

Detecting and monitoring the symptoms of Parkinson’s disease using smartphones: a pilot study

S. Arora, PhD^{1,2}; V. Venkataraman, BSE³; S. Donohue⁴; K.M. Biglan, MD, MPH⁵; E.R. Dorsey MD, MBA^{5,6}; M.A. Little, PhD^{1,7}

1. Nonlinearity and Complexity Research Group, Aston University, Birmingham, United Kingdom
2. Somerville College, University of Oxford, Oxford, United Kingdom
3. Duke University School of Medicine, Durham, NC, United States
4. Johns Hopkins University, Baltimore, MD, United States
5. Department of Neurology, University of Rochester Medical Center, Rochester, NY, United States
6. Center for Human Experimental Therapeutics, University of Rochester Medical Center, Rochester, NY, United States
7. Media Lab, Massachusetts Institute of Technology, Boston, MA, United States

Corresponding author

Ray Dorsey
University of Rochester Medical Center
265 Crittenden, Box CU420694
Rochester, NY 14642
585.276.6824 (phone)
585.461.4594 (fax)
ray.dorsey@chert.rochester.edu

Counts

Title: 93 characters
Abstract: 243 words
Manuscript: 1637 words
Tables and Figures: 2 + 4 supplemental tables
References: 12

Keywords

Parkinson’s disease; motor assessment; diagnostic test assessment; smartphone; speech

Study sponsorship

This study received no external funding

ABSTRACT

Background: Remote, non-invasive and objective tests that can be used to support expert diagnosis for Parkinson's disease (PD) are lacking.

Methods: Participants underwent baseline in-clinic assessments, including the Unified Parkinson's Disease Rating Scale (UPDRS), and were provided smartphones with an Android operating system that contained a smartphone application that assessed voice, posture, gait, finger tapping, and response time. Participants then took the smart phones home to perform the five tasks four times a day for a month. Once a week participants had a remote (telemedicine) visit with a Parkinson disease specialist in which a modified (excluding assessments of rigidity and balance) UPDRS performed. Using statistical analyses of the five tasks recorded using the smartphone from 10 individuals with PD and 10 controls, we sought to: (1) discriminate whether the participant had PD and (2) predict the modified motor portion of the UPDRS.

Results: Twenty participants performed an average of 2.7 tests per day (68.9% adherence) for the study duration (average of 34.4 days) in a home and community setting. The analyses of the five tasks differed between those with Parkinson disease and those without. In discriminating participants with PD from controls, the mean sensitivity was 96.2% (SD 2%) and mean specificity was 96.9% (SD 1.9%). The mean error in predicting the modified motor component of the UPDRS (range 11-34) was 1.26 UPDRS points (SD 0.16).

Conclusion: Measuring PD symptoms via a smartphone is feasible and has potential value as a diagnostic support tool.

INTRODUCTION

Accurate diagnosis of Parkinson's disease (PD) and other neurological disorders is critical for care and research. However, clinical diagnostic accuracy by a neurologist is limited as up to 20% of cases may be misdiagnosed by clinicians when compared to post mortem pathology.¹ New imaging modalities, such as SPECT scans, can assist, but their utility is limited by cost, time, and availability.

Previous studies have objectively characterized pathological changes in motor symptoms² such as voice production (using a microphone),³ posture and gait (using accelerometers),⁴ tremor (using finger tapping tasks),⁵ and cognitive performance (using reaction times)⁶ in PD.

Smartphones, which are increasingly inexpensive and ubiquitous, have built-in voice recorders, accelerometers, and touch screens, and can potentially be used to objectively measure PD severity symptoms. For continuous objective motor assessment of PD (such as, freezing of gait, balance deficits and falls), the potential of wearable sensors has been discussed recently.⁷

Although smartphones are more ubiquitous than wearable sensors, their use by individuals in home settings to differentiate those with PD from controls and to evaluate disease severity has not been previously investigated.

METHODS

Study design and study participants

Individuals with PD diagnosed clinically by a movement disorder specialist and control participants were recruited from an academic movement disorder clinic (Johns Hopkins) and all participants provided informed consent. The study was reviewed and approved by the Johns

Hopkins Institutional Review Board. Control participants were spouses, caregivers, relatives, or colleagues of an individual with PD and did not have any known neurological disorder. All participants were provided identical LG Optimus S smartphones (cost: \$30 each) with Android OS (Google, Inc.) loaded with specialized software developed by the team (M.A.L.)

(Supplemental Figure 1).

The smartphone software prompted participants to perform the following short tests (less than 5 minutes): (1) (*voice test*) say the sustained phonation ‘aaah’ for as long and as steadily as possible; (2) (*posture test*) stand upright unaided for thirty seconds; (3) (*gait test*) walk twenty steps forward, turn around, and return back to the starting position; (4) (*finger tapping test*) tap the screen alternately keeping a regular rhythm; and (5) (*reaction time test*) press and hold the on-screen button as soon as it appears and release it as soon as it disappears.

At an in-person baseline visit, participants were given the smartphone, trained how to use the software application, and instructed how to upload phone data via Wi-Fi to a secure database. The participants were asked to conduct the above specified tests four times daily: just before taking their first (morning) dose of levodopa (or in one case, rasagiline), one hour later, mid-afternoon, and before going to bed. Demographic information, medical history, and a current medication list were also collected in addition to completing the motor portion of the UPDRS,⁸ and the Parkinson Disease Questionnaire 39 (PDQ-39), a disease-specific quality of life instrument. The specialists also conducted a remote assessment of the modified UPDRS⁸ (that excluded assessments of rigidity and balance) via web-based video conferencing once per week for four weeks at approximately the same time each week. To map the summary measures to UPDRS, we derived the UPDRS values corresponding to each summary measure using piecewise linear interpolation, going exactly through the actual UPDRS assessments.³

Statistical analysis

Voice impairment may be one the earliest indicators of PD⁹ and is typically characterized by reduced loudness, breathiness, roughness, monopitch, and exaggerated vocal tremor. We extracted a range of summary measures that quantify roughness, aperiodicity, aeroacoustic noise in voice (due to incomplete vocal fold closure), and placement of the articulators (mouth, teeth, tongue and lips), some of which are known to be affected in PD.³

For posture and gait, using the three dimensional accelerometry time traces we calculated different summary measures, ranging from basic statistics, such as mean, standard deviation, and median acceleration, to more sophisticated measures, such as the Teager-Kaiser energy operator and detrended fluctuation analysis, that characterize changes in body motion.⁴

For finger tapping, we recorded the screen pixel position (x, y coordinates) and the time of finger touch. Finger tapping tasks have been used to assess arrhythmokinesis, which can include hastening, faltering or freezing in the tapping pattern in PD.⁵ Using this finger tap data, we developed novel summary measures that quantify the tapping speed, rhythm, inter-tap interval, fatigue, and tremor.

For the reaction time test, we measured the elapsed time between the stimulus (appearance/disappearance of a screen button) and the corresponding response (press/release of the button). This type of reaction time is also referred to as non-cued simple reaction time and is linked with pre-movement central neural processing.⁶ Using this reaction time data, we calculated a range of different summary measures such as mean, coefficient of variation, correlation, and fatigue.

Supplementary Tables 1, 2, 3 and 4 provides an exhaustive list of the summary measures extracted using the voice, posture, gait, finger tapping, and reaction time data, respectively, along with a brief description of each.

The study's overall methodological approach is summarized in **Figure 1**. Using random forests,¹⁰ a powerful statistical machine learning technique that is commonly used to separate generic data into several different classes, we sought to discriminate participants with PD from controls and predict the modified motor UPDRS scores, solely on the basis of the summary measures derived from the smartphone recordings. Statistical analysis of the smartphone data was performed using the Matlab® software (version 2013a).

Validation

To validate the method, we used a standard 10-fold, randomized cross-validation method with 100 repetitions. This validation method involves repetitive splitting of the dataset into non-overlapping “training” and “validation” sets. In total, we identified 1772 recordings that were of decent quality for the analysis. The training set (consisting of 90% of randomly selected smartphone recordings, n=1595 recordings) allows the random forest method to find discriminating patterns in the summary measures. The validation set (consisting of the remaining 10% of recordings, n=177 recordings) is used to assess the method's accuracy; effectively, the random forest method is blinded to these data during training. This randomized selection of training and validation sets is repeated 100 times, leading to different accuracy on each repetition. The mean and standard deviation of the accuracy is then computed over all 100 repetitions. The actual and interpolated UPDRS scores were used in the validation. Furthermore,

we performed two statistical tests against the null hypothesis that the discrimination results obtained were no better than randomized determinations of whether one has PD.

RESULTS

Twenty participants (10 individuals with PD and 10 controls) enrolled and completed the study, and their baseline characteristics are provided in **Table 1**. The 20 participants performed an average of 2.7 tests per day (68.9% adherence) for the study's duration (average of 34.4 days). Those with PD performed an average 2.6 tests per day, and controls completed 2.8 tests per day.

The summary measures differed between participants with PD and controls. Using random forests, the mean sensitivity was 96.2% (SD 2%) and mean specificity was 96.9% (SD 1.9%) in discriminating participants with PD from controls. These sensitivity and specificity results differed from comparable results obtained from randomized predictions about which participants had PD or were controls ($p < 0.001$, two-sided Kolmogorov-Smirnov test).

To predict ongoing symptom severity, our final objective was to replicate the clinician's symptom severity assessment as closely as possible, using only the summary measures. The performance of random forest in predicting symptom severity was assessed using the mean absolute error. The average mean absolute error in predicting the modified motor UPDRS (which ranged from 11-34, mean 19.6, SD 6.7) was 1.26 UPDRS points (SD 0.16). The average mean absolute error in predicting the modified motor UPDRS for female participants (UPDRS range 14-29, mean 22.1, SD 4.0) was 1.02 UPDRS points (SD 0.2), while for male participants (UPDRS range 11-34, mean 20.3, SD 5.2), the average mean absolute error was 1.29 (SD 0.2). The UPDRS errors were lower than the inter-rater variability of the UPDRS (4-5 points).¹¹

DISCUSSION

In this pilot study, we demonstrated that consumer-grade smartphones can be used to accurately differentiate individuals with PD from age-matched controls and potentially predict disease severity, which could be used to monitor disease progression. While the results remain to be confirmed in larger scale studies, they open the door for smartphones to be potentially used as diagnostic and disease management support devices.

Future analyses will quantify response to treatment and intra-day symptom variation, which are often impractical to perform in most clinical settings. Furthermore, future assessments could passively record accelerometry time traces continuously in the background, an approach which has shown good success in a recent study.²

Despite this promise, considerable work remains. This study was limited by size ($n = 20$ participants) and by the range of PD symptom severity (individuals with mild to moderate PD). The characteristics (e.g., age, sex) of the control participants did not match well those of the participants, which will need to be addressed in future studies. Larger-scale assessments of individuals earlier in their disease course compared to individuals with other disorders that cause changes in voice, impairment in cognitive skills, or result in tremor/gait deficits will better assess the discriminatory potential of such smartphone applications. Importantly, this study did not include individuals with other parkinsonian or tremor disorders that may be more difficult to differentiate from PD. This study was also limited by the variability in the home conduct of the tests and the remote assessments of the UPDRS.

Nearly two centuries after the seminal description of PD and five decades after the development of a highly effective, inexpensive therapy, many individuals in the world with PD have not been diagnosed or received treatment. Simple, inexpensive, widely available

technologies are needed to facilitate diagnosis and care. As this study demonstrates, smartphones – whose numbers are projected to reach 3 billion by 2017¹² – may represent an effective tool for detection, assessment, and potentially care, of this increasingly common disorder.

ACKNOWLEDGEMENTS

The authors wish to thank Andong Zhan, for his contribution towards the development of smartphone application used in this study, Reham Bedawy, for her help in identifying relevant literature for this work, and Ervin Sejdić for providing us the code for a few summary measures used for gait analysis. We also extend our sincere gratitude to all the individuals who participated in this study and made this research possible.

Financial Disclosures of all Authors for the Past Year

Dr. Siddharth Arora receives research support from The Michael J. Fox Foundation for Parkinson's Research.

Mr. Vinayak Venkataraman has no relevant financial disclosures.

Mr. Donohue has no relevant financial disclosures.

Dr. Kevin Biglan received compensation for consulting activities from Lundbeck and UCB, research support from the Davis Phinney Foundation, National Institute of Neurological Disorders and Stroke, Michael J. Fox Foundation, National Parkinson Foundation, Patient-Centered Outcomes Research Institute, Huntington Disease Society of America, and Lundbeck.

Dr. Ray Dorsey has received compensation for consulting activities from Amgen, Clintrex, Lundbeck, mc10, Medtronic, and the National Institute of Neurological Disorders and Stroke, research support from Davis Phinney Foundation, Great Lakes Neurotechnologies, Huntington Study Group, Lundbeck, Michael J. Fox Foundation, Patient-Centered Outcomes Research Institute, Prana Biotechnology, Sage Bionetworks, stock options from Grand Rounds, and compensation for expert testimony.

Dr. Max Little receives research support from the Wellcome Trust and The Michael J. Fox Foundation for Parkinson's Research.

REFERENCES

1. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992; 55:181–4.
2. Brunato M, Battiti R, Pruitt D, Sartori E. Supervised and unsupervised machine learning for the detection, monitoring and management of Parkinson's disease from passive mobile phone data. In: predicting Parkinson's disease progression with smartphone data, Kaggle Competition. Available at: <https://kaggle2.blob.core.windows.net/prospector-files/1117/958625cf-3514-4e64-b0e7-13ebd3cf9791/kaggle.pdf>. Accessed October 22, 2014.
3. Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc Interface* 2011; 8:842–55.
4. Sejdić E, Lowry KA, Bellanca J, Redfern MS, Brach JS. A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. *IEEE Trans Neural Syst and Rehabil Eng* 2014; 22:603–12.
5. Taylor Tavares AL, Jefferis GS, Koop M, Hill BC, Hastie T, Heit G, et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. *Mov Disord* 2005; 20:1286–98.
6. Gauntlett-Gilbert J, Brown VJ. Reaction time deficits and Parkinson's disease. *Neurosci Biobehav Rev* 1998; 22:865–81.

7. Maetzler W, Domingos J, Srulijes K, Ferreira JJ, Bloem BR. Quantitative wearable sensors for objective assessment of Parkinson's disease. *Mov Disord* 2013; 28:1628–37.
8. Abdolahi A, Scoglio N, Killoran A, Dorsey ER, Biglan KM. Potential reliability and validity of a modified version of the unified Parkinson's disease rating scale that could be administered remotely. *Parkinsonism Relat Disord* 2013; 19:218–21.
9. Harel B, Cannizzaro M, Snyder PJ. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. *Brain and Cogn* 2004; 56:24–9.
10. Breiman L. Random Forests. *Machine Learning* 2001; 45:5–32.
11. Post B, Merkus MP, de Bie RMA, de Haan RJ, Speelman JD. Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?. *Mov Disord* 2005; 20:1577–84.
12. Pramis J. By 2017 there will almost be as many smartphones as literate adults. In: *Digital Trends*. Available at: <http://www.digitaltrends.com/mobile/2017-literate-adults-smartphones/?tru=bHNoSD#ixzz2d1dtkRDg>. Accessed October 22, 2014.

Table 1. Baseline characteristics of study participants.

Characteristic	Parkinson's disease participants (N=10)	Control participants (N=10)
Age (SD)	65.1 years (9.8)	57.7 years (14.3)
Percent women	30%	60%
Percent taking levodopa	90%	0%
Percent with high school education	100%	100%
Baseline motor Unified Parkinson's Disease Rating Scale score (SD)	19.6 (6.7)	NA
Baseline Parkinson's Disease Questionnaire 39 score (SD)	18.5 (16.9)	NA

NA = not applicable; SD = standard deviation

Supplementary Table 1: Brief description of summary measures extracted from the voice recordings.

<i>Summary Measure</i>	<i>Brief Description</i>
Detrended Fluctuation Analysis	Characterizes the changing detail of aero-acoustic breath noise
Recurrence Period Density Entropy (RPDE)	Quantifies any ambiguity in fundamental pitch (RPDE is zero for perfectly periodic signals and one for purely stochastic signals)
Teager-Kaiser Energy Operator	Measures the instantaneous changes in voice energy (takes into account both amplitude and frequency)
Jitter	Quantifies the instabilities of the oscillating pattern of the vocal folds by measuring cycle-to-cycle changes in the fundamental frequency (measure of roughness in voice)
Shimmer	Quantifies the instabilities of the oscillating pattern of the vocal folds by measuring cycle-to-cycle changes in the amplitude (measure of roughness in voice)
Pitch Period Entropy (PPE)	Measures the impaired control of stable pitch, a property common in PD
Harmonics-to-Noise Ratio (HNR)	Quantifies noise in the speech signal, caused mainly due to incomplete vocal fold closure
Noise-to-Harmonics Ratio (NHR)	Quantifies noise in the speech signal, caused mainly due to incomplete vocal fold closure
Mel Frequency Cepstral Coefficients (MFCCs)	Computes the contribution of the energy of the speech signal at each frequency band (are aimed at detecting subtle changes in the motion of the articulators)
Glottis Quotient (GQ)	Quantifies properties of the vocal folds (when glottis is open and closed)
F0 contour features	Measures based the summary statistics of the fundamental frequency
Glottal to Noise Excitation (GNE) ratio	Quantifies the extent of noise in speech using linear and nonlinear energy measures
Vocal Fold Excitation Ratios (VFER)	Quantifies the extent of noise in speech using energy (linear and nonlinear) and entropy based measures
Perturbation Quotient (PQ)	Quantifies variations in speech signal
Wavelet related measures	Variants of above discussed summary measures applied to wavelet coefficients of the speech signal

Supplementary Table 2: Brief description of summary measures extracted from the tri-axial accelerometer time series along the x-axis (a_x), y-axis (a_y), and z-axis (a_z), recorded during the postural sway and gait tests. We also extracted summary measures along the resultant absolute acceleration value (aav), computed as $aav = \sqrt{a_x^2 + a_y^2 + a_z^2}$.

<i>Summary measure</i>	<i>Brief Description</i>
μ	Average
σ	Standard deviation
Q_1	25th percentile
Q_3	75th percentile
IQR	Inter-quartile range ($Q_3 - Q_1$)
M	Median
M_0	Mode
R	Data range (maximum – minimum)
s	Skewness
k	Kurtosis
x_{mse}	Mean squared energy
S	Entropy
$H(x)$	Entropy rate
$r_{x,y}$	Cross-correlation between the acceleration in x- and y-axis
$I(x; y)$	Mutual information between the acceleration in x- and y-axis
$S(x, y)$	Cross-entropy between the acceleration in x- and y-axis
1DFA	Extent of randomness in body motion
mean 2TKEO	Instantaneous changes in energy due to body motion
ϕ_1	Autoregression coefficient at time lag 1
zcr	Zero-crossing rate
F_d	Dominant frequency component using Lomb-Scargle periodogram
r	Radial distance
θ	Polar angle
φ	Azimuth angle

¹DFA stands for Detrended Fluctuation Analysis.

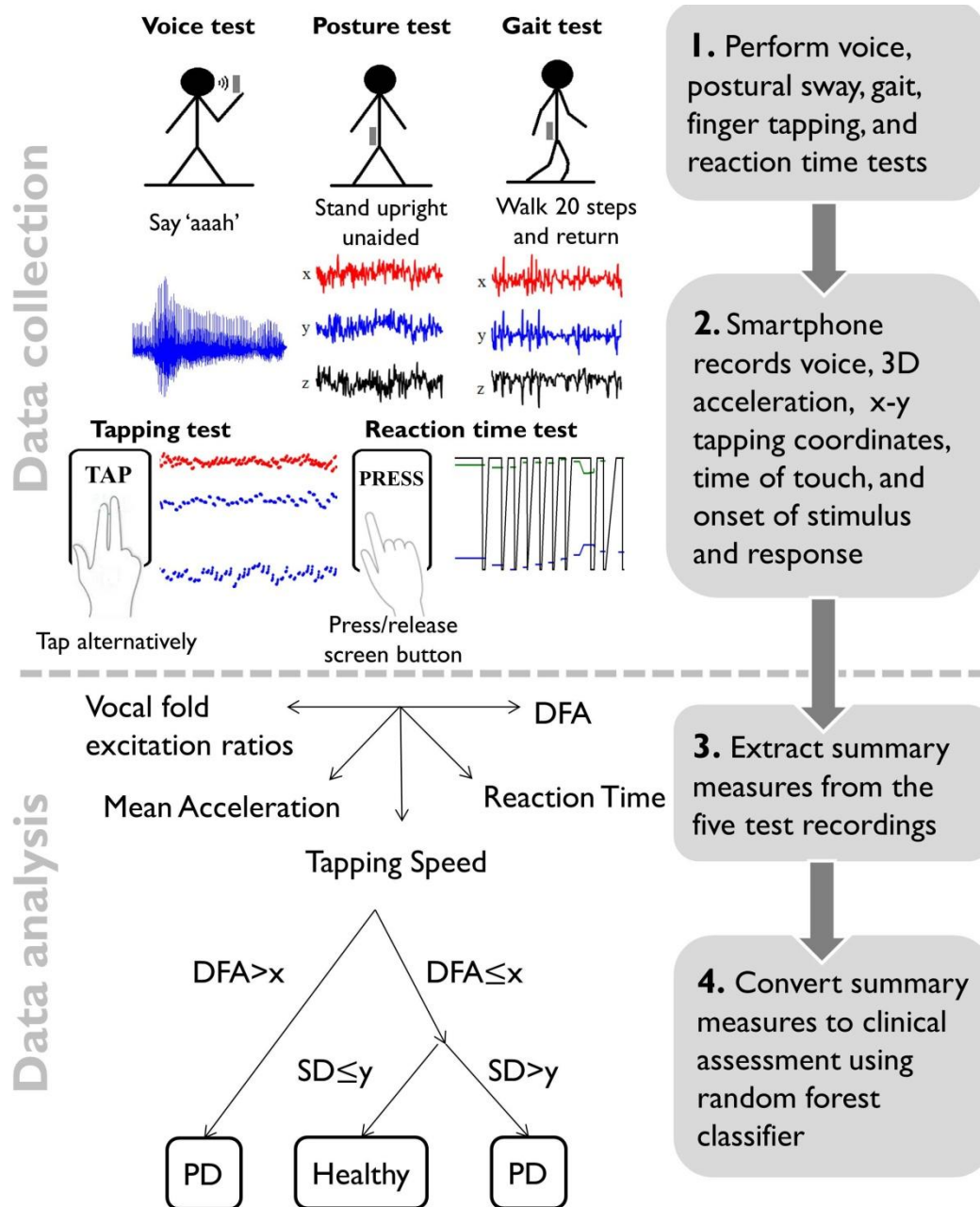
²TKEO stands for Teager-Kaiser Energy Operator.

Supplementary Table 3: Brief description of features extracted from the finger tapping data.

<i>Summary measure</i>	<i>Brief description</i>
TS	Tapping speed (number of finger taps per second)
μ_{ITI}	Average inter-tap interval (average elapsed time between two consecutive taps)
σ_{ITI}	Standard deviation of inter-tap intervals
CV_{ITI}	Coefficient of variation of inter-tap intervals
MSE_{ITI}	Mean squared energy of inter-tap intervals
$TKEO_{ITI}$	Mean Teager-Kaiser energy operator of inter-tap intervals (quantifies energy variation)
$AR1_{ITI}$	Autoregressive coefficient of inter-tap intervals (characterizes relationship between inter-tap intervals at lag 1)
$AR2_{ITI}$	Autoregressive coefficient of inter-tap intervals (characterizes relationship between inter-tap intervals at lag 2)
DFA_{ITI}	Detrended fluctuation analysis of inter-tap intervals (characterizes changes in inter-tap interval)
$Fatigue_{ITI(10\%)}$	Increase in the average inter-tap interval from the first 10% to the last 10% of finger taps during a test (measure of slowing down in tapping speed)
$Fatigue_{ITI(25\%)}$	Increase in average inter-tap interval from the first 25% to the last 25% of finger taps during a test
$Fatigue_{ITI(50\%)}$	Increase in average inter-tap interval from the first 50% to the last 50% of finger taps during a test
$Tremor_{intra-tap}$	Features (mean, standard deviation, median, min, max and range) computed using the intra-tap tremor (displacement in position of the finger during a tap)
$ND_{intra-tap}$	Net displacement in finger position during a tap
$Tremor_{inter-tap}$	Features based on inter-tap tremor (drift in the finger position between consecutive taps)
\emptyset	Finger opening angle estimate

Note: 1) Inter-tap interval was computed as the elapsed time between consecutive taps. 2) Displacement in finger position coordinates was quantified using the Euclidean distance.

Figure 1. Schematic diagram illustrating the procedure for collecting voice recordings (sustained vowel ‘aaah’), acceleration time traces during postural sway and gait tests, finger tapping coordinates and the time of touch, and reaction time (onset of stimulus and response), using commercially-available smartphones, along with the major steps involved in the data analysis.



Abbreviations: 3D = three dimensional; DFA = detrended fluctuation analysis; PD = Parkinson's disease; SD = standard deviation

Supplemental Figure 1. Image of Android smartphone and software application used in the study.

