


Optical Hand Tracking: A Novel Technique for the Assessment of Bradykinesia in Parkinson's Disease

Paulina J.M. Bank, PhD,^{1,*}  Johan Marinus, PhD,¹ Carel G.M. Meskers, MD, PhD,^{2,3} Jurriaan H. de Groot, PhD,⁴ Jacobus J. van Hilten, MD, PhD¹

Abstract: **Background:** Evaluation of therapies for Parkinson's disease (PD) may benefit from objective quantification of the separate movement components of bradykinesia (i.e., velocity, amplitude, and rhythm). This study evaluated the sensitivity and reliability of parameters derived from recently available optical hand tracking techniques for patient-friendly, automated quantification of bradykinesia of the upper extremity in PD. **Methods:** Fifty-seven patients with PD and 57 healthy individuals (controls) performed repetitive finger tapping (RFT), alternating hand movements (AHM), and alternating forearm movements (AFM). Movement components of bradykinesia (i.e., velocity, frequency, amplitude, hesitations, and halts) were quantified using optical hand tracking. Reliability was quantified using intraclass correlation coefficients in a subgroup of 12 patients with PD and 12 controls (test-retest) and in all 57 controls (intra-trial). **Results:** RFT and AHM were successfully recorded in 94% of all participants. Movement components differed between patients with PD and controls and were correlated with clinical ratings. Velocity and halt duration appeared to be most useful (i.e., the largest difference between the PD and control groups, good reliability) for the quantification of RFT, whereas frequency appeared to be most useful for the quantification of AHM. Other variables, such as frequency and amplitude of RFT, showed poor test-retest reliability, because they were susceptible to changes in movement strategy. AFM was excluded from the analysis because of problems with hand recognition. **Conclusion:** Novel optical hand tracking techniques yield promising results for patient-friendly quantification of bradykinesia of the upper extremity in PD. Future work should aim to optimize optical hand tracking and reduce susceptibility to changes in strategy.

Bradykinesia is the cardinal motor manifestation and is required to diagnose Parkinson's disease (PD).¹ It is characterized by slowness of movement, associated with a decrement in amplitude or speed, and with hesitations or halts as movements are continued.¹

Motor impairments in PD are commonly assessed with the Unified PD rating scale (International Parkinson and Movement Disorders Society [MDS]-UPDRS),² which addresses bradykinesia of the upper limbs by means of 3 items (Items 3.4–3.6) on repetitive finger and hand movements. For each item, a

combination of movement parameters (i.e., [progressive reduction of] speed and amplitude, and the occurrence of interruptions of rhythm) is simultaneously rated on a 5-point scale.

Recent work shows that the translation of what meets the eye into 5-point rating categories is rather ambiguous. A comparison of the separate evaluation of 3 component scores of movement (speed, amplitude, and rhythm) with the MDS-UPDRS ratings of bradykinesia revealed that the majority of clinicians based their clinical ratings mainly on amplitude,

¹Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands; ²Department of Rehabilitation Medicine, VU University Medical Center, Amsterdam, the Netherlands; ³MOVE Research Institute, Amsterdam, the Netherlands; ⁴Department of Rehabilitation Medicine, Leiden University Medical Center, Leiden, the Netherlands

*Correspondence to: Dr. Paulina J.M. Bank, Department of Neurology, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands; E-mail: p.j.m.bank@lumc.nl

Keywords: bradykinesia, Movement Disorders Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS), optical hand tracking, Parkinson's disease, test-retest reliability.

Relevant disclosures and conflicts of interest are listed at the end of this article.

Supporting information may be found in the online version of this article.

Received 27 March 2017; revised 19 July 2017; accepted 8 August 2017.

Published online 13 September 2017 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/mdc3.12536

whereas others based their judgment on speed or rhythm.³ Such different rating strategies may influence inter-rater reliability and the evaluation of treatment effects, because different components of movement may show a differential response to treatment (e.g., speed is the component most responsive to dopaminergic medication).⁴ Based on these findings, Heldman et al.³ argued that quantifying the subcomponents of motion separately may enable more accurate evaluation of interventions in clinical care and research.

Although clinical scales and patient-reported outcomes are still the primary assessment tools or endpoints in PD clinical care and research, there is a growing awareness that “technology-obtained measures” may contribute to patient management by improving the sensitivity, accuracy, reproducibility, and feasibility of objectively capturing the full complexity and diversity of changes in motor behavior.⁵ New technologies for the objective quantification of bradykinesia in PD thus should aim for higher resolution and improved sensitivity to capture the nuances of movement. A recent study demonstrated that, compared with clinical ratings (MDS-UPDRS and the Modified Bradykinesia Rating Scale⁶), inertial measurement units (IMUs) attached to the thumb and index finger were more sensitive to subtle changes in motor performance associated with changes of stimulation parameters of the subthalamic nucleus.⁷

A broad range of currently available techniques can objectively quantify bradykinesia in the laboratory or in ambulatory settings.^{5,8–10} However, most of these techniques require the attachment of sensors or markers to the fingers and/or hand, which can affect the movement, requires preparation time, and may depend on the correct execution of a calibration procedure. Other techniques based on a keyboard^{11–13} or tablet¹⁴, are easy to use but only provide information about the contact phase rather than the full movement cycle.

In this study, we evaluated the potential of optical hand tracking (OHT) for separately quantifying the movement components of bradykinesia in PD, which may improve progression monitoring and enable more accurate evaluation of (novel) therapies.

Patients and Methods

Participants

For this cross-sectional study, we recruited 57 patients with PD who fulfilled the UK PD Brain Bank criteria¹⁵ from the outpatient clinic of the Department of Neurology at Leiden University Medical Center (for patient characteristics, see Table 1). Patients were excluded if they had disorders of the central nervous system other than PD or other conditions that could affect motor function of the upper extremity. All patients were allowed to take their routine PD medications. Fifty-seven healthy controls (23 women; mean \pm standard deviation age, 63.8 ± 7.6 years) who were sex-matched and age-matched (± 3 years) to the patients, were recruited through advertisements and from a database of volunteers who had participated in previous studies. Healthy controls had no history of disorders affecting the function of the upper extremities.

TABLE 1 Patient characteristics

Characteristic	Total	Test-retest
Total no.	57	12
Sex: Men/women, no.	36/21	6/6
Age: Mean \pm SD, y	65.2 \pm 8.8	69.4 \pm 6.5
Disease duration, y	12.6 \pm 6.8	14.3 \pm 6.8
Hoehn and Yahr stage:	3 (1-5)	3 (2-4)
Median (range) ^a		
MDS-UPDRS-III ^b	36.6 \pm 16.3	39.1 \pm 14.6

SD, standard deviation; MDS-UPDRS-III, International Parkinson and Movement Disorders Society-sponsored revision of the Unified Parkinson's Disease Rating Scale, part III (motor evaluation).

^aHoehn and Yahr staging ranges from 0 to 5, with higher stages indicating worse disease.

^bMDS-UPDRS-III scores range from 0 to 132, with higher scores indicating worse disease.

Test-retest reliability was evaluated in a subgroup of 24 participants (12 patients with PD and 12 controls) who repeated the test after 1 week at the same hour of the day. Informed consent was obtained according to the Declaration of Helsinki. The Ethical Committee of the Leiden University Medical Center approved of the study's protocol.

Measurement Instruments and Data-Collection Procedure

Clinical assessment

In all 57 patients, repetitive finger tapping with the thumb and index finger (RFT) (Item 3.4 of the MDS-UPDRS), alternating hand movements (AHM) (opening and closing of the hand; Item 3.5), and alternating forearm movements (AFM) (pronation and supination; Item 3.6) were each scored according to the MDS-UPDRS guidelines by a certified researcher. Scores were rated on a scale ranging from 0 (normal; no problems) to 4 (severe; cannot or can only barely perform the task because of slowing, interruptions, or decrements). Participants who repeated the test after 1 week rated the perceived change in function of the arm/hand on an 11-point numeric rating scale (0, much worse; 5, unchanged; 10, much better).

Optical hand tracking

The 3 tasks (RFT, AHM, and AFM) were evaluated in all participants using OHT (see Fig. 1A). The system comprised a sensor (ASUS Xtion Pro 2.0; ASUSTeK Computer Inc., Taipei, Taiwan), consisting of an infrared laser transmitter and an infrared camera providing depth data, in combination with a 3Gear software development kit (SDK v 0.9.32; 3Gear Systems Inc., San Francisco, CA). The sensor was mounted on an adjustable pole hanging from the ceiling, about 0.70 m above the hands of the participants when arms were outstretched in front of the shoulders. Computer graphics algorithms within the SDK, which were trained to recognize the shape and pose of a hand from the 3-dimensional (3D)-depth data, provided real-time 3D-coordinates of the wrist, the finger joints, and the finger tips at a sampling rate of 30 Hz. D-flow software¹⁶ (Motekforce

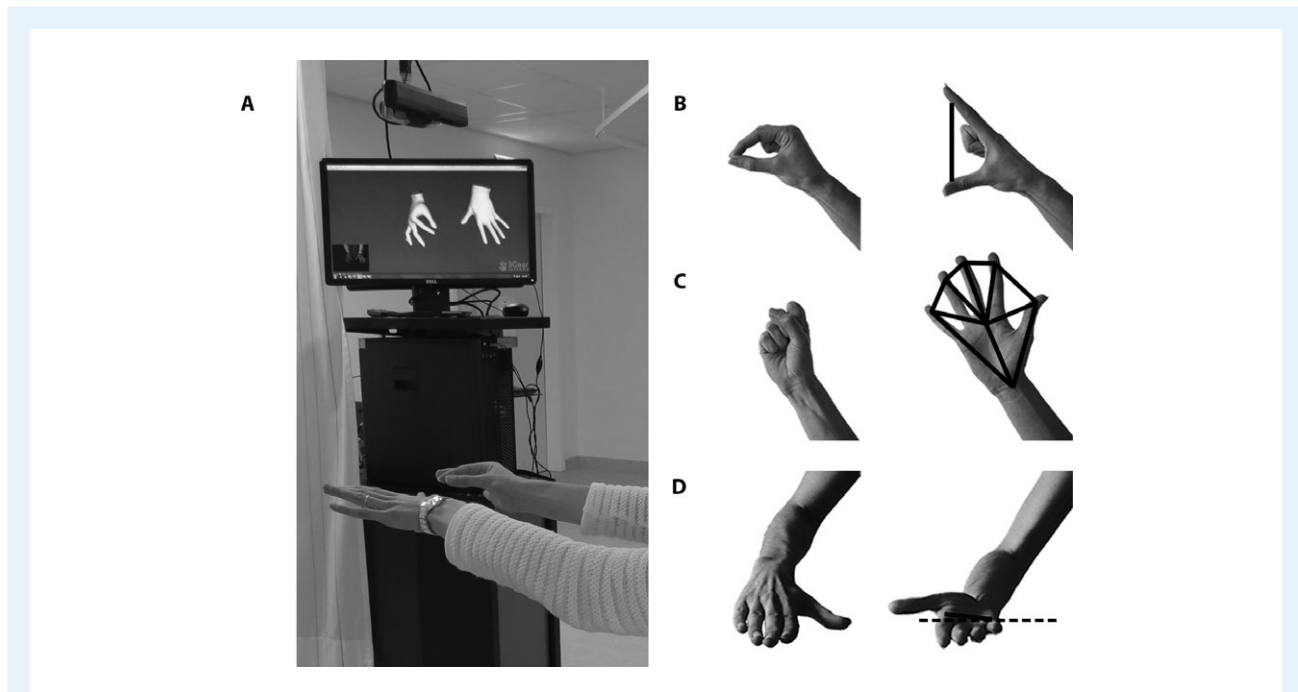


FIG. 1. Impression of (A) the optical hand tracking setup and (B–D) the 3 tasks: (B) repetitive finger tapping, an analysis based on the distance between distal joints of the thumb and index finger (in cm); (C) alternating hand movements, an analysis based on the “surface” of the hand (in cm^2); and (D) alternating forearm movements, an analysis based on the orientation of the line between the base of the index finger and the base of the pinky in the frontal plane (in degrees).

Link, Amsterdam, the Netherlands), expanded with a data fusion component (NCF; Noldus, Wageningen, the Netherlands), was used for controlling the experiment and data storage.

Throughout the experiment, participants sat upright in a chair with their feet supported. Before each task, instructions were presented on a 60-inch LED TV (LC-60LE652E; Sharp Electronics Europe Ltd., Usbridge, UK) placed approximately 1.5 m in front of the participant’s head. The task was demonstrated both on screen and live by the experimenter. Participants first performed a 10-second practice trial at comfortable speed with the widest possible amplitude, followed by a short break, then the actual 20-second measurement. For the actual measurement, participants were instructed to perform the movement as fast as possible and with the widest possible amplitude.

Patients performed the tasks with their most affected hand (i.e., the hand with the highest total score on the Items 3.4 through 3.6 of the motor evaluation subscale [part III] of the MDS-UPDRS). Controls were randomly assigned to perform the tasks with either their dominant ($N = 29$) or nondominant ($N = 28$) hand (as assessed using a Dutch version of the Edinburgh Handedness Questionnaire¹⁷). All participants were instructed to keep both hands in view of the sensor, holding their inactive hand still under the sensor with palm down and fingers extended and slightly spread, to improve detection of the contralateral, active hand. To ensure optimal hand detection, RFT was performed at an angle of approximately 45 degrees relative to the camera, and AHM was performed with the palm down. During the actual measurement, the estimated

hand model was checked in real time by the experimenter on a second screen, which was not visible to the participant. The task was repeated in case there were problems with hand detection. The order of the 3 tasks was randomized across participants. For participants in the “test-retest” study, the same order was kept between measurements.

Data Analysis

Occasionally, the software failed to fit a hand model. Data segments were considered representative of the participant’s performance and selected for analysis if they met the following requirements: a minimum duration of 5 seconds, a minimum of 80% of valid data points, and free from episodes of 4 or more consecutive missing data points. In approximately 76% of all trials, this yielded 1 valid data segment (average duration \pm standard deviation, 16.3 ± 4.2 seconds). In approximately 8% of the trials, multiple shorter segments (each >5 seconds) were identified. For these trials, outcome parameters were calculated per segment, and their weighted average was calculated (based on segment duration). Sixteen percent of all trials were excluded from further analyses because no segment met the requirements (the large majority concerned AFM) (Table 2).

Data analysis was performed using custom-made scripts in MATLAB (The MathWorks Inc., Natick MA; version R2016a; interested readers are invited to contact the corresponding author for the post-processing scripts). Data from selected segments were resampled to a uniformly distributed discrete time

TABLE 2 Overview of excluded trials and their distribution over experimental tasks

Experimental task	No. in study part	No. in analysis (%)	No. not in analysis		
			Task not performed	Technical issues	No segment
<i>Repetitive finger tapping</i>					
Cross-sectional					
PD	57	52 (91)	2	1	2
Control	57	53 (93)	0	1	3
Test-retest ^a					
PD	12	10 (83)	0/0	0/2	0/0
Control	12	11 (92)	0/0	0/0	0/1
<i>Alternating hand movements</i>					
Cross-sectional					
PD	57	52 (91)	2	2	1
Control	57	53 (93)	0	2	2
Test-retest ^a					
PD	12	8 (67)	0/0	0/2	0/2
Control	12	12 (100)	0/0	0/0	0/0
<i>Alternating forearm movements</i>					
Cross-sectional					
PD	57	40 (70)	2	4	11
Control	57	24 (42)	0	0	33
Test-retest ^a					
PD	12	6 (50)	0/0	3/2	1/2
Control	12	3 (25)	0/0	0/2	7/7

PD, Parkinson's disease.

^aReported values for the number not in analysis are for the first and second measurements, respectively (separated by /).

series (30 Hz) using linear interpolation before a low-pass filter was applied (fourth-order bidirectional Butterworth, 10 Hz). For RFT, the distance (cm) between the distal joints of the thumb and index finger was calculated (Fig. 1B). For AHM, the surface of the hand (cm²) was calculated as the total surface of 6 triangles, which were created by connecting 2 adjacent points (i.e., from the distal joints of the fingers and the wrist) with the approximate center of the hand (i.e., the mean of all joint positions) (Fig. 1C). For AFM, the orientation of the line between the base of the index finger and the base of the pinky was calculated in the frontal plane (in degrees), with 0 degrees indicating palm up, and -180 degrees indicating palm down (Fig. 1D).

Separate parameters quantified (progressive reductions in) speed, frequency, and amplitude and interruptions of rhythm. A peak-detection algorithm was used to detect individual movement cycles for the quantification of movement frequency (inverse of cycle duration) and amplitude (difference between the cycle's maximum excursion and the previous minimum). Within each trial, the average movement frequency (f_{mean}) and amplitude (A_{mean}) were calculated. Movement speed was quantified by means of the mean absolute velocity v_{mean} . Changes in frequency (Δf), amplitude (ΔA), and velocity (Δv) were obtained by linear regression over the time course of the trial. Interruptions of rhythm were quantified by means of *hesitations* (i.e., the surplus of times within a cycle that velocity changed sign, calculated as the number of velocity zero crossings minus the 2 reversals per cycle) and *halts* (i.e., the periods longer than the average duration of 2 cycles, based on the mean frequency of that trial, in which all data points fell within a range of 1.5 cm for RFT, 10 cm² for AHM), or 30 degrees for AFM. Halt duration was expressed as a percentage of the total duration of included data segments.

Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics 23.0 (IBM Corporation, Armonk, NY). Analyses were conducted separately for the different tasks. Normality curves were inspected, and Kolmogorov-Smirnov tests were used to assess whether the data were normally distributed.¹⁸ Values of normally distributed variables are presented as means \pm standard deviations, and all other values are presented as medians with interquartile ranges.

Cross-sectional analysis

Independent-samples *t* tests (for normally distributed variables) or Mann-Whitney *U* tests (for other variables) were used to compare outcomes between patients with PD and controls. For *t* tests, degrees of freedom were adjusted if the assumption of homogeneity of variance was violated.¹⁸ Values for Δf , ΔA , and Δv were submitted to 1-sample *t* tests (for normally distributed variables) or Wilcoxon signed-rank tests (for other variables) to determine whether they differed significantly from zero. Significance was set at $P < 0.05$. Within the PD group, Kendall's τ was used to examine correlations between the quantitative kinematic parameters and the score on the corresponding item of the MDS-UPDRS.

Test-retest and intra-trial reliability

Test-retest reliability for kinematic parameters was assessed by means of the intraclass correlation coefficient for absolute agreement (ICC_{A,1})¹⁹ based on all participants in the test-retest study (i.e., controls and patients with PD combined). With data from 24 participants, a good ICC of 0.65 would at least be detected

TABLE 3 Results of cross-sectional analysis

Experimental task	Control, N = 53	PD, N = 52	P value	Correlation with MDS-UPDRS item
Repetitive finger tapping (RFT)				
Velocity (v)				
v_{mean} , cm.s ⁻¹	24.9 ± 6.2	16.1 ± 8.0	<0.001 ^a	-0.37**
Δv , cm.s ⁻²	-0.22 (-0.43, 0.11) ^c	-0.29 (-0.51, -0.08) ^c	0.14 ^b	0.03
Frequency (f)				
f_{mean} , cycles.s ⁻¹	2.5 ± 0.9	2.2 ± 0.8	0.027 ^a	-0.24 *
Δf , cycles.s ⁻²	-0.003 (-0.01, 0.02)	-0.01 (-0.03, 0.01)	0.032 ^b	0.05
Amplitude (A)				
A_{mean} , cm	5.6 ± 1.6	4.1 ± 2.0	< 0.001 ^a	-0.16
ΔA , cm.s ⁻¹	-0.05 (-0.12, 0.01) ^c	-0.06 (-0.13, -0.01) ^c	0.791 ^b	0.04
Interruptions of rhythm				
Hesitations, no.cycle ⁻¹	0.4 (0.2-0.8)	0.8 (0.3-1.6)	0.007 ^b	0.31*
Halts, % of time	0 (0-0)	0 (0-21.6)	<0.001 ^b	0.27*
Clinical assessment				
Total MDS-UPDRS-III	—	36.2 ± 16.1	—	—
RFT item 3.4	—	2 (1-3)	—	—
Alternating hand movements (AHM)				
Velocity (v)				
v_{mean} , cm.s ⁻¹	110.8 ± 37.3	84.6 ± 31.3	<0.001 ^a	-0.41**
Δv , cm.s ⁻²	-0.2 ± 2.0	-0.7 ± 2.5 ^c	0.24 ^a	0.11
Frequency (f)				
f_{mean} , cycles.s ⁻¹	1.8 ± 0.6	1.5 ± 0.4	0.005 ^a	-0.39**
Δf , cycles.s ⁻²	0.0004 ± 0.03	-0.005 ± 0.03	0.42 ^a	0.05
Amplitude (A)				
A_{mean} , cm	34.2 ± 12.2	29.5 ± 9.4	0.03 ^a	-0.17
ΔA , cm.s ⁻¹	-0.21 ± 0.62 ^c	-0.35 ± 0.58 ^c	0.23 ^a	0.11
Interruptions of rhythm				
Hesitations, no.cycle ⁻¹	3.4 (2.9-4.4)	3.7 (3.2-4.5)	0.33 ^b	0.35*
Halts, % of time	0 (0-0)	0 (0-0)	0.04 ^b	0.10
Clinical assessment				
Total MDS-UPDRS-III	—	36.5 ± 16.6	—	—
AHM item 3.5	—	1 (1-2)	—	—

Δ , change; PD, Parkinson's disease; MDS-UPDRS-III, International Parkinson and Movement Disorders Society-sponsored revision of the Unified Parkinson's Disease Rating Scale.

^aThese P values were determined using an independent-sample t test and parameter values are presented as the mean ± standard deviation.

^bThese P values were determined using the Mann-Whitney U test and parameter values are presented as the median (interquartile range).

^cThese values are significantly different from 0 (using the Wilcoxon signed-rank test; $P < 0.05$). Asterisks indicate significant correlations between kinematic parameters and the score on the corresponding MDS-UPDRS item: * $P < 0.05$; ** $P < 0.01$.

as being moderate (>0.4)²⁰ (expected ICC = 0.65; 95% confidence interval width = 0.46). In addition, we assessed intra-trial reliability in all controls who were included in the cross-sectional analysis to evaluate reliability in a situation where variation in performance was expected to be minimal. Assuming that, in controls, task performance was constant over the time course of a trial, we calculated the ICC_{A,1} of outcome parameters derived from the first and last 5 seconds of each trial. ICC_{A,1} values above 0.40 were considered fair, values above 0.60 were considered good, and values above 0.75 were considered excellent.²¹ ICC_{A,1} values were complemented by mean differences and precision values obtained with a Bland-Altman analysis (i.e., the bias and limits of agreement).²²

Results

OHT was successful in evaluating the RFT and AHM tasks in 94% of all participants who performed these tasks. Failure was caused by technical problems or the lack of a data segment that met the requirements; for the AFM task, no data segment could be selected in 44 participants because of problems with hand

recognition during fast alternation between pronation and supination (Table 2). This was particularly evident in the control group, in which only 42% of participants had more than 10 seconds of data available for analysis. Therefore, AFM was excluded from further analysis. Two patients with PD were unable to perform the tasks due to fatigue (caused by the larger study protocol of which this experiment was part). An overview of excluded trials and their distribution over experimental tasks is presented in Table 2.

Cross-sectional Analysis

The patients with PD performed the RFT task with significantly reduced velocity, frequency, and amplitude and with more hesitations and halts than controls (Table 3). In addition, patients displayed a more pronounced decrement of frequency than controls (i.e., Δf was more negative but was not significantly smaller than zero). For the patients with PD, higher scores on the MDS-UPDRS item for RFT were associated with reduced velocity and frequency, more hesitations, and longer halt duration (Table 3).

Patients with PD performed the AHM task with significantly reduced velocity, frequency, and amplitude and with more halts than controls (Table 3). For the patients with PD, higher scores on the MDS-UPDRS item for AHM were associated with lower velocity and frequency and with more hesitations (Table 3).

Test-Retest and Intra-trial Reliability

The self-reported function of the arm and hand was rated 5 (i.e., “unchanged” compared with the first measurement) by all controls and by 11 of 12 the patients with PD who participated in the test-retest study; 1 patient with PD reported a 1-point (out of a possible 5-point) reduction.

For the RFT task, test-retest reliability was fair for amplitude, good for velocity, and excellent for halts, with relatively small bias and narrow limits of agreement for these parameters (Table 4). Test-retest reliability was poor for frequency and for changes of velocity (Δv), frequency (Δf), and amplitude (ΔA), with $ICC_{A,1}$ values <0.20 in combination with large bias and/or wide limits of agreement relative to the average parameter values. For the AHM task, fair-to-good test-retest reliability was observed for frequency, with relatively small bias and narrow limits of agreement (Table 4), whereas test-retest reliability of the other kinematic variables was poor (all $ICC_{A,1}$ values ≤ 0.25).

The intra-trial $ICC_{A,1}$ (based on outcome parameters derived from the first and last 5 seconds of each trial in controls) was excellent for frequency, good for velocity, and fair for amplitude and hesitations on the RFT task (Table 4). Excellent intra-trial reliability was observed for velocity, frequency, and amplitude on the AHM task (Table 4), whereas reliability of hesitations was fair on this task.

Discussion

OHT used for objective assessment of the movement components of bradykinesia of the upper extremity in PD showed promise with regard to finger tapping (RFT) and hand opening and closing (AHM): key kinematic parameters differed between patients with PD and healthy controls and were correlated with corresponding scores on the MDS-UPDRS, indicating that characteristic features of PD were successfully captured in objective parameters.

Test-retest reliability was fair or good for velocity, amplitude, and halts on the RFT item and for frequency on the AHM item. The test-retest reliability of the other kinematic variables, such as amplitude and halts on the AHM item, however, was poor, with $ICC_{A,1}$ values below 0.25. For comparison, test-retest reliability has been reported as poor to fair^{23–25} or good^{7,26} for individual bradykinesia items on the MDS-UPDRS (ICC values ranging from 0.21 to 0.80) and as fair or good for items

TABLE 4 Results for test-retest and intra-trial reliability

Experimental task	Test-retest reliability and agreement ^a		Intra-trial reliability and agreement: First 5 s vs. last 5 s of trial in controls ^b	
	$ICC_{A,1}$	Bias (limits of agreement)	$ICC_{A,1}$	Bias (limits of agreement)
<i>Repetitive finger tapping (RFT)</i>				
Velocity (v)				
v_{mean} , $cm \cdot s^{-1}$	0.72	1.1 (−2.1, 4.3)	0.71	0.5 (−0.8, 1.9)
Δv , $cm \cdot s^{-2}$	0.17	−0.01 (−0.3, 0.3)	—	—
Frequency (f)				
f_{mean} , $cycles \cdot s^{-1}$	0.12	0.4 (−0.1, 0.8)	0.91	0.08 (−0.03, 0.19)
Δf , $cycles \cdot s^{-2}$	0.19	−0.02 (−0.05, 0.01)	—	—
Amplitude (A)				
A_{mean} , cm	0.40	−0.08 (−1.0, 0.9)	0.48	−0.02 (−0.7, 0.7)
ΔA , $cm \cdot s^{-1}$	−0.01	0.05 (−0.03, 0.12)	—	—
Interruptions of rhythm				
Hesitations, $no \cdot cycle^{-1}$	0.18	−0.4 (−1.0, 0.1)	0.42	−0.04 (−0.3, 0.2)
Halts, % of time	0.90	−3.1 (−8.6, 2.3)	—	—
<i>Alternating hand movements (AHM)</i>				
Velocity (v)				
v_{mean} , $cm \cdot s^{-1}$	0.22	−7.8 (−30.0, 14.3)	0.76	4.8 (−4.2, 13.8)
Δv , $cm \cdot s^{-2}$	0.17	−0.8 (−2.3, 0.7)	—	—
Frequency (f)				
f_{mean} , $cycles \cdot s^{-1}$	0.59	0.1 (−0.1, 0.4)	0.81	−0.004 (−0.1, 0.1)
Δf , $cycles \cdot s^{-2}$	−0.01	0.01 (−0.02, 0.04)	—	—
Amplitude (A)				
A_{mean} , cm	0.25	−4.2 (−9.8, 1.4)	0.82	−3.8 (−6.5, −1.0)
ΔA , $cm \cdot s^{-1}$	0.07	−0.02 (−0.6, 0.6)	—	—
Interruptions of rhythm				
Hesitations, $no \cdot cycle^{-1}$	0.11	−0.2 (−1.0, 0.5)	0.41	−0.7 (−1.6, 0.2)
Halts, % of time	0.21	−1.1 (−3.3, 1.1)	—	−1.6 (−3.9, 0.6)

Δ , change; $ICC_{A,1}$, intraclass correlation coefficient for absolute agreement.

^aBased on $N = 21$ for RFT and $N = 20$ for AHM.

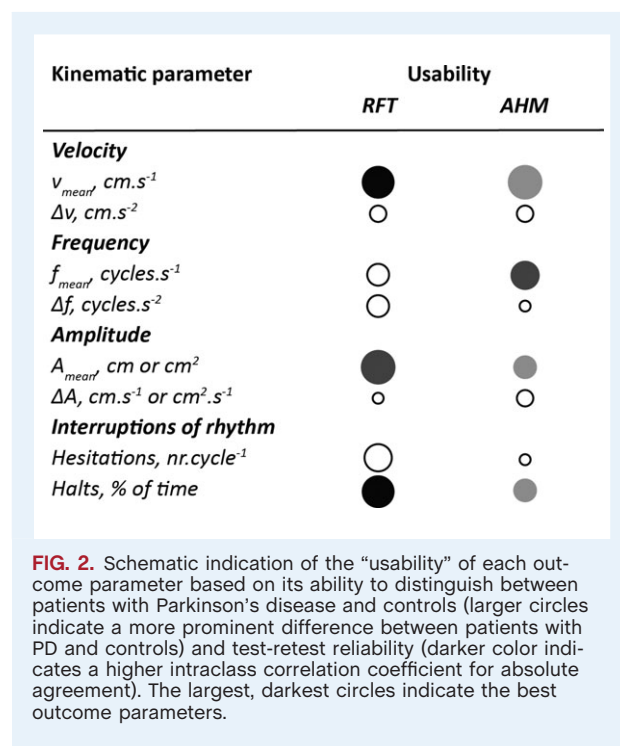
^bBased on $N = 53$ for RFT and AHM. Reported variables were assumed to be constant over the time course of a trial in controls. Variables Δv , Δf , and ΔA were excluded from this analysis, because the 5-second segments were considered too short to quantify changes in velocity, frequency, and amplitude.

on the Modified Bradykinesia Rating Scale (ICC values ranging from 0.45 to 0.83).^{3,7} To date, reports on test-retest reliability of technology-based sensors for bradykinesia assessment are scarce. Heldman et al.⁷ reported a within-session ICC for consistency of 0.63 for rhythm and 0.94 for speed and amplitude of RFT using IMU technology. These values are slightly higher than the “stricter” ICC values for absolute agreement obtained in the present study (i.e., 0.90 for halts, 0.72 for velocity, and 0.40 for amplitude), where the time interval between 2 measurements was considerably longer.

Three possible reasons can be put forward for the disappointing test-retest reliability of several kinematic parameters in the current study: (1) instrument-based error, (2) change in health status between measurements, and (3) change in movement strategy. With regard to the potential role of instrument-based error, it is unfortunate that there are no reports on the validity and accuracy of the 3D hand model used for OHT. Also, in the current study, the instrument-based error could not be evaluated by means of a direct comparison of OHT with a *gold-standard* reference system, because performance of the OHT algorithm would be compromised by the attachment of markers or sensors to the hand. However, the high intra-trial reliability (i.e., ICC_{A,1} values for outcome parameters derived from the first and last 5 seconds of each trial in controls) for both RFT and AHM indicates that the setup measured similar kinematics when performance was assumed to be constant. This finding strongly suggests that the low test-retest reliability for amplitude, velocity, frequency, and hesitations observed in this study is not likely caused by instrument-based error. It is also not plausible that the low test-retest reliability is attributable to changes in the participants' health status, given the 1-week interval between measurements and the fact that most participants reported no change in function of the arm and hand. Moreover, the same order of tasks was retained so that the effects of fatigue would be similar, and measurements were performed at the same time of the day (and thus at the same time after medication intake) to minimize the effects of motor fluctuations in patients. Despite these measures, it cannot be fully ruled out that non-constant effects of PD medication in patients might have led to slight underestimation of test-retest reliability and correlations with the MDS-UPDRS scores. However, it seems more likely that participants applied different strategies during the 2 measurement sessions, because the task instructions were open to different interpretations (i.e., according to the MDS-guidelines, the movement should be performed *as wide and as fast as possible*). This hypothesis of strategy change is supported by a strong negative correlation between the *changes in amplitude and frequency from the first to the second measurement* for RFT ($r = -0.69$), but not for AHM ($r = -0.07$) (see Fig. S1). The negative correlation for RFT indicates that participants may have put more emphasis on a high movement frequency (at the cost of smaller amplitude) during one measurement and more emphasis on movement amplitude (at the cost of lower movement frequency) during the other measurement. This strategy change between measurements would result in low test-retest reliability of these individual variables.

Because velocity reflects a combination of amplitude and frequency, one might expect that velocity would be less susceptible to changes in strategy than amplitude and frequency separately, which was indeed the case for RFT. For the AHM task, a low ICC_{A,1} was also observed for velocity, possibly because velocity was more strongly associated with amplitude ($r = 0.88$) than with frequency ($r = 0.45$), whereas amplitude in particular appeared to be subject to test-retest variation (as was indicated by lower ICC_{A,1} values) (Table 4). In line with this finding, velocity and halts appeared to be the most useful variables for quantification of RFT, whereas frequency appeared most the useful for quantification of AHM (Fig. 2). These findings suggest that quantification of these movement components is recommendable for monitoring patients or evaluating treatment effects.

Against this background, the absence of strong correlations between kinematic parameters and clinical scores may not be surprising, particularly because clinical scores may reflect different combinations of multiple movement components. Notably, for technological developments, a strong agreement between quantitative measures and subjective clinical scores, the so-called “clinimetric validation pitfall”, is not necessarily required. A “perfect” objective measurement should have a complicated quantitative match with an “imperfect” subjective measurement (Espay et al.,⁵ p. 1279). Technology-based outcome measures should show improved resolution, reliability, and/or responsiveness relative to the widely used clinical rating scales. Our findings indicate that individual parameters are susceptible to variations in movement strategy, which may suggest that task performance is better captured by a combination of kinematic



outcome parameters. We have explored several options in this regard. The strong covariance between velocity, frequency, and amplitude did not allow an analysis of the combination of variables by either logistic regression models, principal component analysis, or linear discriminant analysis. Therefore, as an alternative, we explored the potential of “sum scores” based on cutoff values for each outcome parameter (see the online Supporting Materials and Tables S1–S4). This seemed promising for the identification of severely affected patients with PD, but it could not distinguish between mildly affected patients with PD and healthy participants.

OHT for the objective quantification of bradykinesia has a great advantage in terms of patient-friendly assessment, because it does not require preparation time or attachment of sensors, which can affect movements of the fingers and/or hands. The results seem to be unaffected by slight variations in orientation of the hand relative to the sensor, and the sampling rate appears to be sufficient for evaluating RFT and AHM. Current limitations in post-processing (e.g., the manual correction of peak-detection errors that was required in some cases) need to be further improved and automated in the future. The system appeared insufficiently able to evaluate the AFM item, because rapid alterations between pronation and supination hampered hand recognition. Consequently, this task may only discriminate the more affected, slower moving patients with PD (see Table 2). Currently, approaches that are not based on algorithms for hand recognition in real time (e.g., using an IMU,^{3,4,27,28} gyroscope,^{29,30} or accelerometer³¹) appear to be more suitable for the quantification of pronation and supination.

In conclusion, OHT yielded promising results for the assessment of bradykinesia of the upper extremity in PD. Separate quantification of velocity and halts for RFT and frequency for AHM is recommended for monitoring patients or evaluating treatment effects. Future work should address approaches to combining kinematic parameters to better understand and reduce the measurement's susceptibility to changes in strategy and to enhance responsiveness to changes in the patient's actual health status (e.g., induced by dopaminergic medication).

Authors' Roles

1. Research Project: A. Conception, B. Organization, C. Execution; 2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique; 3. Manuscript Preparation: A. Writing the First Draft, B. Review and Critique.

P.J.M.B.: 1A, 1B, 1C, 2A, 2B, 3A

J.M.: 1A, 1B, 2C, 3B

J.H.d.G.: 1A, 2C, 3B

C.G.M.M.: 1A, 1C, 2C, 3B

J.J.v.H.: 1A, 3B

Acknowledgments

We thank Leo Casteel, MSc, Rosanne van Tol, MSc, and Elma Ouweland, MSc, for their help in collecting the data.

Disclosures

Ethical Compliance Statement: We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

Funding Sources and Conflicts of Interest: This work is part of the research program IMDI Neurocontrol (NeurAS project 10-10400-98-008) financed by the Netherlands Organisation for Health Research and Development and the research program Technology in Motion (TIM) (628.004.001), which is financed by the Netherlands Organisation for Scientific Research. The authors declare that there are no conflicts of interest relevant to this work.

Financial Disclosures for the past 12 months: P.J.M. Bank is employed by the Leiden University Medical Center based on funding by the Netherlands Organisation for Scientific Research (NWO) (628.004.001), and received a grant from the “Stichting Wetenschapsfonds Dystonie.” J. Marinus is employed by the Leiden University Medical Center, received a grant from the “Stichting ParkinsonFonds” and received honoraria from the International Parkinson and Movement Disorders Society. J.H. de Groot is employed by the Leiden University Medical Center, received a grant from the Technology Foundation STW, and is member of the advisory board of the Netherlands Society of Rehabilitation Medicine-Research Commission. C.G.M. Meskers is employed by the VU University Medical Center, received grants from the Technology Foundation STW and the Netherlands Organisation for Health Research and Development, received a personal grant from the Dutch Brain Foundation, and is member of the scientific advisory board of Motekforce Link. J.J. van Hilten is employed by the Leiden University Medical Center; he received grants from the Alkemade-Keuls Foundation, the “Stichting ParkinsonFonds,” the Parkinson Vereniging, the Dystonie Vereniging and the Netherlands Organisation for Health Research and Development.

Orcid

Paulina J.M. Bank  <http://orcid.org/0000-0002-3127-398X>

References

1. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2015;30:1591–1601.
2. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008;23:2129–2170.
3. Heldman DA, Giuffrida JP, Chen R, et al. The Modified Bradykinesia Rating Scale for Parkinson's disease: reliability and comparison with kinematic measures. *Mov Disord* 2011;26:1859–1863.
4. Espay AJ, Giuffrida JP, Chen R, et al. Differential response of speed, amplitude, and rhythm to dopaminergic medications in Parkinson's disease. *Mov Disord* 2011;26:2504–2508.
5. Espay AJ, Bonato P, Nahab FB, et al. Technology in Parkinson's disease: challenges and opportunities. *Mov Disord* 2016;31:1272–1282.

6. Kishore A, Espay AJ, Marras C, et al. Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: differential effects on bradykinesia. *Mov Disord* 2007;22:328–333.
7. Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP. Clinician versus machine: reliability and responsiveness of motor endpoints in Parkinson's disease. *Parkinsonism Relat Disord* 2014;20:590–595.
8. Sánchez-Ferro Á, Elshehabi M, Godinho C, et al. New methods for the assessment of Parkinson's disease (2005 to 2015): a systematic review. *Mov Disord* 2016;31:1283–1292.
9. Maetzel W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. *Mov Disord* 2013;28:1628–1637.
10. Del Din S, Godfrey A, Mazzà C, Lord S, Rochester L. Free-living monitoring of Parkinson's disease: Lessons from the field. *Mov Disord* 2016;31:1293–1313.
11. Goetz CG, Stebbins GT, Wolff D, et al. Testing objective measures of motor impairment in early Parkinson's disease: feasibility study of an at-home testing device. *Mov Disord* 2009;24:551–556.
12. Homann CN, Suppan K, Wenzel K, et al. The Bradykinesia Akinesia Incoordination Test (BRAIN TEST), an objective and user-friendly means to evaluate patients with Parkinsonism. *Mov Disord* 2000;15:641–647.
13. Taylor Tavares AL, Jefferis GS, Koop M, et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Mov Disord* 2005;20:1286–1298.
14. Memedi M, Khan T, Grenholm P, Nyholm D, Westin J. Automatic and objective assessment of alternating tapping performance in Parkinson's disease. *Sensors (Basel)* 2013;13:16965–16984.
15. Gibb W, Lees A. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J Neurol Neurosurg Psychiatry* 1988;51:745–752.
16. Geijtenbeek T, Steenbrink F, Otten B, Even-Zohar O. D-flow: immersive virtual reality and real-time feedback for rehabilitation. In: Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry; December 22–12, 2011; Hong Kong, China. New York, NY: Association for Computing Machinery; 2011:201–208.
17. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113.
18. Field A. *Discovering Statistics Using SPSS*. London, UK: Sage Publications; 2009.
19. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033–1039.
20. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. *Stat Methods Med Res* 2004;13:251–271.
21. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 1994;6:284–290.
22. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–160.
23. Camicioli R, Grossmann SJ, Spencer PS, Hudnell K, Anger WK. Discriminating mild parkinsonism: methods for epidemiological research. *Mov Disord* 2001;16:33–40.
24. Marinus J, Visser M, Stiggelbout AM, et al. A short scale for the assessment of motor impairments and disabilities in Parkinson's disease: the SPES/SCOPA. *J Neurol Neurosurg Psychiatry* 2004;75:388–395.
25. Bennett DA, Shannon K, Beckett LA, Goetz C, Wilson R. Metric properties of nurses' ratings of parkinsonian signs with a modified Unified Parkinson's Disease Rating Scale. *Neurology* 1997;49:1580–1587.
26. Siderowf A, McDermott M, Kiebertz K, et al. Test-retest reliability of the unified Parkinson's disease rating scale in patients with early Parkinson's disease: results from a multicenter clinical trial. *Mov Disord* 2002;17:758–763.
27. Mera TO, Heldman DA, Espay AJ, Payne M, Giuffrida JP. Feasibility of home-based automated Parkinson's disease motor assessment. *J Neurosci Methods* 2012;203:152–156.
28. Piro NE, Piro LK, Kassubek J, Blechschmidt-Trapp RA. Analysis and visualization of 3D motion data for UPDRS rating of patients with Parkinson's disease. *Sensors (Basel)* 2016;16: pii:E930.
29. Tabbal SD, Ushe M, Mink JW, et al. Unilateral subthalamic nucleus stimulation has a measurable ipsilateral effect on rigidity and bradykinesia in Parkinson disease. *Exp Neurol* 2008;211:234–242.
30. Jun JH, Kim JW, Kwon Y, et al. Quantification of limb bradykinesia in patients with Parkinson's disease using a gyrosensor—improvement and validation. *Int J Precis Eng Manuf* 2011;12:557–563.
31. Jia X, Duroseau N, Chan V, et al. Objective quantification of upper extremity motor functions in Unified Parkinson's Disease Rating Scale Test. *Conf Proc IEEE Eng Med Bio Soc* 2014;2014:5345–5348.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Cutoff values for repetitive finger tapping (RFT) and alternating hand movements (AHM).

Table S2. Sum score results for repetitive finger tapping (RFT) and alternating hand movements (AHM).

Table S3. Correlations of between-measurement changes in kinematic parameters and in calculated scores for repetitive finger tapping (RFT) and alternating hand movements (AHM)

Table S4. Sensitivity and specificity of cutoff scores for discriminating between patients with Parkinson's disease and controls

Appendix S1. “Sum scores” to combine outcome parameters.

Figure S1. Correlation between the changes in amplitude and frequency from the first (test) to the second (retest) measurement.