson's Disease Rating Scale (UPDRS) or the Hoehn and Yahr scale (H&Y) [2,3]. The UPDRS is the most widely used and tested scale and consists of an impairment and disability section. The H&Y scale is the most commonly used method to assess the severity of the disease [4]. However, rating scales highly

depend on the experience and interpretation of the physician performing the assessment and have limited precision for quantifying upper limb motor skill. To support the clinical diagnosis, a trial dose of levodopa should result in an improvement of the clinical symptoms. The clinical diagnosis can also be supported by radiotracer neuroimaging techniques such as positron emission tomography or single photon emission computed tomography, in which a presynaptic dopaminergic deficit can be demonstrated [5].

Early diagnosis of PD is very important, because it allows early intervention and management toward an improved overall outcome for the patient [6]. Currently, no definite methods for an early, objective and quantitative diagnosis are available, but several methods that provide quantitative measures for motor symptoms of PD have been studied. For example, handwriting tasks and systems have been used for this purpose [7–13]. Bajaj et al. [7] used handwritten samples to differentiate PD patients from patients with other tremors. They provided an objective measure for micrographia, but their analysis was time consuming, because script height and length were measured manually. An electronic pen and digitizer tablet were used in other studies to distinguish PD from healthy control (HC) participants [8,10,12]. However, Alty et al. [8] only studied bradykinesia and Van Gemmert et al. [12] only studied micrographia. Broderick et al. [10] studied both micrographia and bradykinesia, but the shoulder and elbow of participants were fixated, which resulted in a constrained, rather unnatural movement. Ünlü et al. [9] used an electronic pen as well and showed that several features can be computed to distinguish PD from HC. One of the features was related to tremor, but the remaining features were not related to a symptom of PD. Rosenblum et al.[13] also used handwriting to distinguish PD from HC analyzing movement speed and size of writing. They did not assess tremor. Thus, each of the systems provided useful measures to distinguish PD from HC participants, but most of them focused on just one of the motor symptoms of PD and none of these studies included a task to measure rest tremor. For early differential diagnosis a system which provides quantitative measures for several motor symptoms of PD simultaneously would be beneficial [14].

The aim of the present study was to determine whether standardized handwriting can provide quantitative measures to assess multiple important motor symptoms simultaneously to distinguish patients diagnosed with PD from age- and gender matched HC participants. The study focused on two important motor symptoms of PD, bradykinesia and micrographia. Additionally, rest tremor was investigated. The design of the present study was exploratory and therefore a small group of PD patients and HC participants was included and a large number of features was produced, to examine which features can best be used to distinguish PD from HC.

Several handwriting and geometric tasks, based on tasks used in previous studies, were evaluated. Ünlü et al. [9] used the writing of l-loops and a complete sentence. In a study of Ponsen et al.[15] participants wrote a complete sentence and the authors showed that letter height decreased in PD patients as writing progressed. Also Bajaj et al.[7] assessed micrographia in PD by analyzing a handwritten sentence. The present study includes the writing of e- and l-loops and a complete sentence to assess micrographia.

Besides writing tasks, geometric tracing tasks were included in this study, based on previous findings. For example, Kereszténvi et al.[11] used a circle tracing task to show that PD patients were significantly slower than HC. Other studies [16,17] also investigated a circle drawing task to compare PD with HC. Saunders-Pullman et al.[18] showed a correlation between spiral analysis and the UPDRS score and Stanley et al.[19] described that spiral analysis may be more sensitive than the UPDRS for detecting early changes in motor performance. Dounskaia et al.[16] showed that drawing lines in different directions differentiated between PD and HC. For example, line drawing variability was higher in PD than in HC. Therefore, in the present study line drawing in eight different directions was included in addition to circle and spiral tracing tasks. A rest task was added as well, based on the task used by Scanlon et al.[20] to measure rest tremor.

To summarize, the present study aimed to provide quantitative measures to evaluate bradykinesia, micrographia and tremor in one assessment by recording pen tip movement during handwriting tasks, including tracing geometric figures and actual writing. We additionally assessed whether these features allowed distinguishing PD patients from HC participants.

Methods

Ethics Statement

The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen.

Participants

Ten patients with PD (mean age 69.0 years; range 63–81, 6 male) and ten gender- and age- matched HC participants (mean age 68.1 years; range 61–78, 6 male) participated. Patients, who are clinically diagnosed with PD by a neurologist (according to the United Kingdom Parkinson's Disease Society Brain Bank Diagnostic Criteria for Parkinson's Disease [21]) and who are under treatment at the movement disorders clinic in the University Medical Center Groningen (UMCG) were contacted retrospectively. Since the patients had to be able to hold a pen for 30 minutes and perform tracing and writing tasks, PD patients in relatively early stages of the disease (modified H&Y stage 1–2.5 [3,22]) were selected. The first ten patients who replied positively and met the inclusion criteria were included. The healthy participants were recruited from the general population and were matched to the patients by their age and gender. All participants were right-handed according to the Annett handedness scale [23] and signed informed consent before participation. All PD patients complied with overnight withdrawal of PD-related medication. Exclusion criteria were a history of epileptic seizures, head injury, neurological disorders (other than PD for the patients), the use of medication affecting movement, or a low (<26) score on the Mini Mental State Examination (MMSE). Patients who suffered from a severe tremor in the hands (score ≥ 3 on the UPDRS-III) were excluded from the study, because this study mainly focused on bradykinesia and micrographia. Table 1 shows a summary of the patient characteristics.

Experimental design

Participants were seated in front of a table in a comfortable position to write. As was shown before [9,14,16], a digitizer pen and tablet are suitable to record handwriting. A graphic tablet[a] (WACOM Intuos 2) and a modified digitizer pen were used. The position of the pen-tip on the tablet during movement was recorded using the MovAlyzeR software[b] (Neuroscript LLC, USA) with a sampling frequency of 100 Hz. The pen had a wired connection to an operator computer where MovAlyzeR was installed. Participants performed five drawing and writing tasks (see below) using the digitizer pen. The examiner was seated behind the operator computer and determined whether the participants executed the tasks correctly. If a task was executed incorrectly, the recording was stopped and restarted after re-

Table 1. PD patient characteristics.

Patient no.	Age (years)	Gender	Disease duration (years)	Modified Hoehn and Yahr scale	UPDRS score (last visit)
1	63	M	5	2	6
2	81	M	5	2.5	35
3	79	M	3	2	20
4	78	M	4	2	18
5	62	F	4	1	11
6	64	M	4	1.5	12
7	67	F	8	1.5	13
8	67	M	5	1	*
9	65	F	2	1.5	11
10	64	F	4	1.5	11

*No UPDRS score was available for this patient.
doi:10.1371/journal.pone.0097614.t001

instruction. An example of incorrect task execution would be moving the pen in the wrong direction or starting the task too early.

Tasks

Each participant performed five tasks in the same order, to limit variability in task results. Participants were instructed to start the task at a signal of the examiner and to perform the tasks at a comfortable speed, allowing them to write and draw as smoothly as individually possible. First, a rest recording (30 seconds) was performed prior to the writing and drawing tasks to measure pen movement at rest. The participants were instructed to touch the tablet with the pen-tip, with the lower right arm resting on the table [20]. Next, the participants traced geometric shapes on templates; a circle, a star and a spiral (Figure 1). The templates were printed on A4 paper and placed on the tablet under a transparent sheet.

Circle drawing. In this task, participants had to continuously trace a circle ten times in a clockwise direction starting from the 12 o'clock position (Figure 1).

Star drawing. Straight lines orientated in eight different directions, set at 45 degrees to each other and forming an eight pointed star, were traced in this task (Figure 1). The lines had to be repeatedly traced from the central point of the star to each endpoint and back, ten times without interruption, starting with the upward direction and then proceeding clockwise.

Spiral drawing. In this task the participants traced a spiral (Figure 1) clockwise from inside to outside. Each participant performed ten consecutive spiral tracing trials.

During the last two tasks, the participants wrote a particular phrase ten times. The texts were chosen such that symbols and words were written repetitively and texts were nonsensical in one case and meaningful in the other.

'elel' character writing. In this task the participant wrote the 8 character text sequence '*elelelel*' ten times with each phrase starting at the left side of the tablet.

Sentence writing. In this task, the participant wrote the sentence: 'veel te veel felle schelle zon' ('way too much bright, shrill sun' in Dutch), ten times.

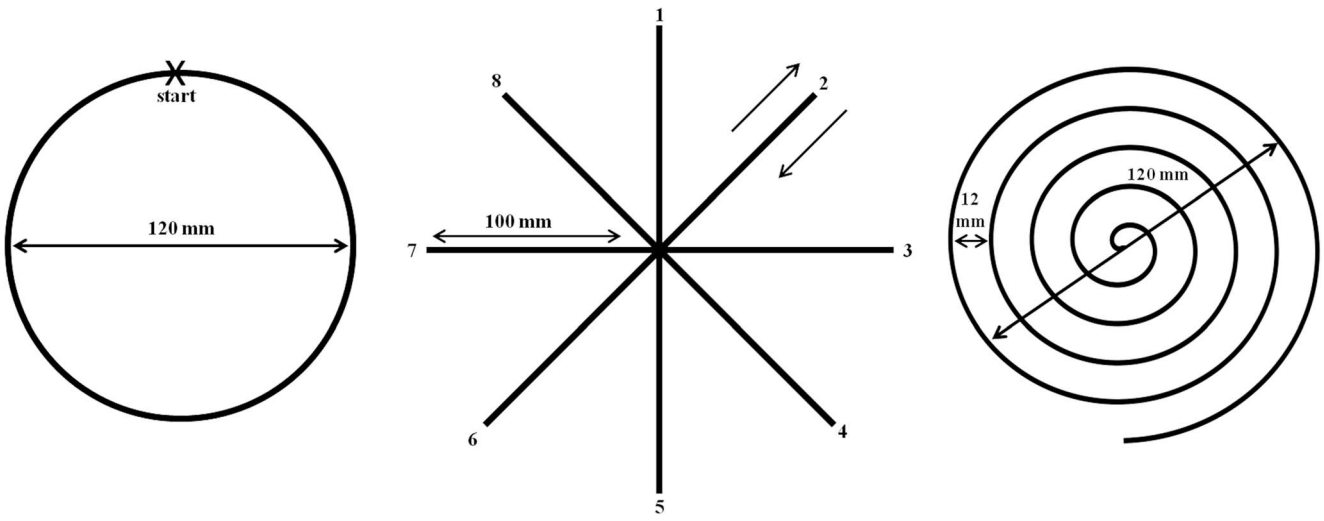


Figure 1. Templates used for tracing geometric shapes; circle, star and spiral. The dimensions of the templates are indicated in the figure.
doi:10.1371/journal.pone.0097614.g001

Data analysis

Using custom made scripts in Matlab 7.4.0 (R2007a) the drawing and writing tasks were analyzed to evaluate the speed of movement to assess bradykinesia and the size of writing to assess micrographia. Additionally, a frequency analysis was performed to assess rest-tremor. The data were preprocessed to allow for evaluation of each separate trial as well as of the whole task (see methods S1 and figure S1). The pen position data for the star task were divided into four main directions, for comparison. Directions 1 and 5 (see Figure 1) were taken together as the vertical direction, 3 and 7 as the horizontal direction, 2 and 6 as diagonal1 and 4 and as diagonal2. The data points were assigned to the main directions (see methods S1 and figure S2). Separating each line of the '*elel*' task and recognizing the individual letters was done using a state vector machine (see methods S1 and figure S3). The start and end points of each sentence were selected manually.

Bradykinesia assessments. To assess bradykinesia, features concerning movement speed were defined. Total movement time was calculated for the circle, spiral and star task. Median time for each trial was calculated for the circle, spiral and sentence task. Median velocity and acceleration were calculated for all tasks. For the star task median time for each line was calculated for the whole task as well as for the four main directions. Finally, for the '*elel*' task median times for writing an 'e' or an 'l' were calculated yielding 23 bradykinesia features in total.

Micrographia assessment. To assess micrographia, writing size was investigated. For the '*elel*' task median width and height of the individual letters '*e*' and '*l*' were calculated. For the sentence task median script height and median sentence length were calculated, yielding six micrographia features in total.

Tremor assessment

Data collected during the rest task were used to investigate the presence of a rest-tremor. To detect the tremor, the data of the pen tip location (x and y) were analyzed. First, the difference signals, dx and dy, for x and y were computed according to:

$$dx(n) = x(n) - x(n-1)$$
$$dy(n) = y(n) - y(n-1)$$

where n is the signal's sample index. Then a principal component analysis was performed and the first principal component of dx and dy was selected, to take into account all possible directions for rest-tremor. The first principal component is the linear combination of dx and dy with the highest variance. The power spectrum of the first principal component was computed using Welch's method. Finally, the spectral maximum was identified and the power spectral density (PSD) and frequency at the peak were determined.

Statistical analysis

Statistical analyses were conducted using SPSS 20.0.0.1. First, it was tested whether features were normally distributed by the Shapiro-Wilk test. For both groups, all features were described by their mean and standard deviation when normally distributed, or median and interquartile range (iqr), when not normally distributed. Since the goal was to derive quantitative measures for bradykinesia, micrographia and rest tremor and to assess whether these features could be used to distinguish PD patients from HC participants, the bradykinesia and micrographia features were compared between the two groups. Since only a few patients had rest tremor, related features were not compared further. To compare the bradykinesia and micrographia features between the two groups, multiple independent t-tests were performed for the features which were normally distributed and the Mann Whitney test was used when normality assumptions were violated. The statistical analyses were corrected for multiple comparisons by applying a Bonferroni correction. After Bonferroni correction a probability value (p) of ≤ 0.0014 (0.05/35) was considered significant for the bradykinesia and micrographia assessments. Additionally, to investigate the progressive reduction in writing size the difference between the first and last trial was computed for the width and height of the letters 'e' and 'l' and the length and height of the sentence and also compared between the two groups with multiple independent t-tests. After Bonferroni correction a probability value (p) of ≤ 0.0014 (0.05/35) was considered significant. Median time per line, which was normally distributed over participants, was compared between the four main directions of the star task according to a repeated measures ANOVA with between-subjects factor Group (PD and HC) and within-subject factor Direction (four main directions).

Results

All participants completed each of the writing and drawing tasks. Median disease duration of the PD patients was 4.4 years (range 2–8) and nine PD patients normally used Parkinsonian medication.

Bradykinesia assessments

Table 2 provides the test statistics for the bradykinesia features. Four bradykinesia features (median time per repetition, median velocity and median acceleration of the sentence and median velocity of the 'elel' task) differed significantly between PD and HC (all $p \leq 0.0014$). The remaining features also showed large differences between the two groups, although significance did not survive correction for multiple comparisons. Median time per line differed significantly between the four main directions of the star ($F(3,16) = 9.35$, $p = 0.001$), because median time per line was significantly higher in diagonal2 (0.81 s.) compared to diagonal1 (0.71 s.). No significant interaction was found.

Micrographia assessments

The test statistics for the micrographia assessments are also shown in Table 2. Sentence length and sentence script height did not differ significantly between PD and HC. The width of the letter 'e' was significantly smaller in PD than in HC ($p \leq 0.0014$). The height of the letter 'e' and the width and height of the letter 'l' in the 'elel' task were smaller in PD compared to HC, although significance did not survive correction for multiple comparisons (see Figure 2 for an example of writing). No other significant effects were found concerning writing size and progressive reduction in writing size.

Tremor assessments

The PSD at the peak was higher (>30 $(mm/s^2)/Hz$) for three PD patients who were clinically assessed as having rest-tremor, than for all other participants (<2 $(mm/s^2)/Hz$). The peak frequencies for these patients were between 4.4 and 8 Hz (PD2 8.0 Hz; PD3 5.3 Hz; PD7 4.4 Hz).

Discussion

The present study showed that handwriting tasks can provide objective measures for bradykinesia, micrographia and rest tremor that distinguish PD from HC.

Corresponding to earlier studies [8,10,11,13,15] results from the current study showed that PD patients perform movements

Table 2. Summary of test statistics of the bradykinesia and micrographia features, mean (SD) values for both groups are provided in case of a normal distribution, otherwise Median (iqr) values are shown; for the normal distributed features an independent t-test was performed, otherwise a Mann Whitney U test was performed.

Task	Feature	PD	HC	t-value #	p-value	
Circle	Total Movement time (s)	37.27 (13.08)	22.87 (5.99)	3.17	0,0077	
Circle	Median time per repetition (s)	3.24 (2.26)°	2.19 (0.59)°	18 #	0,0150	
Circle	Median velocity (m/s)	0.11 (0.05)	0.18 (0.06)	−2.94	0,0087	
Circle	Median Acceleration (m/s^2)	0.29 (0.32)°	0.47 (0.40)°	19 #	0,0190	
Cross	Total Movement time (s)	175.76 (66.55)	106.38 (36.52)	2.89	0,0098	
Cross	Median time per line (all) (s)	0.94 (0.34)	0.56 (0.21)	3.04	0,0070	
Cross	Median time per line (diagonal 1) (s)	0.89 (0.29)	0.54 (0.20)	3.09	0,0067	
Cross	Median time per line (diagonal 2) (s)	1.03 (0.43)	0.58 (0.25)	2.82	0,0112	
Cross	Median time per line (horizontal) (s)	0.91 (0.29)	0.62 (0.22)	2.58	0,0187	
Cross	Median time per line (vertical) (s)	0.98 (0.40)	0.53 (0.20)	3.20	0,0072	
Cross	Median Velocity (m/s)	0.11 (0.06)	0.17 (0.06)	−2.48	0,0234	
Cross	Median Acceleration (m/s^2)	0.41 (0.44)°	0.96 (1.44)°	22 #	0,0350	
Spiral	Total Movement time (s)	122.69 (59.00)°	83.39 (40) °	14 #	0,0050	
Spiral	Median time per repetition (s)	10.36 (5.36)°	6.79 (3.79)°	16 #	0,0090	
Spiral	Median velocity (m/s)	0.10 (0.05)	0.15 (0.05)	−2.40	0,0274	
Spiral	Median Acceleration (m/s2)	0.29 (0.27)°	0.54 (0.79)°	19 #	0,0190	
Sentence	Median time per repetition	16.30 (4.94)°	11.18 (2.92)°	3 #	0,0000	*
Sentence	Median velocity (m/s)	0.05 (0.02)	0.08 (0.02)	−4.22	0,0005	*
Sentence	Median Acceleration (m/s^2)	0.78 (0.32)	1.77 (0.39)	−6.23	0,0000	*
Elel	Median velocity (m/s)	0.07 (0.04)	0.15 (0.04)	−4.18	0,0006	*
Elel	Median Acceleration (m/s^2)	0.61 (0.36)	1.52 (0.71)	−3.62	0,0030	
Elel	Median duration letter e (s)	0.48 (0.13)	0.37 (0.10)	2.17	0,0441	
Elel	Median duration letter l (s)	0.74 (0.26)	0.51 (0.14)	2.53	0,0209	
	Micrographia features					
Elel	Median Width of the e (mm)	7.67 (3.42)	14.16 (3.34)	−4.29	0,0004	*
Elel	Median Height of the e (mm)	16.30 (6.57)	24.29 (5.90)	−2.86	0,0104	
Elel	Median Width of the l (mm)	12.82 (4.83)	19.15 (5.40)	−2.76	0,0129	
Elel	Median Height of the l (mm)	42.64 (15.45)	59.86 (13.82)	−2.63	0,0171	
Sentence	Median Script Height (mm)	13.46 (5.91)	18.03 (4.41)	−1.96	0,0660	
Sentence	Median Sentence Length (mm)	228.36 (116.19)°	275.96 (30.20)°	25 #	0,0630	
Elel	Difference first-last trial Width e (mm)	0.0030 (0.36)	0.13 (0.28)	−0,89	0,3870	
Elel	Difference first-last trial Height e (mm)	−0.032 (0.44)	0.37 (0.38)	−2,21	0,0410	
Elel	Difference first-last trial Width l (mm)	−0.12 (0.69)	0.25 (0.38)	−1,48	0,1560	
Elel	Difference first-last trial Height l (mm)	−0.89 (1.48)	−0.13 (0.71)	−1,46	0,1620	
Sentence	Difference first-last trial Script Height (mm)	−2881.50 (2387.90)	−898.60 (2138.36)	−1,96	0,0660	
Sentence	Difference first-last trial Script Length (mm)	−271.00 (297.80)	48.50 (410.81)	−1,99	0,0620	

°Median (iqr).
SD = Standard Deviation.
iqr = interquartile range.
#The values which are marked with a # are the U-values of the Mann Whitney U test, otherwise a t-value is shown.
*indicates a Bonferroni corrected significant result at $\alpha = 0.0014$.
doi:10.1371/journal.pone.0097614.t002

significantly slower than HC. PD patients were likely slower than HC, because of bradykinesia [10]. However, some caution is needed when drawing this conclusion, because it is crucial to distinguish bradykinesia from simple age-related slowness [1]. However, the groups were age-matched, which suggests that the decreased movement speed in PD patients reflects bradykinesia rather than just age-related slowness. All bradykinesia features showed large differences between the two groups, but only four features were significantly different between the two groups. These four features were derived from data obtained during the writing tasks, which were more complex than the tracing tasks. Moroney et al. [24] also showed in a simulation model that PD patients were slower than HC in both simple and complex movements, but slowness increased with increased movement complexity.

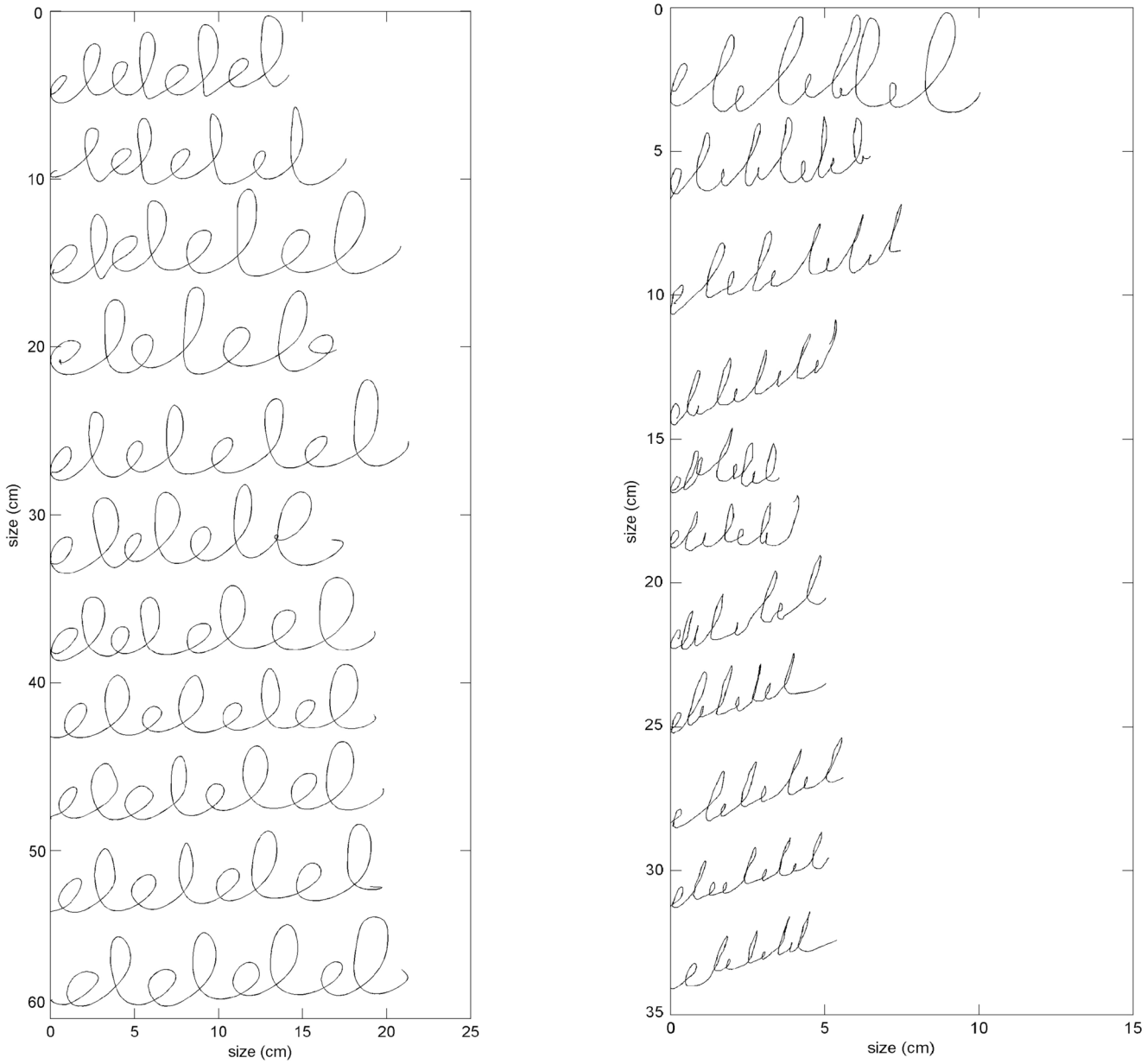


Figure 2. An example of the '*elel*' task is shown for a HC participant (left) and a PD patient (right). Each line of the writing task was shifted vertically so that individual trials are visible. Note the differences in the x-axis and y-axis between the left and right figure.
doi:10.1371/journal.pone.0097614.g002

Writing size was examined to find objective measures for micrographia. Micrographia is a symptom frequently associated with PD and is reflected in smaller sized writing patterns [25,26], but has also been defined as a progressive reduction in amplitude during a writing task [26]. In the current study PD patients produced smaller handwriting than HC as represented by smaller average width and height of the letters 'e' and 'l' in the '*elel*' task (note that only the width of the letter 'e' differed significantly between groups). This result was similar to the findings of Van Gemmert et al.[12] and Rosenblum et al.[13]. They showed reduced stroke sizes in PD patients compared to HC participants who performed handwriting tasks. We investigated the progressive reduction in writing size during a task as well, and there was a small reduction in size of different letter features, but there were no significant differences between the two groups. This result is in contrast with observations by Ponsen et al.[15], who showed a progressive reduction in writing size in PD patients. The fact that the present study showed no progressive reduction in writing size during the tasks might be due to the lack of visual feedback on the tablet during the tasks as the stylus is non-inking. De Jong et al.[27] described that PD patients drew larger when no visual feedback was available. Ondo et al.[28] also showed that withdrawal of visual feedback during actual writing improved micrographia in PD patients. Therefore, in future studies handwriting with visual feedback should be analyzed, because this might improve the sensitivity of micrographia measures.

In addition, participants were asked how frequently they practiced handwriting in their daily lives to investigate whether

the differences between groups are not a result of a lack of practice. The participants, both PD and HC, only wrote small amounts in their daily lives, such as a shopping-list, so we assume that the differences between the two groups in this study are a result of PD rather than a lack of practice.

Furthermore, rest tremor was detected in the patients who were clinically assessed as having rest tremor by the handwriting system described in this paper. The strength of combining handwriting tasks as was done in the present study is that three important motor symptoms of PD are assessed simultaneously. Handwriting tasks could be useful for screening PD in patients with mild symptoms: they are easily applicable in the clinic, since only a digitizer pen and tablet are needed to perform the measurements. Before such a handwriting system would be implemented in the clinic a future longitudinal study should investigate which participants with a high risk to develop PD, based on the handwriting measurements, will actually develop PD. Furthermore, future studies should investigate whether PD can be distinguished from other movement disorders using these handwriting tasks. Additionally, the custom made Matlab-scripts should be converted to automatic methods, which generate simple outcome measures for the clinician. Finally, handwriting analysis could also be useful for monitoring the effects of rehabilitation programs or other interventions.

One of the limitations of this study was the small sample size, which limits the number of statistically significant results. However, almost all features showed a clear difference between the groups ($p<0.05$), although they did not all survive Bonferroni correction. In addition, this study does not include a comparison with the clinical examination of the motor symptoms of PD. However, previous studies have already demonstrated correlations between separate handwriting tasks and clinical examinations [8,18]. The handwriting test battery presented in this study might be further improved by including a measure for rigidity, which is one of the classical symptoms of PD [1]. However, rigidity is a symptom which is very hard to quantify, because it refers to an increased muscle tone noticed during subjective assessment by a physician during passive movements of, for example, an affected arm [1].

Conclusions

In the present study we showed that standardized handwriting tasks can provide quantitative measures for the assessment of bradykinesia, micrographia and tremor. Several of these measures distinguished clinically diagnosed PD from HC.

Supporting Information

Figure S1 Illustration of the start and end areas in the circle and spiral task. A: Circle task; B: Spiral task. Green: start area; Red: end area.
(TIF)

Figure S2 Illustration of the star task segmentation method. A: the original x (blue) and y (green) coordinate time series recorded by the digitizer. B: the x and y coordinates as a function of the distance travelled by the pen tip. The function fitted to the coordinates is shown in red for two points, one of which is a point where the subject has drawn an acute angle and the other is slightly after such a point. C: turning angle estimated from parameters of the functions fitted to the coordinate series (green) and fitting error of the functions (blue); the local minima of the fitting error are shown as red circles in the angle series. D: the turning points detected by the algorithm are marked in the original time series by red squares.
(TIF)

Figure S3 Two samples of the 'e' and 'l' in the elel task. Left: A sample of text containing one 'e' and one 'l', including the recognized characteristic points (red dots). The numbered black arrows show the states of the state vector machine. Right: An example of a real detected letter 'e'. The light blue box indicates detected letters 'e'. The line color indicates the state of the algorithm; black: state 1, dark blue: state 2, light green/cyan: state 3, green: state 4, red: state 0/error. Markers indicate state changes; blue upward arrow indicates transition from state 1 to 2, blue leftward arrow indicates transition from state 2 to 3, blue downward arrow indicates transition from state 3 to 4, a green circle indicates a transition from state 4 to state 1 and a red cross indicates a transition from any state to state 0 (the points were an error is recognized).
(TIF)

Methods S1 Segmentation of the Digitizer data.
(DOCX)

Acknowledgments

A poster presentation concerning this research was presented at the Movement Disorders Congress in Dublin (2012) and at the 8th FENS Forum of Neuroscience in Barcelona (2012) and an oral presentation was given at the 4th Dutch Bio-Medical Engineering Conference in Egmond aan Zee (2013).

Author Contributions

Conceived and designed the experiments: RZ NM MG ES. Performed the experiments: ES. Analyzed the data: ES AT LC MG NM. Wrote the paper: ES NM AT LC MG BC RZ KL.

References

1. Massano J, Bhatia KP (2012) Clinical approach to parkinson's disease: Features, diagnosis, and principles of management. Cold Spring Harb Perspect Med 2: a008870.
2. Jankovic J (2008) Parkinson's disease: Clinical features and diagnosis. J Neurol Neurosurg Psychiatry 79: 368–376.
3. Hoehn MM, Yahr MD (1967) Parkinsonism: Onset, progression and mortality. Neurology 17: 427–442.
4. Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ (2002) Systematic evaluation of rating scales for impairment and disability in parkinson's disease. Mov Disord 17: 867–876.
5. Eshuis SA, Jager PL, Maguire RP, Jonkman S, Dierckx RA, et al (2009) Direct comparison of FP-CIT SPECT and F-DOPA PET in patients with parkinson's disease and healthy controls. Eur J Nucl Med Mol Imaging 36: 454–462.
6. Grosset DG, Schapira AH (2008) Timing the initiation of treatment in parkinson's disease. J Neurol Neurosurg Psychiatry 79: 615.
7. Bajaj NP, Wang L, Gontu V, Grosset DG, Bain PG (2012) Accuracy of subjective and objective handwriting assessment for differentiating parkinson's disease from tremulous subjects without evidence of dopaminergic deficits (SWEDDs): An FP-CIT-validated study. J Neurol 259: 2335–2340.
8. Alty JE, Jamieson S, Lones MA, Smith SL (2012) How slow is too slow? objective measurement of bradykinesia in parkinson's disease using novel non-invasive devices [abstract]. Movement Disorders 27 Suppl 1: 291.
9. Ünlü A, Brause R, Krakow K (2006) Handwriting analysis for diagnosis and prognosis of parkinson's disease. ISBMDA Lecture Notes in Computer Science 4345: 441–450.
10. Broderick MP, Van Gemmert AW, Shill HA, Stelmach GE (2009) Hypometria and bradykinesia during drawing movements in individuals with parkinson's disease. Exp Brain Res 197: 223–233.
11. Keresztenyi Z, Cesari P, Fazekas G, Laczko J (2009) The relation of hand and arm configuration variances while tracking geometric figures in parkinson's disease: Aspects for rehabilitation. Int J Rehabil Res 32: 53–63.

12. Van Gemmert AW, Teulings HL, Stelmach GE (2001) Parkinsonian patients reduce their stroke size with increased processing demands. Brain Cogn 47: 504–512.
13. Rosenblum S, Samuel M, Zlotnik S, Erikh I, Schlesinger I (2013) Handwriting as an objective tool for parkinson's disease diagnosis. J Neurol 260: 2357–2361.
14. Zietsma RC (2010) Designing a comprehensive system for analysis of handwriting biomechanics in relation to neuromotor control of handwriting. (Ph D thesis) Department of Bioengineering, University of Strathclyde, Glasgow: http://www.strath.ac.uk/library/.
15. Ponsen MM, Daffertshofer A, Wolters EC, Beek PJ, Berendse HW (2008) Impairment of complex upper limb motor function in de novo parkinson's disease. Parkinsonism Relat Disord 14: 199–204.
16. Dounskaia N, Van Gemmert AW, Leis BC, Stelmach GE (2009) Biased wrist and finger coordination in parkinsonian patients during performance of graphical tasks. Neuropsychologia 47: 2504–2514.
17. Teulings HL, Contreras-Vidal JL, Stelmach GE, Adler CH (1997) Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control. Exp Neurol 146: 159–170.
18. Saunders-Pullman R, Derby C, Stanley K, Floyd A, Bressman S, et al. (2008) Validity of spiral analysis in early parkinson's disease. Mov Disord 23: 531–537.
19. Stanley K, Hagenah J, Bruggemann N, Reetz K, Severt L, et al. (2010) Digitized spiral analysis is a promising early motor marker for parkinson disease. Parkinsonism Relat Disord 16: 233–234.
20. Scanlon BK, Levin BE, Nation DA, Katzen HL, Guevara-Salcedo A, et al. (2013) An accelerometry-based study of lower and upper limb tremor in parkinson's disease. J Clin Neurosci 20: 827–830.
21. Hughes AJ, Daniel SE, Kilford L, Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic parkinson's disease: A clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 55: 181–184.
22. Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, et al. (2004) Movement disorder society task force report on the hoehn and yahr staging scale: Status and recommendations. Mov Disord 19: 1020–1028.
23. Annett M (1970) A classification of hand preference by association analysis. Br J Psychol 61: 303–321.
24. Moroney R, Heida C, Geelen J (2008) Increased bradykinesia in parkinson's disease with increased movement complexity: Elbow flexion-extension movements. J Comput Neurosci 25: 501–519.
25. McLennan JE, Nakano K, Tyler HR, Schwab RS (1972) Micrographia in parkinson's disease. J Neurol Sci 15: 141–152.
26. Wagle Shukla A, Ounpraseuth S, Okun MS, Gray V, Schwankhaus J, et al. (2012) Micrographia and related deficits in parkinson's disease: A cross-sectional study. BMJ Open 2: 10.1136/bmjopen-2011-000628.
27. de Jong BM, Frackowiak RS, Willemsen AT, Paans AM (1999) The distribution of cerebral activity related to visuomotor coordination indicating perceptual and executional specialization. Brain Res Cogn Brain Res 8: 45–59.
28. Ondo WG, Satija P (2007) Withdrawal of visual feedback improves micrographia in parkinson's disease. Mov Disord 22: 2130–2131.

Measurement 114 (2018) 177–184

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Reproducibility of standardized fine motor control tasks and age effects in healthy adults

Esther J. Smits[a,*], Antti J. Tolonen[b], Luc Cluitmans[b], Mark van Gils[b], Rutger C. Zietsma[c], Marina A.J. Tijssen[a], Natasha M. Maurits[a]

[a] University Medical Center Groningen, University of Groningen, Department of Neurology, AB51, PO Box 30001, 9700 RB Groningen, The Netherlands
[b] VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 VTT, Tampere, Finland
[c] Manus Neurodynamica Ltd, 4a John Street, Cullercoats, North Shields, Tyne and Wear NE30 4PL, United Kingdom

ARTICLE INFO

Keywords:
Fine motor control
Drawing
Handwriting
Movement time
Accuracy
Reproducibility
Intraclass correlation

ABSTRACT

Graphical tasks can provide objective measures of important motor symptoms of movement disorders such as Parkinson's disease (PD). These tasks could potentially be useful in clinical settings for (early) diagnosis and monitoring of such diseases. However, before such tasks can be used clinically, reproducibility needs to be investigated. The present study assesses the reproducibility of these graphical tasks including age-effects in healthy adults. Overall, performance on circle, spiral and zigzag tracing tasks and a writing task showed good reproducibility (intraclass correlation coefficients (ICC) > 0.7). Reproducibility was similar to the reproducibility of the Purdue pegboard task, which is an already validated fine motor control task. Reproducibility for the modified Fitts' task was moderate (ICC = 0.6). Reproducibility was higher in older participants compared to younger participants. To conclude, performance on graphical tasks, especially tracing and writing tasks, was reproducible in healthy adults, which is essential for future diagnostic and monitoring purposes in patients.

1. Introduction

Despite the increased use of computers, the use of a pen for handwriting and drawing is still an important skill in daily life that everyone is expected to master. Holding a pen and performing handwriting and drawing is one of the most complex fine motor functions of humans [1], involving a cooperation between the central nervous system (CNS) and the musculoskeletal system [2]. Therefore, deficits in brain function or in the musculoskeletal system due to a disease, such as movement disorders [3] or trauma could cause deterioration in handwriting and drawing ability [2]. Even though handwriting and drawing are complex functions, these graphical tasks entail overlearned skills [4]. Therefore, once mastered, performance on such tasks is expected to not considerably improve or deteriorate over time anymore [4]. Because of this expected stability in performance, graphical tasks are interesting to study to gain more insight into the changes in motor control due to a movement disorder, such as Parkinson's disease (PD) [5–7], to evaluate treatment effects [8,9] or to study fine motor control in general [1,10–14]. A system to record graphical tasks, like handwriting and drawing, was developed in a European project to aid in the diagnostic process of PD (the DiPAR project: funded by the EC under the FP7-SME-201001 programme, grant agreement 262291). This system consists of a pen and tablet and custom software, based on a concept by Manus Neurodynamica Ltd. In a previous study we showed that a set of standardized graphical tasks, recorded with this newly developed system, could provide objective measures of important motor symptoms of PD and allowed distinguishing between PD patients and gender and age-matched healthy control participants [15]. In addition, we showed that performance on these tasks improved after taking dopaminergic medication in PD patients [16], indicating validity/responsiveness of the tasks. Before graphical tasks such as those used in the newly developed system can actually be used clinically for diagnosis, screening or monitoring, their characteristics and added value should be assessed [17]. According to Van den Bruel et al. [17], several steps should be followed in this process. Besides assessment of validity, another important step is to examine the reproducibility of the results, defined as the ability to achieve the same test results on repeated testing [17]. Therefore, the goal of the present study was to investigate the reproducibility of this set of graphical tasks, executed with the newly developed system, with a one-week interval in healthy participants of

Abbreviations: PD, Parkinson's disease; HC, Healthy Control; MMSE, Mini Mental State Examination; MT, Movement Time; ICC, Intraclass correlation

* Corresponding author.
E-mail addresses: Smits.E.J@gmail.com (E.J. Smits), Antti.Tolonen@vtt.fi (A.J. Tolonen), Luc.Cluitmans@vtt.fi (L. Cluitmans), Mark.vanGils@vtt.fi (M. van Gils), R.C.Zietsma@manusneuro.com (R.C. Zietsma), M.A.J.de.Koning-Tijssen@umcg.nl (M.A.J. Tijssen), N.M.Maurits@umcg.nl (N.M. Maurits).

http://dx.doi.org/10.1016/j.measurement.2017.09.011
Received 29 May 2017; Received in revised form 27 July 2017; Accepted 11 September 2017
Available online 19 September 2017
0263-2241/

different ages.

Previously, reproducibility of similar graphical tasks has been investigated [14,18,19]. However, the scope of these previous studies was limited. In addition, reproducibility needs to be investigated for each newly developed system. Mergl et al. [14] investigated reproducibility in young adults (n = 21) only and their measures focused on movement speed. Erasmus et al. [18] only investigated reproducibility of drawing precision between two consecutive days. Finally, Feys et al. [19] focused on tremor measures and only investigated short-term test-retest reliability for a spiral drawing task in multiple sclerosis patients with tremor. Additionally, in the present study intraclass correlation coefficients (ICC) were used to determine the reproducibility, instead of Pearson or Spearman correlations which were used in the earlier studies. For the ICC, the data are centered and scaled using a pooled mean and standard deviation, whereas for the Pearson or Spearman correlation coefficient, each variable is centered and scaled by its own mean and standard deviation. Because measurements on repeated testing are of the same quantity and unit, the ICC is a better measure to examine reproducibility than the Pearson or Spearman correlation coefficient. Furthermore, since some movement disorders are typically diagnosed in specific age groups - e.g., PD is typically diagnosed in persons older than 60 years - in this study also the influence of age on reproducibility of these tasks was examined.

The graphical tasks which appeared to be the most useful (in terms of their ability to distinguish between PD patients and controls and their validity) in our previous studies [15,16] were included in the present study. This set of tasks consisted of circle, spiral and zigzag tracing tasks, an *'elelelel'* writing task and a modified Fitts' task. These tasks are easy to perform and cover a large range of upper limb functioning. The modified Fitts' task was included to assess the speed-accuracy trade off, which may be impaired in PD patients [20]. We explored whether these tasks, performed with the newly developed system, show good reproducibility in healthy adults. To show that this set of graphical tasks is able to serve its intended goal, it is important to compare its performance to that of an existing test of fine motor control [17]. Therefore, the reproducibility on the graphical tasks was compared to the reproducibility of the Purdue pegboard test, since both tests measure aspects of fine motor control. Over the years, the Purdue pegboard test has been used in neuropsychological assessments and rehabilitation contexts [21] and has been shown to be reliable [21–23].

To summarize, the present study investigated the reproducibility of a set of graphical tasks employing a newly developed system. The reproducibility of these graphical tasks was compared to the reproducibility of an independent measure for manual dexterity, the Purdue pegboard task. Additionally, the influence of age on performance of these tasks was investigated.

2. Methods

2.1. Participants

Thirty-six healthy volunteers, recruited from the general population, participated in this study. The handwriting and drawing tasks in this study entail overlearned skills, and once mastered, performance on such tasks is expected to not considerably change over time. Given the small expected variability in the healthy population, the sample size was considered adequate. The only inclusion criteria were perceived health and being 18 years or older. After data collection was completed, the participants were divided into three age-groups to investigate the effect of age on reproducibility. The age ranges of the groups were chosen to generate three equally-sized groups of 12 participants. The young group consisted of participants aged 20–29 years (mean age 26.3, sd 2.5, 7 males), the middle-aged group was aged 30–55 years (mean age 42.0, sd 6.4, 8 males) and participants in the older group were aged 56–75 years (mean age 64.7, sd 6.2, 8 males). All participants provided informed consent and completed the tasks twice with one week in between. Exclusion criteria were a history of epileptic seizures, head injury, neurological or motor disorders, the use of medication affecting movement, or a low (< 26) score on the Mini Mental State Examination (MMSE). The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen.

2.2. Experimental design

Participants were seated in front of a table in a comfortable position to write. A tablet computer (ASUS Eee Slate EP121) and a newly developed digital pen with custom software were used. The position of the pen-tip on the tablet during movement was recorded at a sampling frequency of 200 Hz. The pen was not inking and had a wired connection to the tablet. The pen has a length of 170 mm and a barrel width of 20 mm. The pen-tip was a custom-made resonant circuit which communicates with the Eee Slate, similar to the original Eee Slate stylus. The size of the pen-tip was similar to the tip of a normal inking pen to resemble writing on paper. The spatial resolution of the pen was 0.01 mm and the temporal resolution 0.005 s. The system made use of a "Wacom enabled" touchscreen that incorporates an electromagnetic array in the screen for highly accurate tracking of a stylus. The Wacom technology along with appropriate Linux device drivers and a proprietary software implementation allowed to monitor the variability of the sampling rate of pen-tip tracking and subsequently oversample by "latching" into 200 Hz. The resulting constant temporal accuracy was validated by comparing event markers for contact between pen-tip and tablet, for concurrent recording from the touchscreen with those from the proprietary digital pen, over the full duration of trials. Participants performed eight tasks (see Section 2.3.) with the digitizer pen on the tablet using their dominant hand. Additionally, participants performed the Purdue pegboard test. The examiner was seated behind an operator computer to start and stop the recordings. The complete experiment lasted approximately thirty minutes and participants were allowed to have a break in between tasks. The participants were allowed to perform a few practice trials to get used to the system.

2.3. Tasks

Each participant performed several graphical tasks in the same order to limit variability in task results. In addition, task order was maintained because the newly developed system is intended to be used in clinical practice, where a fixed task order will be used. Participants were instructed to start the task at a signal of the examiner and to perform the tasks at a comfortable speed, allowing them to move as smoothly as individually possible. The newly developed system might be used in the future in home-based settings for testing and monitoring and therefore the instructions were kept as simple as possible. In addition, in a home-based situation it is difficult to implement and verify an unnatural way of writing and thus the participants were not instructed to keep their arm in a specific orientation, to not interfere with their natural way of writing. The participants first traced geometric shapes; a circle, a spiral and a zigzag figure which were displayed on the tablet (see Fig. 1). The circle and spiral were traced ten times in a clockwise direction, starting from the 12 o'clock position (circle) or from inside to outside (spiral). The zigzag was traced five times, from left to right and back. During the tracing tasks the participants did not receive visual feedback on the screen of the tablet, since we intended for them not to be distracted by the traces of previous trials. In addition, since the goal was to perform the tasks as smoothly as individually possible, the participants were instructed to trace the figures, but were not explicitly told to trace the figures as accurate as possible.

The next task consisted of writing '*elelelel*' five times with each phrase starting at the left side of the tablet. During this task the participants were provided with visual feedback on the screen to resemble natural writing on paper. An example of the *'elelelel'* sentence was

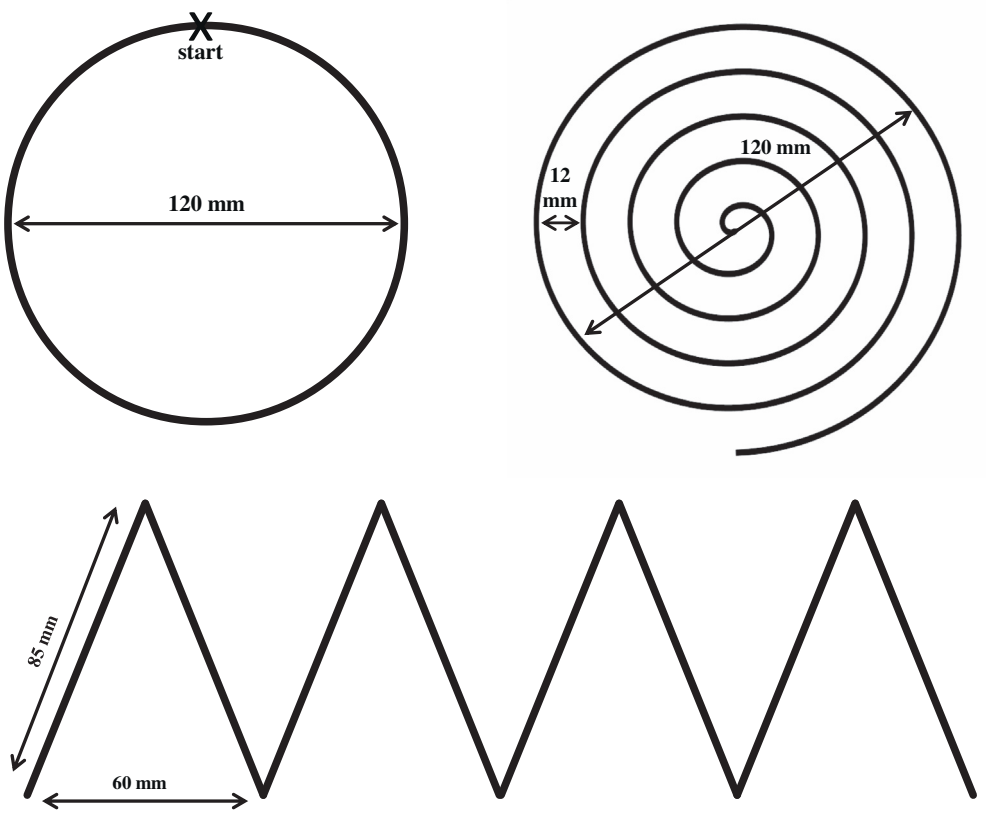


Fig. 1. Templates and their dimensions for the tracing and drawing tasks: a circle, spiral and zigzag figure.

printed on paper and placed on the table above the tablet. Thereafter, a modified Fitts' task was performed, which was similar to Fitts' original task [24] but adapted to the dimensions of our system. Participants were asked to tap into two targets (filled circles, placed on an imaginary horizontal line in the middle of the tablet) alternately with the pen-tip as fast and as accurately as possible during 20 s. In eight subtasks the difficulty of the tasks was altered by varying the distance between targets and varying the diameter of the targets. The varying distances and diameters were chosen according to the dimensions, to allow determination of the relationship between movement time and difficulty of tasks (see Section 2.4). In the first four subtasks (1–4), the distance between the center of the targets was kept constant at 7 cm, while the diameter of the targets was increased (0.7, 1.3, 1.9, 2.5 cm). In another four subtasks (5–8), the distance between the center of the targets was kept constant at 20 cm, while the diameter of the targets was increased (0.7, 1.3, 1.9, 2.5 cm). Finally, the Purdue pegboard test (PPT) was performed, that employed a board, pins, collars and washers. The board contains two vertically oriented parallel rows with 25 holes in each row and the pins, collars and washers are located in reservoirs at the top of the board. Four subtests were performed in an order according to the instructions [21]. In the first three subtests the participant was instructed to place as many pins as possible in the holes within 30 s, first with the dominant hand, then with the other hand and finally with both hands simultaneously. In the last subtest (assembly) the participant used alternate hands to make as many assemblies as possible within 60 s. An assembly consisted of a pin, washer, collar and a second washer. In accordance with the instructions, the participants were allowed to practice before each subtest [21].

2.4. Data analysis

The drawing and tracing tasks were analyzed using custom made scripts in Matlab 7.4.0 (R2007a). Since movement time (MT) was an important measure of speed to distinguish PD patients from HC participants [15], in the present study we also calculated mean MT per repetition for the circle, spiral, and zigzag tracing tasks. MT in seconds was calculated by dividing the total number of samples for each repetition by 200 (sampling frequency was 200 Hz). Mean MT was then calculated as the average MT over all repetitions per participant. We also calculated the mean deviation (in mm) from the template for the circle, spiral and zigzag tracing tasks (mean error) as a measure of accuracy. For each sample the pen-tip position (x and y coordinates) was compared to the x and y coordinate of the template and the Euclidean distance was calculated in mm. The deviation from the template was calculated for each sample and repetition of the circle, spiral and zigzag, and then averaged over all samples and repetitions for each participant to derive the mean error for each task. The pen-tip position data (x and y coordinates) of the *'elelelel'* writing task were preprocessed to detect the letters 'e' and 'l'. For each letter MT was calculated and then averaged over all 'e's' and 'l's', separately (see Appendix A). Similarly, the mean width and height of the letters was calculated.

The modified Fitts' task was analyzed according to Fitts' law [24]. The tradeoff between speed and accuracy was modeled by Fitts [24] in the time required for movement (T):

$$T = a + b\left(\log\frac{2A}{D}\right)$$

Here, A is the distance between targets and D the target diameter. The part log(2A/D) is known as the index of difficulty (ID). When multiple IDs are available, a and b can be estimated by linear regression. In our modified Fitts' task eight IDs could be determined, since the task consists of eight subtasks, with varying difficulty. For each participant the mean T for each ID (each subtask) was calculated as the average time needed to move the pen from one target to the other, to allow determination of the relationship between movement time and ID. A linear curve was then fitted to the data points and a least squares calculation was used to determine the goodness of fit (R2). The R2 refers to

the degree of compliance with Fitts' law and was determined for each participant. The slope of the fitted curves describes the extent to which the performance becomes slower with an increase in ID and was calculated for each participant, as well. These calculations resulted in two measures: FittsSlope and FittsR2.

The results of the Purdue pegboard test were analyzed in accordance with the instructions [21]. The score on the first two subtests was equal to the number of pins inserted in the holes within 30 s. The score on the third subtest equaled the number of pairs of pins inserted in the holes and the assembly score equaled the sum of the number of assembled parts. Also a sum score was computed by adding the scores obtained in the first three subtests (right hand + left hand + both hands).

All measures were determined for the first and second measurement day.

2.5. Statistical analysis

2.5.1. Reproducibility

Statistical analyses were conducted using SPSS 20.0.0.1. Since the goal was to investigate the reproducibility of the tasks, the scores on the first and second measurement day were compared for each task and subtask. The handwriting system was primarily designed for diagnostic purposes and therefore, the relative reliability for all tasks was determined by the intraclass correlation coefficient (ICC). This was done for the whole group as well as for each age-group separately. In this study we used the two-way random ICC with absolute agreement, to take into account systematic and random errors [25]. The ICC ranges from 0 to 1. According to Andresen [26] an ICC between 0 and 0.40 signifies poor reliability, between 0.40 and 0.74 moderate reliability and an ICC between 0.75 and 1.00 signifies excellent reliability.

2.5.2. Differences between measurement days and age-groups

Differences between measurement days and age-groups were tested with a mixed factors ANOVA, with between subjects factor *group* (3 levels; younger, middle-aged and older) and within subjects factor *time* (2 levels; measurement day 1 and day 2). Even though not all variables were normally distributed an ANOVA was performed as ANOVAs are in general quite robust for distribution violations and especially since the data on day 1 was assumed to have the same distribution as on day 2, this test was chosen. A Bonferroni correction was applied to correct for multiple comparisons, which resulted in an alpha of 0.0036 (0.05/14) for the graphical tasks and an alpha of 0.01 (0.05/5) for the pegboard tasks.

3. Results

All participants (n = 36; mean age: 44.3; sd: 16.8; 23 male; 13 female) completed each of the tracing and writing tasks, the modified Fitts' task and the Purdue pegboard test twice with exactly one week in between. The two measurements were performed at approximately the same time, but at least within a range of three hours on both days. None of the participants indicated that they were fatigued during the tasks or that they needed a break in between tasks.

3.1. Reproducibility

Mean MT and mean error for each of the tracing and drawing tasks are given in Table 1 for the total group and for each of the age groups. Agreement between the first and second measurement for the total group was moderate to excellent for mean MT on the circle, spiral and zigzag tasks (circle tracing: ICC = 0.69; spiral tracing: ICC = 0.77; zigzag tracing: ICC = 0.89). Agreement for the total group between the first and second measurement for mean error was excellent for spiral and zigzag tracing (ICC = 0.85 and ICC = 0.82) and moderate for circle tracing (ICC = 0.47).

Table 1
Statistical results for the tracing tasks. Intra Class Correlation (ICC) coefficients and mean movement time per repetition are displayed for the whole group as well as for the three groups separately (mean(sd)). The results of the mixed factors ANOVA (F-value and p-value) are also shown in this table.

	Day 1 (mean (sd))	Day 2 (mean (sd))	ICC	F	p
Circle mean MT (s)					
Group I (20–29 years)	3.6 (1.3)	2.3 (0.8)	0.21		
Group II (30–55 years)	3.1 (1.2)	2.3 (1.1)	0.76		
Group III (56–75 years)	4.5 (2.6)	3.7 (2.4)	0.78		
Total	3.7 (1.9)	2.8 (1.7)	0.69	22.96	0.00
Spiral mean MT (s)					
Group I (20–29 years)	7.4 (1.9)	6.0 (1.2)	0.41		
Group II (30–55 years)	7.5 (2.4)	5.9 (1.7)	0.61		
Group III (56–75 years)	10.8 (6.6)	9.1 (4.4)	0.81		
Total	8.6 (4.4)	7.0 (3.1)	0.77	18.91	0.00
ZigZag mean MT (s)					
Group I (20–29 years)	7.4 (1.6)	6.6 (1.9)	0.66		
Group II (30–55 years)	7.4 (2.2)	5.9 (1.7)	0.62		
Group III (56–75 years)	11.8 (7.1)	10.1 (6.1)	0.91		
Total	8.9 (4.8)	7.5 (4.1)	0.89	22.45	0.00
Circle mean Error (mm)					
Group I (20–29 years)	2.1 (0.5)	2.5 (1.0)	0.26		
Group II (30–55 years)	2.3 (0.6)	2.6 (1.1)	0.54		
Group III (56–75 years)	2.7 (1.5)	2.7 (1.8)	0.51		
Total	2.4 (1.0)	2.6 (1.3)	0.47	1.10	0.30
Spiral mean Error (mm)					
Group I (20–29 years)	2.5 (0.4)	2.6 (0.6)	0.73		
Group II (30–55 years)	2.5 (0.5)	2.6 (0.6)	0.85		
Group III (56–75 years)	2.6 (0.7)	2.7 (0.9)	0.92		
Total	2.5 (0.6)	2.6 (0.7)	0.85	6.05	0.02
ZigZag mean Error (mm)					
Group I (20–29 years)	2.1 (0.4)	2.4 (0.5)	0.59		
Group II (30–55 years)	2.4 (0.8)	2.8 (0.8)	0.89		
Group III (56–75 years)	1.8 (0.7)	2.0 (0.8)	0.77		
Total	2.1 (0.7)	2.4 (0.8)	0.82	18.69	0.00

Table 2
Statistical results for the modified Fitts task. IntraClass Correlation coefficients (ICC) and descriptive values for both measures of the Fitts' task (FittsSlope and FittsR2) are displayed for the whole group as well as for the three groups separately (mean(sd)). The results of the mixed factors ANOVA are also shown (F-value and p-value).

	Day1 (mean (sd))	Day2 (mean (sd))	ICC	F	p
FittsSlope[a]					
Group I (20–29 years)	0.09 (0.02)	0.09 (0.02)	0.53		
Group II (30–55 years)	0.09 (0.03)	0.08 (0.03)	0.62		
Group III (56–75 years)	0.11 (0.02)	0.09 (0.04)	0.58		
Total	0.10 (0.03)	0.09 (0.03)	0.58	5.06	0.03
FittsR2[b]					
Group I (20–29 years)	0.91 (0.12)	0.93 (0.05)	0.05		
Group II (30–55 years)	0.94 (0.04)	0.90 (0.09)	0.33		
Group III (56–75 years)	0.93 (0.05)	0.88 (0.15)	0.00		
Total	0.93 (0.08)	0.90 (0.10)	0.02	1.17	0.29

[a] FittsSlope represents the extent to which performance becomes slower with an increase in difficulty of the task (scores range from 0.02 to 0.14, where a lower score means that performance becomes less slow with an increase in difficulty compared to a higher score, i.e. lower scores indicate better performance).

[b] FittsR2 represents the degree of compliance with Fitts' law (scores range from 0.55 to 0.99, where a higher score indicates better compliance with Fitts' law, i.e. better performance, than a lower score).

Table 3
Statistical results for the elel writing task. ICC coefficients and descriptive values for performance on the elel task are displayed for the whole group as well as for the three groups separately (mean(sd)). The results of the statistical analysis for differences between measurement days are also shown (mixed factors ANOVA (F-value and p-value)).

	Day1 (mean (sd))	Day2 (mean (sd))	ICC	F	p
Mean MT (s) letter 'e'					
Group I (20–29 years)	0.32 (0.06)	0.26 (0.03)	0.32		
Group II (30–55 years)	0.29 (0.06)	0.24 (0.05)	0.56		
Group III (56–75 years)	0.35 (0.15)	0.32 (0.13)	0.94		
Total	0.32 (0.10)	0.27 (0.09)	0.81	46.0	0.00
Mean MT (s) letter 'l'					
Group I (20–29 years)	0.41 (0.09)	0.35 (0.07)	0.59		
Group II (30–55 years)	0.41 (0.09)	0.36 (0.07)	0.67		
Group III (56–75 years)	0.49 (0.19)	0.46 (0.17)	0.90		
Total	0.44 (0.13)	0.39 (0.12)	0.82	23.2	0.00
Width letter 'e' (mm)					
Group I (20–29 years)	11.69 (3.03)	13.14 (3.50)	0.27		
Group II (30–55 years)	12.19 (3.88)	11.24 (2.86)	0.78		
Group III (56–75 years)	8.80 (4.11)	8.92 (3.68)	0.93		
Total	10.89 (3.90)	11.10 (3.71)	0.73	0.21	0.65
Height letter 'e' (mm)					
Group I (20–29 years)	23.53 (10.12)	24.92 (10.50)	0.89		
Group II (30–55 years)	21.06 (9.21)	20.15 (7.78)	0.93		
Group III (56–75 years)	16.40 (7.43)	18.13 (8.65)	0.87		
Total	20.33 (9.23)	21.07 (9.25)	0.90	0.21	0.65
Width letter 'l' (mm)					
Group I (20–29 years)	18.33 (5.36)	21.82 (7.70)	0.30		
Group II (30–55 years)	19.73 (5.59)	19.70 (4.56)	0.83		
Group III (56–75 years)	13.01 (6.53)	14.13 (6.84)	0.92		
Total	17.03 (6.39)	18.55 (7.12)	0.70	3.31	0.08
Height letter 'l' (mm)					
Group I (20–29 years)	48.07 (20.30)	51.42 (22.82)	0.88		
Group II (30–55 years)	50.91 (16.69)	51.54 (15.75)	0.91		
Group III (56–75 years)	38.62 (17.75)	43.15 (20.37)	0.90		
Total	45.87 (18.56)	48.70 (19.70)	0.89	3.31	0.08

The descriptive values for FittsSlope and FittsR2 are shown in Table 2, for the total group, as well as for each age group separately. Agreement for the total group between the two measurement days for FittsSlope was moderate (ICC = 0.58) and poor for FittsR2 (ICC = 0.02, see Table 2).

The descriptive values for mean MT, width and height of the letters 'e' and 'l' for the *'elelelel'* task are shown in Table 3 for the total group, as well as for each age group separately. Agreement between the two measurement days was moderate for the width of the letter 'e' and 'l' (ICC = 0.73 and ICC = 0.70, respectively) and excellent for mean MT of the letters 'e' and 'l' (ICC = 0.81 and ICC = 0.82, respectively) and height of the letters 'e' and 'l' (ICC = 0.90 and ICC = 0.89, respectively).

Mean values for the scores on the Purdue pegboard task are shown in Table 4. Agreement between the first and second measurement was excellent for the both hands score, the sum score, and the assembly score (ICC = 0.77, ICC = 0.78 and ICC = 0.90, respectively) (see Table 4). The scores for the right hand and left hand resulted in a moderate ICC (ICC = 0.50 and ICC = 0.71, respectively) (see Table 4).

3.2. Differences between measurement days and age-groups

The mixed factors ANOVA showed that there was a significant difference between measurement days for mean MT on the circle, spiral and zigzag tracing tasks (all $p < 0.0036$, see Table 1). Mean MT on all tracing tasks was significantly lower on day 2 compared to day 1. Mean Error on the circle and spiral tracing task was not significantly different between the two measurement days. Mean Error on the zigzag tracing task was higher on the second measurement day compared to the first ($p < 0.0036$, see Table 1). FittsSlope and FittsR2 were not significantly different between the two measurement days. Mean MT of the letters 'e' and 'l' for the *'elelelel'* task was significantly lower on the second measurement day compared to the first ($p < 0.0036$, see Table 3). The width and height of the letter 'e' and 'l' were not significantly different between the two measurement days (see Table 3). There were no significant differences between age-groups on all of the graphical tasks.

Table 4
Statistical results for the Purdue pegboard task (PPT). ICC coefficients and descriptive values for performance on the PPT task are displayed for the whole group as well as for the three groups separately (mean(sd)). The results of the statistical analysis for differences between measurement days are also shown (mixed factors ANOVA (F-value and p-value)).

	Day1 (mean (sd))	Day2 (mean (sd))	ICC	F	p
Right hand score					
Group I (20–29 years)	14.92 (1.78)	15.92 (1.24)	0.23		
Group II (30–55 years)	14.42 (2.02)	15.08 (1.93)	0.77		
Group III (56–75 years)	13.67 (1.30)	14.50 (1.45)	0.17		
Total	14.33 (1.76)	15.17 (1.63)	0.50	9.34	0.00
Left hand score					
Group I (20–29 years)	13.17 (1.03)	13.92 (1.08)	0.48		
Group II (30–55 years)	13.92 (2.11)	14.92 (1.93)	0.84		
Group III (56–75 years)	12.83 (1.34)	13.50 (1.68)	0.56		
Total	13.31 (1.58)	14.11 (1.67)	0.71	20.9	0.00
Both hands score					
Group I (20–29 years)	11.17 (1.27)	11.67 (0.89)	0.41		
Group II (30–55 years)	11.58 (2.15)	11.50 (2.20)	0.88		
Group III (56–75 years)	9.83 (1.27)	10.25 (1.29)	0.57		
Total	10.86 (1.74)	11.14 (1.64)	0.77	2.14	0.15
Sum Score					
Group I (20–29 years)	39.25 (3.05)	41.50 (2.68)	0.46		
Group II (30–55 years)	39.92 (5.63)	41.50 (5.57)	0.90		
Group III (56–75 years)	36.33 (2.57)	38.25 (3.52)	0.65		
Total	38.50 (4.18)	40.42 (4.28)	0.78	24.8	0.00
Assembly score					
Group I (20–29 years)	37.67 (4.44)	39.42 (4.12)	0.69		
Group II (30–55 years)	34.25 (8.36)	34.92 (7.22)	0.94		
Group III (56–75 years)	29.17 (4.04)	29.83 (4.80)	0.80		
Total	33.69 (6.77)	34.72 (6.69)	0.90	4.52	0.04

The right hand score, left hand score and the sum score on the Purdue pegboard test were significantly higher on the second measurement day compared to the first, according to the mixed factors ANOVA ($p < 0.01$, see Table 4). The both hands score and assembly score were not significantly different between the two measurement days. The assembly score was significantly lower in the older age group compared to the younger age group ($p < 0.01$, see Table 4). The other scores on the Purdue pegboard test were not significantly different between age groups.

4. Discussion

The aim of this study was to investigate the reproducibility of a set of graphical tasks using a newly developed system, consisting of a digital pen and tablet. Overall, the performance measures derived for the tasks showed moderate to excellent test-retest reliability. Additionally, this study showed that in general test-retest reliability increased with age.

We showed that reproducibility on this set of graphical tasks, which

measures aspects of fine motor control, was similar to the reproducibility of an already validated fine motor control task, the Purdue pegboard test. This suggests that the set of graphical tasks studied here provides a reliable method to measure aspects of fine motor control. Mean MT for the tracing and writing tasks and mean error per repetition for the tracing tasks were well reproducible, in line with previous studies which also reported high test-retest reliability for MT on a circle drawing task and some handwriting tasks [14] and mean drawing error on a tracing task [18], although Pearson or Spearman correlation coefficients were used in those studies to assess reproducibility [14]. In our study, participants were significantly faster on a few tasks on the second measurement day compared to the first, which suggests a learning effect. According to Longstaff and Heath [4] handwriting and drawing are overlearned skills and are not expected to considerably improve or deteriorate over time. However, a possible learning effect could be stronger in simple tasks compared to more complex tasks. It is, for example, easier to increase speed on a simple circle tracing task than on a spiral tracing task, as the spiral tracing task requires more accuracy. In the present study, the more complex tracing tasks – the spiral and zigzag task - indeed showed better reproducibility than the simpler circle tracing task. This confirms that there might be a smaller learning effect for the spiral and zigzag tasks. In addition, the mean error for the spiral and zigzag tracing tasks showed higher reproducibility than the mean error for the circle tracing task. Similar results were found for the Purdue pegboard test, for which reproducibility was also better on the complex task than on the simple tasks. This finding suggests that complex tasks are more reliable than simple tasks to assess fine motor control, since learning effects between two measurements are smaller. However, the tasks were performed in a fixed order and 'getting used to the system' could also be a reason for the circle task being less reproducible than the spiral and zigzag tasks. However, since the participants were allowed to perform a few practice trials before the actual experiment and indicated not to need any more practice trials at the second measurement, we believe that 'getting used to the system' is not the reason for the circle task being less reproducible. For the Purdue pegboard task it was also allowed to practice before the actual measurement started.

A seemingly conflicting finding in this study is the moderate to excellent reproducibility as expressed by the moderate to high ICC values while performance on some of the tasks was significantly different between measurement days. This could be explained by the fact that the differences were in the same direction for almost all participants, i.e. all improved a bit on the second day. In such cases, when all participants behave similarly, even small mean differences can be statistically different [25], but might be less relevant. However, these small differences could indicate a learning effect and should be investigated further in clinical populations if the system would be used for monitoring purposes.

The modified Fitts' task was analyzed according to Fitts' law, deriving two measures 'FittsSlope' and 'FittsR2'. FittsSlope, representing the extent to which performance becomes slower with an increase in difficulty of the task, showed moderate reproducibility. FittsR2 represents the degree of compliance with Fitts' law and showed poor reproducibility. This suggests that FittsSlope is a better measure for performance on the modified Fitts' task than FittsR2. However, the low ICC on FittsR2 could also be explained by the fact that all scores are very close to the maximum value of 1.00. A very small difference between the measurement days would have great impact on the ICC, which might make the ICC less suitable to investigate reproducibility for FittsR2. Furthermore, the modified Fitts' task could be improved by a larger range of difficulties of the different subtasks, which could lead to an improvement in the reproducibility. However, the modified Fitts' task was already adapted to the size of the tablet and adding more varying subtasks, which are significantly more or less difficult than the current subtasks, is not possible with the current system. More difficult subtasks could for example be generated by creating a larger distance between the targets, but this would only be possible by using a larger tablet than the current tablet. This indicates that the modified Fitts' task, performed with the current system, is less suited for diagnostic and monitoring purposes in patients with movement disorders.

Movement time on all tasks increased with age, consistent with previous studies [14,27]. However, the results of the older group were generally more reproducible than the results of the other two groups, which suggests that performance (in terms of speed) in the older group is more stable over time. An explanation may be that older adults payed more attention to performing the tasks correctly while the younger and middle-aged groups were more focused on finishing the tasks quickly, which may cause a larger learning effect in the latter groups. High reproducibility of the speed at which fine motor control tasks are executed in the older group indicates that these tasks might be particularly suited for application in movement disorders such as PD, which is typically diagnosed in people older than 60 years, but less suited in movement disorders which are diagnosed at very young or across all ages. On the contrary, mean error for the spiral and zigzag tracing tasks and writing size for the '*elelelel*' tasks were highly reproducible for all ages suggesting that drawing error and writing size might be more suited for diagnostic or monitoring applications in movement disorders that occur across all ages.

Graphical tasks have previously been proposed as an aid in the diagnostic work-up of movement disorders, since differences in performance on these tasks have been found between patients with movement disorders and healthy controls [5–7,15,28–30] and between patients with different movement disorders [31]. However, none of these studies included reproducibility testing, which is a necessary step before introducing a test in clinical practice [17]. In the present study we showed high reproducibility of several graphical tasks executed with a newly developed system. Previously we already showed differences between PD patients and HC participants based on these graphical tasks [15] and that performance on these tasks improved after taking dopaminergic medication in PD patients [16]. Further testing of the set of graphical tasks and the system used in the present study is still needed with additional analyses, for example to show that PD patients can be distinguished from patients with other movement disorders, such as essential tremor. To investigate whether these graphical tasks are suitable for long term monitoring a longitudinal study should be performed in which PD patients will be followed for a longer time-period. Additionally, performance on such graphical tasks should be validated against current gold standards in movement disorders, such as the Unified Parkinson's Disease Rating Scale.

5. Conclusions

To conclude, this study shows that a set of graphical tasks, which measures fine motor control, has moderate to high reproducibility and that reproducibility is similar to the reproducibility of another fine motor control task, the Purdue pegboard task. The modified Fitts' task seems less suitable for clinical testing with the newly developed system since the measures in the current setup only showed poor to moderate reproducibility. We propose that more complex tracing tasks, such as spiral and zigzag tracing and a letter writing task are more suitable for clinical testing, because such tasks provide measures which are more reliable than the measures provided by the simpler tasks.

Acknowledgements

This study is part of the DiPAR project for which funding was received from the EC in the FP7-SME-201001 programme (grant agreement 262291). The hardware of the measurement setup was provided by Fraunhofer IPMS (Dresden, Germany) and the data acquisition software by Fraunhofer IPA (Stuttgart, Germany), based on a concept by Manus Neurodynamica Ltd (patent WO/2011/141734). We thank Florian Dennerlein for his contribution to this manuscript.

Competing interests

One of the authors, RCZ, holds a directorship at Neurodynamica Ltd, UK. The other authors have no competing interests to declare.

Appendix A

Elelelel writing task: Letter shape recognition analysis

To calculate the mean width and height of each letter in the '*elelelel*' writing task, the pen tip position data (x and y coordinates) were pre-processed. First, the data were split into separate segments, where each segment represented one line of text. This was done using an 'in range' signal, which indicates whether or not the pen is in detection range of the tablet employing that, after writing one line of text, the patient lifts the pen so that it is outside the detection range of the tablet. Subsequently, the segments corresponding to an 'e' or an 'l' were identified. The shapes in each line were recognized by using a state machine that employs the direction of change of the pen tip position as input (similar to the method used in Smits et al. [15]).

The data was processed according to the following steps.

1. The direction of change (Δx, Δy) of the pen tip position was approximated by dividing the difference between samples that are 20 samples apart by 20. A sample distance of 20 rather than 1 is used as a data smoothing method thereby filtering irregularities in the input signal. The distance of 20 samples corresponds to a time span of 100 ms, because the signals were sampled at 200 Hz.
2. The signs of Δx and Δy were used to drive a state machine. For a 'perfect signal', the state's cycle through the following states in order (see also Fig. A1):
   - State 1: $\Delta x > 0$, $\Delta y > 0$: the pen is moving right and up, from the start of the curve toward the rightmost point.
   - State 2: $\Delta x < 0$, $\Delta y > 0$: the pen is moving further up but leftward, from the rightmost point to the top.
   - State 3: $\Delta x < 0$, $\Delta y < 0$: the pen is moving further left but downward, from the top to the leftmost point.
   - State 4: $\Delta x > 0$, $\Delta y < 0$: the pen is moving further down but rightward again, from the leftmost point to the bottom.

Since not all signals were perfect, the actual state machine was designed to detect errors and correct for these imperfections. Several additions were implemented:

- During normal operation the state can only change from state N to state N + 1 (or from state4 to state 1). For each of these state changes there is only one component that changes, and that is the only change the algorithm looks for. For example, in state 1 the algorithm only searches for a time point when Δx becomes negative, and then the state changes to state 2.
- If the algorithm would try to go back one state (moving in the wrong direction) it stays in the current state; if it would try to go back yet another state (both x and y going in the wrong direction) an error state is entered. If it recovers from the initial "wrong direction" it updates the starting point of the current state.
- A fifth state, state 0, was included which indicates an error or initial state. When in this state, the next state (1, 2, 3 or 4) is selected based on the signs of Δx and Δy directly (the state machine stays in state 0 in case either or both components are 0).
- If the direction component that is not expected to change in a state does change, the state changes to the error state, and recognition of the current shape is cancelled: the 'current' curve is skipped. For instance, this error handling mechanism is evoked if the state is in state 1 (the pen is in the lower right quadrant of the shape, moving right and up) and a downward move is detected.

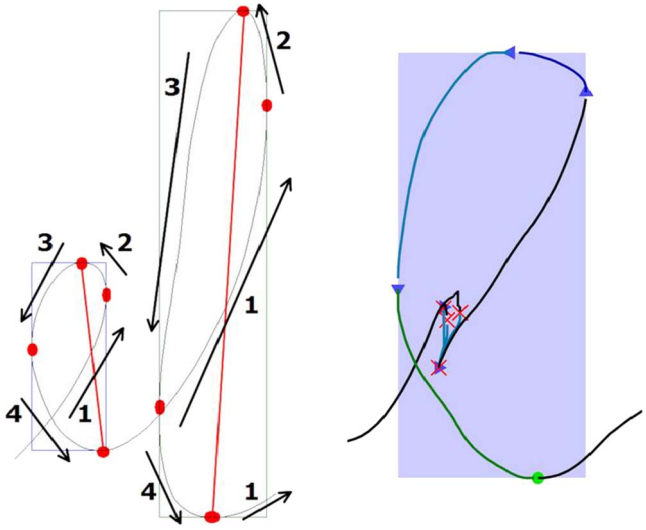


Fig. A1. Left: A sample of text containing one 'e' and one 'l', including the recognized characteristic points (red dots). The numbered black arrows show the states of the state machine. Right: An example of a real detected letter 'e'. The light blue box indicates detected letters 'e'. The line color indicates the state of the algorithm; black: state 1, dark blue: state 2, light green/cyan: state 3, green: state 4, red: state 0/error. Markers indicate state changes; blue upward arrow indicates transition from state 1 to 2, blue leftward arrow indicates transition from state 2 to 3, blue downward arrow indicates transition from state 3 to 4, a green circle indicates a transition from state 4 to state 1 and a red cross indicates a transition from any state to state 0 (the points where an error is recognized). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. A shape is considered recognized if it went through states 1, 2, 3, 4 and into the next state 1 without errors. The four characteristic points are at the four samples where the state changes occurred.
4. Recognized "shapes" that are very narrow (width < 0.7) or low (height < 2.0) are discarded. These limits were empirically determined. This step was added to the algorithm to discard small movements that were sometimes classified as a letter.

For each recognized segment in the line the rightmost, leftmost and bottommost points were saved and each of these points was characterized by an x coordinate, y coordinate and a timestamp. Then the letters were classified as an 'e' or an 'l' according to the height of the segment. A letter was classified as an 'e' when the height of the segment was below the mean letter height and a letter 'l' was classified when the height of the segment was above the mean letter height. To finish the analysis, width and height were calculated for each letter.

References

[1] R. Blank, V. Miller, H. Von Voû, Human motor development and hand laterality: a kinematic analysis of drawing movements 295 (2000) 89–92.
[2] R.C. Zietsma, Designing a Comprehensive System for Analysis of Handwriting Biomechanics in Relation to Neuromotor Control of Handwriting (PhD thesis), University of Strathclyde, Glasgow, 2010.
[3] J. Jankovic, Parkinson's disease: clinical features and diagnosis, J. Neurol. Neurosurg. Psychiatry 79 (4) (2008) 368–376.
[4] M.G. Longstaff, R.A. Heath, Spiral drawing performance as an indicator of fine motor function in people with multiple sclerosis, Hum. Mov. Sci. 25 (4–5) (2006) 474–491.
[5] M.P. Broderick, A.W.A. Van Gemmert, H.A. Shill, G.E. Stelmach, Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease, Exp. Brain Res. 197 (3) (Aug. 2009) 223–233.
[6] M.M. Ponsen, A. Daffertshofer, E.C. Wolters, P.J. Beek, H.W. Berendse, Impairment of complex upper limb motor function in de novo Parkinson's disease, Parkinsonism Relat. Disord. 14 (3) (2008) 199–204.
[7] N. Dounskaia, A.W.A. Van Gemmert, B.C. Leis, G.E. Stelmach, Biased wrist and finger coordination in Parkinsonian patients during performance of graphical tasks, Neuropsychologia 47 (2009) 2504–2514.
[8] Z. Kereszténvi, P. Cesari, G. Fazekas, J. Laczkó, The relation of hand and arm configuration variances while tracking geometric figures in Parkinson's disease: aspects for rehabilitation, Int. J. Rehabil. Res. 32 (2009) 53–63.
[9] M.S. Bryant, D.H. Rintala, E.C. Lai, E.J. Protas, An investigation of two interventions for micrographia in individuals with Parkinson's disease, Clin. Rehabil. 24 (2010) 1021–1026.
[10] S. Sülzenbrück, M. Hegele, H. Heuer, G. Rinkenauer, Generalized slowing is not that general in older adults: evidence from a tracing task, Occup. Ergnomics 9 (2010) 111–117.
[11] S. Sülzenbrück, M. Hegele, G. Rinkenauer, H. Heuer, M. Hegele, G. Rinkenauer, H. Heuer, S. Sandra, The death of handwriting: secondary effects of frequent computer use on basic motor skills, J. Mot. 43 (3) (2011) 247–251.
[12] N. van Drempt, A. McCluskey, N.A. Lannin, Handwriting in healthy people aged 65 years and over, Aust. Occup. Ther. J. 58 (4) (2011) 276–286.
[13] M. Bashir, F. Kempf, Advanced biometric pen system for recording and analyzing handwriting, J. Signal Process. Syst. 68 (1) (2011) 75–81.
[14] R. Mergl, P. Tigges, A. Schröter, H.J. Möller, U. Hegerl, Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives, J. Neurosci. Methods 90 (2) (1999) 157–169.
[15] E.J. Smits, A.J. Tolonen, L. Cluitmans, M. Van Gils, B.A. Conway, R.C. Zietsma, K.L. Leenders, N.M. Maurits, Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease, PLoS One 9 (5) (2014) 1–8.
[16] E.J. Smits, A.J. Tolonen, L. Cluitmans, M. Van Gils, R.C. Zietsma, R.W.K. Borgemeester, T. van Laar, N.M. Maurits, Graphical tasks to measure upper limb function in patients with Parkinson's disease: validity and response to dopaminergic medication, IEEE J. Biomed. Heal. Inform. 2194(c) (2015) 1–1.
[17] A. Van den Bruel, I. Cleemput, B. Aertgeerts, D. Ramaekers, F. Buntinx, The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed, J. Clin. Epidemiol. 60 (11) (2007) 1116–1122.
[18] L. Erasmus, S. Sarno, H. Albrecht, M. Schwecht, Measurement of ataxic symptoms with a graphic tablet: standard values in controls and validity in Multiple Sclerosis patients 108 (2001) 25–37.
[19] P. Feys, W. Helsen, A. Prinsmel, S. Ilsbroukx, S. Wang, X. Liu, Digitised spirography as an evaluation tool for intention tremor in multiple sclerosis, J. Neurosci. Methods 160 (2007) 309–316.
[20] M.K. Rand, G.E. Stelmach, J.R. Bloedel, Movement accuracy constraints in Parkinson's disease patients, Neuropsychologia 38 (2000) 203–212.
[21] J. Desrosiers, R. Hébert, G. Bravo, E. Dutil, The Purdue Pegboard Test: normative data for people aged 60 and over, Disabil. Rehabil. 17 (5) (1995) 217–224.
[22] J. Tiffin, E.J. Asher, The Purdue pegboard; norms and studies of reliability and validity, J. Appl. Psychol. 32 (1948) 234–247.
[23] T. Müller, S. Schäfer, W. Kuhn, H. Przuntek, Correlation between tapping and inserting of pegs in Parkinson's disease, Can. J. Neurol. Sci. 27 (2000) 311–315.
[24] P.M. Fitts, The information capacity of the human motor system in controlling the amplitude of movement, J. Exp. Psychol. 47 (6) (1954) 381–391.
[25] J.P. Weir, Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM, J. Strength Cond. Res. 19 (1) (2005) 231–240.
[26] E.M. Andresen, Criteria for assessing the tools of disability outcomes research, Arch. Phys. Med. Rehabil. 81 (2000) 15–20.
[27] L. Holper, D.C. Kiper, K. Eng, S. Vuillermot, An extended drawing test for the assessment of arm and hand function with a performance invariant for healthy subjects 177 (2009) 452–460.
[28] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, I. Schlesinger, Handwriting as an objective tool for Parkinson's disease diagnosis, J. Neurol. 260 (9) (2013) 2357–2361.
[29] A. Ünlü, R. Brause, K. Krakow, Handwriting analysis for diagnosis and prognosis of Parkinson's disease, ISBMDA Lect. Notes Comput. Sci. 4345 (2006) 441–450.
[30] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R.B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, S.L. Pullman, Validity of spiral analysis in early Parkinson's disease, Mov. Disord. 23 (4) (2008) 531–537.
[31] N.P.S. Bajaj, L. Wang, V. Gontu, D.G. Grosset, P.G. Bain, Accuracy of subjective and objective handwriting assessment for differentiating Parkinson's disease from tremulous subjects without evidence of dopaminergic deficits (SWEDDs): an FP-CIT-validated study, J. Neurol. 259 (11) (2012) 2335–2340.

Distinguishing Parkinson's disease from other syndromes causing tremor using automatic analysis of writing and drawing tasks

Antti Tolonen[1], Luc Cluitmans[1], Esther Smits[2], Mark van Gils[1], Natasha Maurits[2] and Rutger Zietsma[3]

***Abstract*— An easily performed and objective test of patients fine motor skills would be valuable in the diagnosis of Parkinson's disease (PD). In this study we present a set of automatic methods for quantifying the motor symptoms of PD and show that these automatically extracted features can be used to distinguish PD from other movement disorders causing tremor, namely essential tremor (ET), functional tremor (FT) and enhanced physiological tremor (EPT). The classification accuracies (mean of sensitivity and specificity) for separating PD from the other tremor syndromes were 82.0 % for ET, 69.8 % for FT and 72.2 % for EPT.**

I. INTRODUCTION

Parkinson's disease (PD) is a progressive neurological disorder which causes severe dysfunctionality of the motor system. The four cardinal motor symptoms of PD are: tremor at rest, rigidity, bradykinesia (slowness of movement) and postural instability, also micrographia (decreased size of handwriting) is a common motor symptom of PD. [1]

In this study we evaluated the usefulness of a newly developed measurement system, consisting of an electronic pen and a digitizer tablet, in separating PD patients from patients with other movement disorders causing tremor. We developed a set of fully automatic feature extraction methods for quantifying motor symptoms of PD and we tested the usefulness of these features using statistical machine learning methods.

Similar devices and methods have been studied also before, mainly for the separation of PD from ET.

Muthuraman et al. [2] studied 41 ET patients and 39 tremor-dominant PD patients by tremor accelerometry on the most affected side. The study patients were chosen so that the two groups were indistinguishable using tremor frequency, peak power or number of harmonic peaks. In their study, mean harmonic peak power of all harmonics of the tremor frequency separated PD from ET with 94 % accuracy.

Hossen et al. [3] studied postural tremor from the more affected side using an accelerometer attached to the hand, to distinguish PD from ET patients. Their method achieved an accuracy of 82.5 % for this two-class problem.

Wile et al. [4] used a smart watch containing an accelerometer and extracted mean power of harmonic peaks similar to the method used by Muthuraman et al. [2]. The study population consisted of 14 ET and 11 PD patients with clinically relevant tremor. Using a threshold optimized for their data set they achieved a sensitivity of 90.9 % and a specificity of 100 %, but using the optimal threshold value obtained from recordings by Muthuraman et al. the sensitivity was 100 % and specificity only 64.3 %.

Bidayasiri et al. [5] studied the separation of PD and ET patients using 3D accelerometer and gyroscope attached to the hand. They found statistically significant differences in multiple features extracted from the accelerometer and gyroscope signals, but no classification accuracy is reported.

Zeuner et al. [6] studied the separation of PD, ET and FT patients from each other using accelerometry in different tapping tests. Features derived from accelerometry were found to differ between FT patients and PD and ET patients, but no accuracy measures were reported.

To extend these previous studies, we studied the separation of multiple tremor syndromes from the PD and our methods combine quantification multiple motor symptoms instead of focusing on only one feature such as spectral content of the tremor.

II. DATA AND METHODS

A. Study participants

The data set consisted of patients suffering from Parkinson's disease, essential tremor (ET), functional tremor (FT) and enhanced physiological tremor (EPT). The number of participants in each category are listed in Table I.

The recordings for the PD and tremor patients were done at University Medical Center Groningen (UMCG) (59 patients) and at Dublin Neurological Institute (DNI) (33 patients).

TABLE I
SUMMARY OF THE STUDY PARTICIPANTS

Group	N	Age (mean ± std.)	Gender (M/F)
PD	58	64.9 ± 9.6	29/29
ET	16	60.7 ± 12.3	12/4
FT	8	58.6 ± 16.3	2/6
EPT	10	42.3 ± 16.3	4/6

The PD patient group was quite diverse. The time since onset of the symptoms ranged from 0 to 18 years (mean 5.0 years and standard deviation 3.5 years). The severity of the symptoms, measured by the Hoehn and Yahr scale [7], also varied (mean 2.2 and standard deviation 0.7). The Hoehn and Yahr scale was only assessed at UMCG, i.e. only for 30 of the 58 PD patients.

[1]VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Tampere, Finland antti.tolonen@vtt.fi; luc.cluitmans@vtt.fi; mark.vangils@vtt.fi

[2] University of Groningen, University Medical Center Groningen, Department of Neurology, AB51 P.O. Box 30001, 9700 RB Groningen, the Netherlands e.j.smits@umcg.nl; n.m.maurits@umcg.nl

[3]Manus Neurodynamica Ltd 4a John Street, Cullercoats, North Shields Tyne and Wear, NE30 4PL, United Kingdom r.c.zietsma@manusneuro.com

B. Experimental design

The drawing and writing tasks were performed using a custom made digitizer pen and a tablet computer (ASUS Eee Slate EP121). The pen tip location on the tablet was recorded and the pen contained a gyroscope measuring the angular velocity of the pen in three directions (pitch, yaw and roll). The sampling frequency for all signals was 200 Hz. In some of the tasks the tablet provided visual feedback for the participants (i.e. the trail left by the pen was drawn on the tablet, simulating the behavior of a real pen) and in some of the tasks visual feedback was turned off.

The participants performed a set of drawing and writing tasks using the electronic pen: a rest task, circle, spiral and zigzag drawing tasks, an "elelelel" writing task, and a modified version of the original Fitts' task [8].

In the rest task the participants held the pen tip still on a target (filled circle, diameter 7 mm). The arm and hand were resting on the table and the tablet.

In the circle and spiral drawing tasks the participants traced a circle or spiral shown on the tablet. The templates and their dimensions are illustrated in Figure 1. No visual feedback was shown to the participants.

The zigzag task consisted of four subtasks, in the first subtask the participants had to trace a shape shown on the tablet (template is shown in Figure 1), in the second subtask the participants were shown the zigzag shape on a paper and they had to copy the zigzag shape on the tablet, in the third subtask the participants had to copy the zigzag shape with their eyes closed, and in the fourth subtask the participants were asked to draw a 90 degrees rotated version of the zigzag shape shown on the paper. Visual feedback was shown to the participants in all but the tracing subtask.

The "elelelel" writing task consisted of two subtasks, the participants had to write the phrase "elelelel" with visual feedback and without visual feedback.

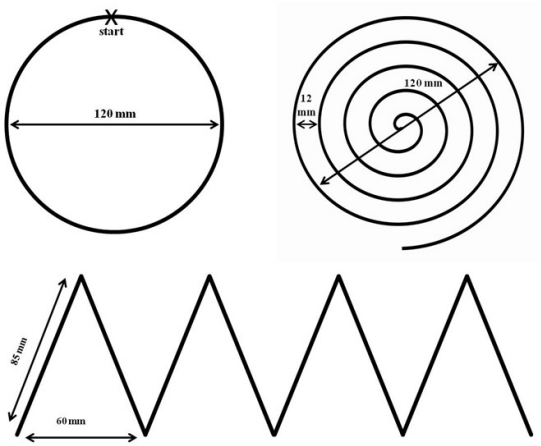
In the Fitts' task the participants had to move the pen tip between two targets (circles with varying size) as fast and accurately as possible for 20 seconds. The Fitts' task consisted of eight subtasks, in which the distance between the targets and the target sizes were varied; distances used were 70 mm and 200 mm and diameters of the targets were 7, 13, 19 and 25 mm.

C. Feature extraction

We extracted a set of features for quantifying different motor symptoms of the PD, such as movement speed and frequency and amplitude of rest and kinetic tremors which are essential for differential diagnosis of PD [1]. All of the feature computations are performed automatically, without for example manually marking time points in the signals. Counting all features from each task separately the total number of features is 118.

From the rest task we computed a set of features quantifying rest tremor. The dominant tremor frequency and the relative power (power in the band divided by the total power) in a 1 Hz band around this frequency were first estimated from the power spectral density (PSD) of the gyroscope signals, and used as features for the classifier. Subsequently the gyroscope and pen tip coordinate signals were filtered with a multiband filter that attenuates everything except the dominant frequency and its first three harmonics. From the resulting signals we computed the root mean square of the tremor amplitude in 1 second long running windows for both the gyroscope and the pen tip coordinate signals. As summary features for the tremor amplitudes we computed the mean of the amplitudes and the relative rate of change of the tremor amplitude (i.e. the coefficient a in $A(t)/\bar{A} = at + b$, where $A(t)$ is the tremor amplitude at time t and $\bar{A}$ is the mean of $A(t)$ and a and b are coefficients chosen to minimize the mean square error $E\{(A(t) - at + b)^2\}$). The mean amplitude and the relative rate of change of the amplitude were computed for both the gyroscope and pen tip coordinate signals. From the pen tip coordinates we also estimated the direction of tremor by computing the first principal component (PC) of the tremor signals; the x and y components of the vector describing the direction of the 1st PC and the relative amount of variance explained by the 1st PC are used as features. The tremor direction can for example be used to differentiate between distal and proximal tremor [9].

Fig. 1. Templates and their dimensions for the tracing and drawing tasks.



For circle, spiral and zigzag tasks we computed the same tremor features as for the rest task (in these tasks the features quantify kinetic tremor). We also located the start and end points of each shape using a segmentation method based on the dynamic time warping (DTW) algorithm. Using this segmentation we computed the average time to draw each shape and the relative rate of change of this time. The relative rate of change is computed similarly as for the tremor amplitude.

For the copy, rotate and blind subtasks of the zigzag task we also computed the mean size (height for the zigzag copy and blind, and width for the zigzag rotate) and the relative rate of change of the mean size.

For the "elelelel" writing task we first located the start and end point of each "e" and "l" using a heuristic segmentation algorithm [10]. Using this segmentation we then computed

the median width of each letter and median time to write each letter as features.

For the Fitts' task we first located the touching point of each touch. Then we computed the mean time to move the pen from one target to the other, and the percentage of under and over shoot touches (i.e. touches that are made before the target or that go beyond the target) as features.

D. Classification

We tested the differentiation capability of the extracted features for a set of classification tasks loosely corresponding to different differential diagnostics scenarios. The first classification task is a four-class problem in which the classifier needs to separate between all four different movement disorders. The subsequent three classification tasks are two-class problems in which the classifier needs to separate between PD and one of the three different tremor syndromes.

We used multinomial and binomial logistic regression for classification. In order to reduce overfitting we estimated the regression coefficients by minimizing a penalized likelihood function; the regularization term used is the L2 norm of the regression coefficients. Before computing the regression coefficients the features have been normalized to have zero mean and unit variance. The samples in each class have been weighed in the penalized likelihood so that the total sum of the samples in each class is the same. This weighting means that the classifier is not trying to achieve higher accuracy on more prevalent classes, which would result in better total accuracy but poorer accuracy on the less prevalent classes. The Glmnet for Matlab implementation of logistic regression was used (http://www.stanford.edu/∼hastie/glmnet_matlab/).

E. Performance evaluation

The classifier performance has been tested using patient-wise leave-one-out (LOO) cross-validation, i.e. when testing the classifier for each patient it is trained using all the other patients as a training set. The weight of the regularization term in the penalized likelihood is chosen automatically using cross-validation within each training set; using all the patients for choosing the weight would cause a small positive bias in the classification accuracy.

III. RESULTS

The confusion matrix and the per class sensitivities (i.e. the proportion of patients classified correctly in each class) for the separation of the PD patients and the patients suffering from other tremor syndromes are shown in Table II. The classifier is able to classify 69.0 % of the PD patients correctly; PD patients are misclassified as ET or EPT patients. The accuracy for the classification for ET and FT patients is low, and only one of the FT patients is classified correctly.

The confusion matrix in Table III is obtained from the more detailed confusion matrix in Table II by pooling all patients from ET, EPT and FT groups together. The classification accuracy for this pooled class is naturally higher than for the three classes separately since for example misclassifying FT as ET is no longer counted as an error.

TABLE II
CONFUSION MATRIX FOR THE SEPARATION OF PD PATIENTS AND THE THREE TREMOR SYNDROMES

Real\Est.	PD	ET	FT	EPT		Sens. [%]
PD	40	11	0	7		69.0
ET	4	7	1	4		43.8
FT	1	6	1	0		12.5
EPT	3	0	0	7		70.0

TABLE III
CONFUSION MATRIX FOR THE SEPARATION OF THE PD PATIENTS FROM ANY TREMOR SYNDROME

Real\Est.	PD	Tremor		Sens. [%]
PD	40	18		69.0
Tremor	8	26		76.5

Table IV lists the sensitivities, specificities and balanced accuracies (mean of sensitivity and specificity) for each of the two-class problems. PD is always the positive class (i.e. the class for which sensitivity is computed) and each of the tremor syndromes is the negative class (i.e. the class for which specificity is computed).

TABLE IV
SENSITIVITIES AND SPECIFICITIES FOR THE TWO-CLASS CLASSIFICATION PROBLEMS

Neg. class	Sens. [%]	Spec. [%]	Bal. acc. [%]
ET	82.8	81.3	82.0
FT	89.7	50.0	69.8
EPT	84.5	60.0	72.2

IV. DISCUSSION AND CONCLUSIONS

The separation of PD from other tremor syndromes can be done with good accuracy using the features extracted from the drawing and writing tasks. The separation of PD patients from the pooled tremor syndrome class was done with an accuracy of 72.7 % (mean of per class sensitivities). The system was less successful in separating the different tremor syndromes from each other, for example only one of FT patients was correctly classified. The separation of the different tremor syndromes is not however what the system was primarily designed for.

The separation of PD patients from each of the tremor syndromes in a pairwise comparison could be done with an accuracy of 69.8 - 82.0 %. The separation of PD from ET is probably clinically most relevant since misdiagnosis of ET as PD is common. In a community-based study of the accuracy of PD diagnosis, ET accounted for 48 % of the misdiagnoses [11]. The separation of PD from ET could be done with an accuracy of 82.0 %.

Muthuraman et al. [2] and Wile et al. [4] have reported higher classification accuracies for the PD vs. ET classification using the mean harmonic peak power of the tremor,

than what was achieved in our study. We also computed for the rest task the mean harmonic peak power similarly as Wile et al. [4], but in our patient population there was no statistically significant difference between the ET and PD patients. Therefore this measure is not included in the feature set used by our classifier. The discrepancy between our study and the earlier two studies is probably due to differences in patient population and test conditions. The rest task used in our study is very different from the outstretched hands position used in these earlier studies, and the PD patients in our study were not selected based on their tremor symptoms.

The classification accuracy could possibly be improved by redesigning the rest task so that it would clearly measure rest tremor. In the current measurement the essential tremor patients seem to have more tremor in the rest task than the PD patients, which could indicate that the rest task is, for at least some subjects, actually measuring postural tremor [12]. The rest tremor is the most common symptom of PD [1]. Therefore redesigning the tasks, so that they would clearly separate between rest and postural tremor, can be expected to improve the accuracy of the differential diagnosis.

ACKNOWLEDGMENTS

This study is part of the European Commission FP7-SME-201001 programme project DiPar (grant agreement 262291). The hardware of the measurement setup was provided by Fraunhofer IPMS (Dresden, Germany) and the data acquisition software by Fraunhofer IPA (Stuttgart, Germany), based on a concept by Manus Neurodynamica Ltd (patent WO/2011/141734).

We would like to thank E. Fallon and Prof. T. Lynch for assisting in recruiting patients and performing measurements for this study at the Dublin Neurological Institute at the Mater Misericordiae University Hospital in Dublin, Ireland.

REFERENCES

[1] J. Jankovic, Parkinsons disease: clinical features and diagnosis., J. Neurol. Neurosurg. Psychiatry, vol. 79, no. 4, pp. 368-76, 2008.

[2] M. Muthuraman, A. Hossen, U. Heute, G. Deuschl, and J. Raethjen, A new diagnostic test to distinguish tremulous Parkinsons disease from advanced essential tremor, Mov. Disord., vol. 26, no. 8, pp. 1548-1552, 2011.

[3] A. Hossen, M. Muthuraman, Z. Al-Hakim, J. Raethjen, G. Deuschl, and U. Heute, Discrimination of Parkinsonian tremor from essential tremor using statistical signal characterization of the spectrum of accelerometer signal., Biomed. Mater. Eng., vol. 23, no. 6, pp. 513-31, 2013.

[4] D. J. Wile, R. Ranawaya, and Z. H. T. Kiss, Smart watch accelerometry for analysis and diagnosis of tremor., J. Neurosci. Methods, vol. 230, pp. 1-4, 2014.

[5] R. Bhidayasiri, Differential diagnosis of common tremor syndromes., Postgrad. Med. J., vol. 81, no. 962, pp. 756-62, 2005.

[6] K. E. Zeuner, R. O. Shoge, S. R. Goldstein, J. M. Dambrosia, and M. Hallett, Accelerometry to distinguish psychogenic from essential or parkinsonian tremor, Neurology, vol. 61, no. 4, pp. 548-550, 2003.

[7] Hoehn, M.M. & Yahr, M.D., 1967. Parkinsonism: onset, progression and mortality. Neurology, 17(5), pp. 427-442.

[8] P. M. Fitts, The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement, J. Exp. Psychol. Gen., vol. 121, no. 3, pp. 262-269, 1992.

[9] S. Wang, P. G. Bain, T. Z. Aziz, and X. Liu, The direction of oscillation in spiral drawings can be used to differentiate distal and proximal arm tremor., Neurosci. Lett., vol. 384, no. 12, pp. 188-92, 2005.

[10] E. J. Smits, A. J. Tolonen, L. Cluitmans, M. van Gils, B. A. Conway, R. C. Zietsma, K. L. Leenders, and N. M. Maurits, Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinsons disease., PLoS One, vol. 9, no. 5, p. e97614, 2014.

[11] J. Meara, B. K. Bhowmick, and P. Hobson, Accuracy of diagnosis in patients with presumed Parkinsons disease, Age Ageing, vol. 28, no. 2, pp. 99-102, 1999.

[12] R. J. Elble and G. Deuschl, An update on essential tremor., Curr. Neurol. Neurosci. Rep., vol. 9, no. 4, pp. 273-277, 2009.

Objective and easily performed assessment of fine motor skills to support the differential diagnosis of Parkinson's disease (PD) and other movement disorders

C. Papasavvas[a], R. Zietsma[a], S. Dodds[d], L. Cluitmans[b], A. B. Deutschlander[c], R. Walker[d]

[a]Manus Neurodynamica Ltd, Edinburgh, UK, [b]VTT, Technical Research Centre of Finland, Finland [c]Mayo Clinic, Florida, USA [d]Northumbria Healthcare NHS Foundation Trust, North Shields, UK

BACKGROUND

The differential diagnosis between Parkinson's disease (PD) and other movement disorders still poses a challenge, especially at the early stages[1]. A correct early diagnosis enables early intervention and lifestyle adaptations with highly improved outcomes for the patients while also making diagnostic and therapeutic processes easier for patients and clinicians.

The current diagnostic toolset has several limitations:

- Symptoms can be mimicked by other movement disorders and diagnosis is mostly based on clinical symptoms with partly subjective criteria[2].
- Brain scans detecting dopamine transporters (DaTSCANs) are frequently used for diagnosis[3], but they are expensive and hard-to-tolerate procedures.
- Current tremor analyses are not accurate enough for differentiation between movement disorders presenting with tremor.

Here, we present a device that quantifies movement abnormalities, expressed as digital biomarkers[4], to quantify the subtlest signs and symptoms and help identify underlying physiological changes to assist in diagnostic decision making of 'difficult to diagnose' patients with PD. The easily performed and objective assessment incorporates additional AI-driven diagnostic decision support[5] at the point of care.

STUDY OBJECTIVE

To assess the ability of a handheld device (*NeuroMotorPen™*) to extract informative features of motor skills and subtle symptoms and distinguish between Parkinson's Disease (PD) and other movement disorders. More specifically, the study aims to asses the ability of the device to detect the presence of PD (sensitivity) and the absence of PD (specificity) among movement disorder patients through cross-validated classification.

Scan to download poster

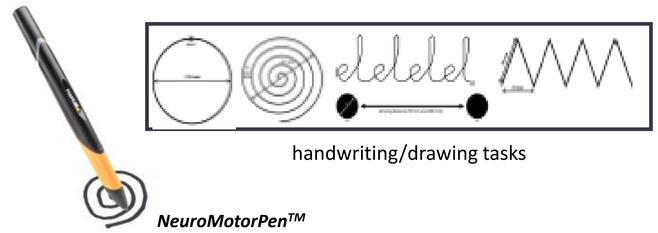
DATA & METHODS

Available Data

We collected data from 76 suspected new PD cases, for whom no clinical unequivocal diagnosis could be made and 49 were later diagnosed with PD. The remaining 27 were diagnosed with other movement disorders (non-PD).

***NeuroMotorPen™* device and tasks**

This proprietary multi-sensor digital pen with a series of standardised handwriting/drawing tasks was designed to measure various clinical factors. In total, 243 features were extracted from the raw data with the majority of them being related to the following symptoms: **tremor**, **bradykinesia**, **micrographia**, **spatial accuracy** and **amplitude of movement**.



handwriting/drawing tasks

NeuroMotorPen™

Feature Selection

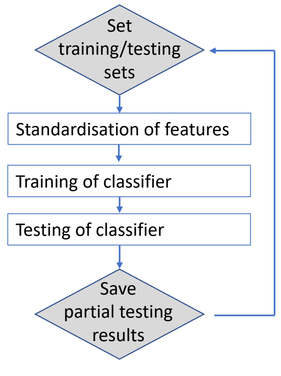
A supervised feature selection process was used to find features with complementary information. It relies on the **Area Under the Curve (AUC)** values of individual features (AUC is a measure of discriminability between two classes, non-PD vs PD in this case).

Classifier

We used a **Adaptive Boosting** classifier from *sklearn* with its default settings but with a sightly lower learning rate.

Cross-validation

We used **leave-one-out** cross-validation. See the diagram of the iterative process and the data used in each step.

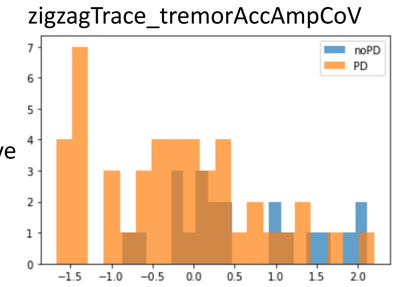


RESULTS

Highly informative features

The feature selection process revealed that 42 of the total 243 features provide potentially useful information for discriminating between the two classes, with AUC > 0.6. These features represent all different tasks and provide measures for all symptoms. Furthermore, some of them exhibit high discriminability with AUC > 0.7 (see example below)

Example of a highly informative feature with AUC = 0.74

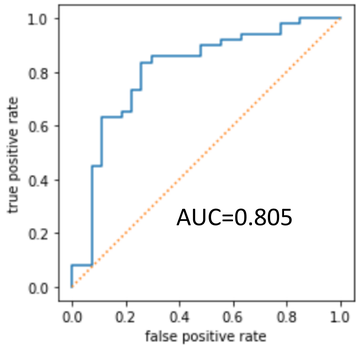


Classification results

The cross-validated classification results for all 76 patients indicate high **sensitivity** (85.7%) and good **specificity** (70.4%), while accuracy was measured at 80.3%. Results are summarised in the confusion matrix:

True Positives = 42
True Negatives = 19
False Positives = 8
False Negatives = 7

predicted class \ true class	PD	nonPD
PD	42	8
nonPD	7	19



ROC curve:
Its shape indicates that the classifier is primarily sensitive rather than specific

DISCUSSION & CONCLUSION

The results indicate that the *NeuroMotorPen™* device extracts multiple informative features of subtle movement abnormalities that are clinically not easily observable and those features are useful in distinguishing between PD and non-PD movement abnormalities. We used a combination of selected features to train a classifier, which achieved high sensitivity; that is, it managed to detect the presence of PD in most of the later-diagnosed PD patients. A highly sensitive classifier can be specifically useful for screening purposes when any indication of PD needs to be detected in early stages for further examination.

The findings suggest the device could support decision making for difficult-to-diagnose cases, potentially avoiding the need for DaTSCAN, thus saving time and money. The device additionally provides opportunities in objective monitoring of symptoms during the course of treatment.

An early and accurate diagnosis is crucial to give patients clarity and access to the range of treatments available and improve their quality of life.

REFERENCES

1. Baumann, C. R. "Epidemiology, diagnosis and differential diagnosis in Parkinson's disease tremor." *Parkinsonism Relat Disord* 18 (2012): S90-S92.
2. Post, B. et al. "Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?." *Mov Disord,* 20.12 (2005): 1577-1584.
3. Gayed, I., et al. "The impact of DaTscan in the diagnosis of Parkinson disease." *Clin Nucl Med* 40.5 (2015): 390-393.
4. Smits, E. J., et al. "Graphical tasks to measure upper limb function in patients with Parkinson's disease: Validity and response to dopaminergic medication." *IEEE J Biomed Health Inform* 21.1 (2015): 283-289.
5. Tolonen, A., et al. "Distinguishing Parkinson's disease from other syndromes causing tremor using automatic analysis of writing and drawing tasks." *2015 IEEE 15th Int Conf Bioinform Biomed Eng*. IEEE, 2015.