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

Smits, Esther J.; Tolonen, Antti J.; Cluitmans, Luc; van Gils, Mark; Zietsma, Rutger C.; Tijssen, Marina A.J.; Maurits, Natasha M.

Published in:
Measurement: Journal of the International Measurement Confederation

DOI:
10.1016/j.measurement.2017.09.011

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

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

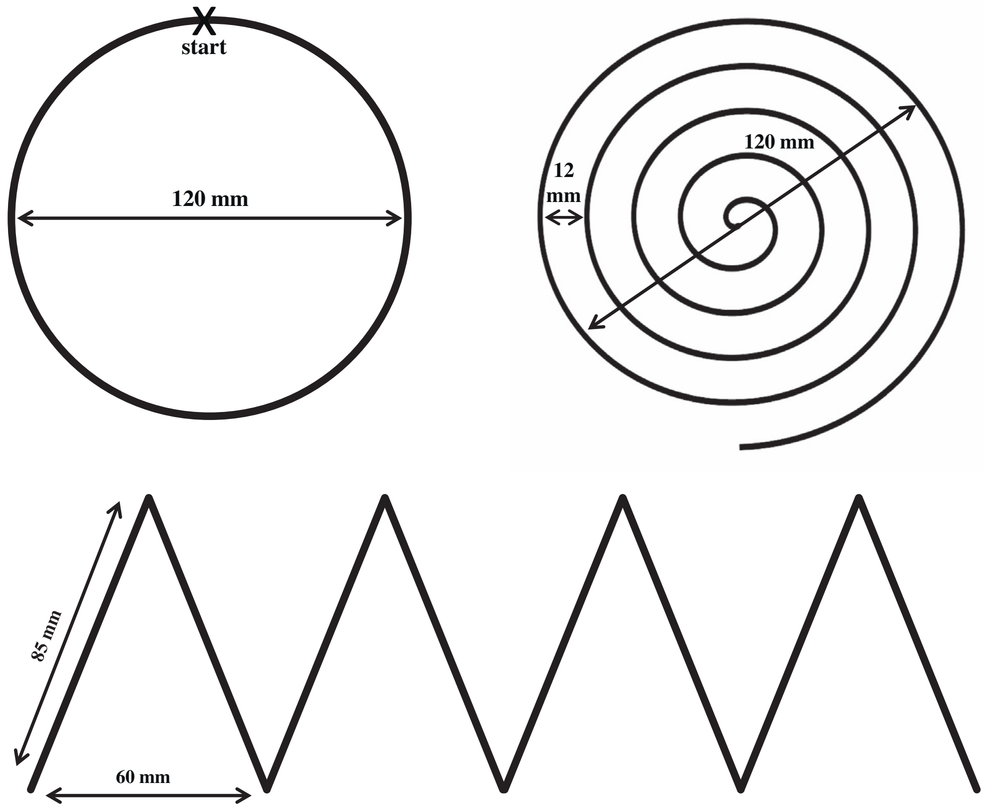

Fig. 1. Templates and their dimensions for the tracing and drawing tasks: a circle, spiral and zigzag figure.

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

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

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.

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.

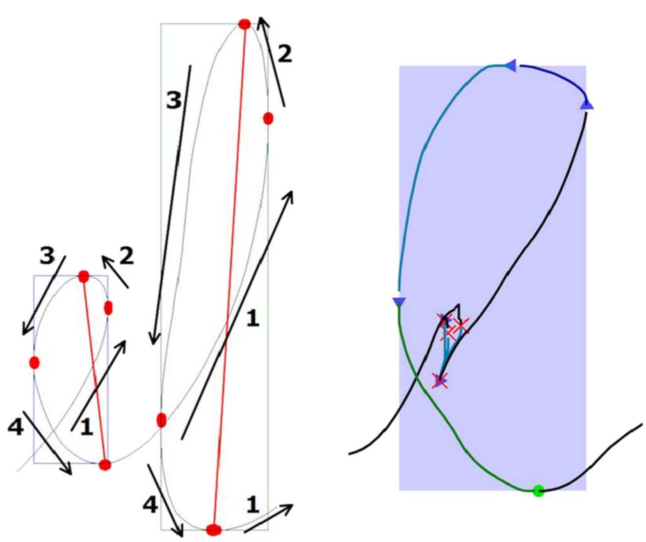

Fig. A1. Left: A sample of text containing one 'e' and one 'l', including the recognized characteristic points (red dots). The numbered black arrows show the states of the state machine. Right: An example of a real detected letter 'e'. The light blue box indicates detected letters 'e'. The line color 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. Keresztényi, 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.