RESEARCH ARTICLE

A Validated Smartphone-Based Assessment of Gait and Gait Variability in Parkinson’s Disease

Robert J. Ellis¹, Yee Sien Ng³, Shenggao Zhu², Dawn M. Tan³, Boyd Anderson¹, Gottfried Schlaug⁴, Ye Wang¹,²*

¹ School of Computing, National University of Singapore, Computing 1, 13 Computing Drive, Singapore, 117417, Singapore, 2 NUS Graduate School for Integrative Sciences and Engineering, 28 Medical Drive, Singapore, 117456, Singapore, 3 Department of Rehabilitation Medicine, Singapore General Hospital, Outram Rd, Singapore, 169608, Singapore, 4 Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Palmer 127, Boston, MA, 02215, United States of America

* wangye@comp.nus.edu.sg

Abstract

Background

A well-established connection exists between increased gait variability and greater fall likelihood in Parkinson’s disease (PD); however, a portable, validated means of quantifying gait variability (and testing the efficacy of any intervention) remains lacking. Furthermore, although rhythmic auditory cueing continues to receive attention as a promising gait therapy for PD, its widespread delivery remains bottlenecked. The present paper describes a smartphone-based mobile application (“SmartMOVE”) to address both needs.

Methods

The accuracy of smartphone-based gait analysis (utilizing the smartphone’s built-in tri-axial accelerometer and gyroscope to calculate successive step times and step lengths) was validated against two heel contact–based measurement devices: heel-mounted footswitch sensors (to capture step times) and an instrumented pressure sensor mat (to capture step lengths). 12 PD patients and 12 age-matched healthy controls walked along a 26-m path during self-paced and metronome-cued conditions, with all three devices recording simultaneously.

Results

Four outcome measures of gait and gait variability were calculated. Mixed-factorial analysis of variance revealed several instances in which between-group differences (e.g., increased gait variability in PD patients relative to healthy controls) yielded medium-to-large effect sizes (eta-squared values), and cueing-mediated changes (e.g., decreased gait variability when PD patients walked with auditory cues) yielded small-to-medium effect sizes—while at the same time, device-related measurement error yielded small-to-negligible effect sizes.
Conclusion
These findings highlight specific opportunities for smartphone-based gait analysis to serve as an alternative to conventional gait analysis methods (e.g., footswitch systems or sensor-embedded walkways), particularly when those methods are cost-prohibitive, cumbersome, or inconvenient.

Introduction
The expected number of individuals living with Parkinson’s disease (PD) will rise sharply by the year 2030, doubling the number of patients living with the disease in the year 2005 to more than 9 million [1]. With much of this increase to be found in rapidly growing countries with still-developing economies such as Brazil, China, and India [2], existing methods for managing the various challenges of PD faced by both individual patients (mental, physical, social, financial) and the medical community (diagnostic methods, therapy delivery) may prove difficult to scale up.

One of the most serious challenges in dealing with the progression of PD is an increase in gait disturbances. “Episodic” disturbances include periods of freezing, festination, or initiation hesitation [3,4]. “Continuous” disturbances affect the step-to-step spatiotemporal dynamics of gait, resulting in increased spatiotemporal gait variability (GV) (for extensive discussions, see [5–7]). The most prevalent outcome measures of GV (for reviews, see [8,9]) are second-moment statistics (i.e., standard deviation or coefficient of variation) of a series of step or stride durations or lengths. Second-moment statistics require precise information about individual gait events (rather than averaged gait events). As such, they are both statistically and conceptually dissociable from first-moment (i.e., mean-based) statistics [6], as supported by large-N factor analytic studies [10–12].

Several classic findings regarding PD and GV have been reported (for detailed reviews, see [5,6]). PD patients show increased GV relative to age-matched healthy elderly (HE) individuals [13–16], particularly when in a dopamine-deplete (off-medication) state [17,18] or when they perform a concurrent cognitive or motor task [19–21]. Conversely, GV can be reduced in PD through the use of external sensory stimulation; in particular, rhythmic auditory cueing (RAC) paradigms (for reviews, see [22–26]). The motor system—from locomotion to manual coordination to speech articulation—is highly adept at synchronizing or entraining to auditory rhythms (e.g., a ticking metronome, or music with a steady beat); an affordance of the intimate auditory–motor pathways in the human brain (for reviews, see [27–29]). When PD patients attempt to synchronize their heel strikes with the auditory beat, however, they show improvements in both first-moment [22,23,26] and second-moment [15,16,21,30] outcome measures of gait.

The significance of GV, however, extends beyond differences between PD and HE or reductions during an RAC paradigm. Importantly, individuals with higher-than-normal GV—both in PD and more broadly—are at increased risk of falling [6,16]; this association has been found using both retrospective [31] and prospective [12,32,33] designs. The consequences of a fall (including a high rate of serious injury [34]) extend beyond the event itself, feeding into a cycle involving fear of falling, immobilization, social isolation, depression, cognitive decline, and increased mortality [3].

In theory, if higher-than-normal GV were detected (e.g., if a GV assessment were incorporated into a regular physical examination), preventive steps (from mental strategies to gait training [35,36]) could be taken to help mitigate fall risk [37,38]. An ideal assessment system...
would contain three core components: the sensing hardware which records the subject’s movement, the analysis software which translates the recorded signal into an outcome measure, and a display unit which communicates the value of that outcome measure. Numerous component-based systems, assembled from third-party sensor, processor, and/or display units, are available. Second-moment outcome measures of GV are most frequently obtained from detected heel contacts (as reviewed in [8,9]), either using either using pressure sensors (footswitches) affixed to the heel [13, 15–17, 20, 30–32, 39, 40] or pressure sensors embedded in a rollable walkway [10–12,14,18,21,33,41,42]. Other measurement approaches (reviewed in [43–45]) center around the use of an inertial measurement unit (comprising a tri-axial accelerometer and/or tri-axial gyroscope) affixed to the torso [46–49], feet [50–54], or multiple locations [55–57], and which compute outcome measures from a series of heel strike “analogues” (i.e., accelerometer waveform events associated with actual heel strikes [58,59]).

With the advent of ubiquitous and powerful smartphones (which contain an inertial measurement unit, a processing core, and a touchscreen), proposed self-contained systems for gait analysis have become more frequent [60–67] (see Table 1), including our own recent investigation [68]. Smartphone-based assessments in PD—including, but not limited to gait—offer numerous potential benefits: in terms of cost savings, portability, customizability, patient tolerance, and deployment scalability [2,69,70]. In reviewing this literature, however, three limitations become apparent.

A first limitation concerns the target sample. Most previous investigations focused either on healthy young [60–62,64,67] or healthy elderly [63,65] individuals. Step detection algorithms trained using healthy subject data may not yield accurate results when tested with PD data, due to important differences in the spatiotemporal dynamics of gait between these two populations.

Table 1. Summary of key features of prior studies of smartphone-mediated gait analysis.

<table>
<thead>
<tr>
<th>First Author (Year)</th>
<th>Group (N)</th>
<th>Age: M (SD)</th>
<th>Recording parameters</th>
<th>Outcome measures derived from</th>
<th>Concurrent validity obtained for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chan (2011) [60]</td>
<td>HY (1a)</td>
<td>n/a</td>
<td>Left pocket 100</td>
<td>Stride ΔM</td>
<td>—</td>
</tr>
<tr>
<td>How (2013) [61]</td>
<td>HY (1a)</td>
<td>n/a</td>
<td>Front waist 60</td>
<td>—</td>
<td>Feetwitches —</td>
</tr>
<tr>
<td>LeMoyne (2011) [62]</td>
<td>HY (1a)</td>
<td>n/a</td>
<td>Left ankle 100</td>
<td>Stride ΔM</td>
<td>—</td>
</tr>
<tr>
<td>Mellone (2012) [63]</td>
<td>HE (49)</td>
<td>59 (16)</td>
<td>Lower back 50</td>
<td>Step ΔM, ΔSD</td>
<td>Accel. —</td>
</tr>
<tr>
<td>Nishiguchi (2012) [64]</td>
<td>HY (30)</td>
<td>20.9 (2.1)</td>
<td>Lower back 33</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Palmerini (2011) [71]</td>
<td>HE (49)</td>
<td>58.9 (16.5)</td>
<td>Lower back 50</td>
<td>Step ΔM, ΔCV</td>
<td>Accel. —</td>
</tr>
<tr>
<td>Yamada (2011) [66]</td>
<td>RA (39) / HE (20)</td>
<td>65.9 (10) / 69.1 (5.8)</td>
<td>Lower back 33</td>
<td>Step ΔM, ΔCV</td>
<td>—</td>
</tr>
<tr>
<td>Yang (2012) [67]</td>
<td>HY (13)</td>
<td>23–36</td>
<td>Lower back 100</td>
<td>Cadence, Step ΔM b</td>
<td>Accel. —</td>
</tr>
<tr>
<td>Zhu (2014) [68]</td>
<td>PD (10)</td>
<td>66.3 (7.8)</td>
<td>Front waist 100</td>
<td>Step ΔM, ΔCV, ΔM, ΔCV</td>
<td>Feetwitches GAITRite</td>
</tr>
<tr>
<td>[Present study]</td>
<td>PD (12) / HE (12)</td>
<td>65.0 (8.4) / 63.1 (7.8)</td>
<td>Front waist 100</td>
<td>Step ΔM, ΔCV, ΔM, ΔCV</td>
<td>Feetwitches GAITRite</td>
</tr>
</tbody>
</table>

**Abbreviations:** Accel.: conventional accelerometer; ΔM: mean inter-event interval; ΔSD: standard deviation of inter-event intervals; ΔCV: coefficient of variation of inter-event intervals; HE: healthy elderly; HY: healthy young; RA: rheumatoid arthritis; PD: Parkinson’s disease; SF: sampling frequency.

a piloted data used to illustrate algorithms or processing steps in a proof-of-concept format; b obtained by dividing the pre-specified walking distance by the number of detected steps.

doi:10.1371/journal.pone.0141694.t001
Conversely, the algorithms developed in our previous paper [68] were exclusively trained on PD data, and have not yet been evaluated on healthy elderly data.

A second limitation concerns the use of concurrent validation; that is, an analysis of the accuracy of outcome measures derived from the smartphone-based system relative to outcome measures derived from a conventional gait measurement system. Specifically, a key question for any novel gait analysis system is whether it accurately detects heel strike analogues in the accelerometer waveform relative to ground truth (i.e., actual heel strike events). In several previous investigations of concurrent validity, however, “ground truth” was a second (non-smartphone) accelerometer [63,64,66,67] rather than actual heel contacts (as in [62,68]), preventing this important question from being answered.

A third limitation concerns the examined outcome measures. Several previous investigations of smartphone-based gait analysis have examined second-moment statistics of GV [63,65,66,68], as reviewed above. With the exception of our previous report [68], however, these studies have exclusively focused on step or stride time measures of GV rather than step or stride length measures of GV. Further evaluations of this latter class of outcome measure are warranted, as they are frequently reported in the literature [10,15,18,31,42].

Thus, missing from the literature is an analysis of the accuracy of smartphone accelerometry–derived outcome measures relative to heel contact–derived outcome measures of gait and gait variability in both PD and HE.

In the present study, relative accuracy was assessed using mixed-factorial analysis of variance (ANOVA). ANOVA enables the primary “novel” source of variance (i.e., device-related measurement error) to be entered into a statistical model alongside “expected” sources of variance (i.e., differences between PD and HE and differences between self-paced versus metronome-cued conditions). By computing and comparing eta-squared ($\eta^2$) effect sizes for each of these ANOVA model components, the relative magnitude of device-related measurement error in the context of widely-used experimental contrasts can be assessed more objectively.

**Methods**

The research study described here was formally approved by the Institutional Review Board (IRB) of Bright Vision Hospital / Singapore General Hospital (approval number 2013/150/F), and conducted according to the principles expressed in the Declaration of Helsinki. Upon arrival at the testing location, all subjects were informed about the purpose of the study and provided written consent.

**1. Participants**

All subjects were recruited through the Singapore General Hospital clinics. A telephone questionnaire was first administered to screen out potential subjects who (1) are not within the age range of 40 to 85; (2) have any problems with their hearing; (3) are not able to walk independently without an aid; (4) have joint problems or other neurological, musculoskeletal or medical problems that can affect walking; (5) have sustained a fall within the past year that continues to affect their walking pattern; (6) have had surgery to implant a device (e.g., deep brain stimulation or pacemaker). Subjects who satisfied all six criteria were invited to participate in the study.

Upon arrival at the testing location, four clinical assessments were administered to PD subjects, beginning 30 to 90 minutes after standard medication intake: the complete Unified Parkinson’s Disease Rating Scale (UPDRS [72]), the modified Hoehn & Yahr stage assessment [73], the Mini Mental State Exam [74], and the Freezing of Gait Questionnaire [75]. Of these
assessments, only the MMSE was administered to HE subjects. All subjects had a MMSE ≥ 24, indicating an absence of cognitive impairment.

Our initial target sample size for this study was 15 PD patients and 15 HE subjects. However, several issues during data collection resulted in incomplete data sets from one or more RAC conditions for 3 PD patients and 3 HE subjects, due to: (1) an insufficiently tight connection between SmartMOVE and the chest (i.e., a tighter elastic band was required than was available at the time); (2) data lost from SmartMOVE during file transfer; (3) poor quality contact between the footswitch sensor and the heel, yielding irregular heel strike timing data; or (4) insufficient data (particularly in GAITRite, which yields the fewest events per trial) due to patient fatigue. Because a complete set of data across all experimental conditions is required for ANOVA (i.e., no missing values are permitted), two options were available: (1) imputation of missing values or (2) casewise deletion of any subject with a missing condition. Because any form of data imputation would introduce its own new set of assumptions (e.g., as reviewed in [76]), we chose the second option for the sake of parsimony and clarity.

Thus, a final sample of 12 PD patients (5 female) and 12 HE subjects (4 female) were analyzed in this study.

Table 2. Demographics of the PD and HE samples, and clinical characteristics PD sample.

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Gender</th>
<th>MMSSE (0 to 30)</th>
<th>Age (yrs)</th>
<th>Gender</th>
<th>MMSSE (0 to 30)</th>
<th>Disease duration (yrs)</th>
<th>UPDRS III (0 to 56)</th>
<th>H &amp; Y (1 to 5)</th>
<th>FOGQ (0 to 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>63.4 M</td>
<td>30</td>
<td>58.8 M</td>
<td>30</td>
<td>7</td>
<td>20</td>
<td>4</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51.0 M</td>
<td>30</td>
<td>72.1 M</td>
<td>30</td>
<td>3</td>
<td>33</td>
<td>2.5</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68.5 M</td>
<td>29</td>
<td>77.5 M</td>
<td>28</td>
<td>8</td>
<td>16</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66.1 F</td>
<td>29</td>
<td>63.4 F</td>
<td>24</td>
<td>14</td>
<td>24</td>
<td>2.5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67.6 M</td>
<td>30</td>
<td>65.5 F</td>
<td>29</td>
<td>6</td>
<td>28</td>
<td>2.5</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64.9 M</td>
<td>30</td>
<td>81.0 M</td>
<td>30</td>
<td>2</td>
<td>37</td>
<td>2.5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59.1 M</td>
<td>29</td>
<td>61.8 F</td>
<td>30</td>
<td>2</td>
<td>36</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60.5 F</td>
<td>26</td>
<td>63.5 M</td>
<td>28</td>
<td>4</td>
<td>14</td>
<td>1.5</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65.8 F</td>
<td>29</td>
<td>63.2 M</td>
<td>29</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79.9 M</td>
<td>29</td>
<td>49.8 M</td>
<td>30</td>
<td>5</td>
<td>38</td>
<td>2.5</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52.0 M</td>
<td>30</td>
<td>60.6 F</td>
<td>30</td>
<td>4</td>
<td>21</td>
<td>2.5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58.2 F</td>
<td>30</td>
<td>62.3 F</td>
<td>28</td>
<td>3</td>
<td>22</td>
<td>3</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Mean]</td>
<td></td>
<td>63.08</td>
<td></td>
<td>29.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[SD]</td>
<td></td>
<td>7.79</td>
<td></td>
<td>1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“UPDRS III” is the Unified Parkinson’s Disease Rating Scale [72] section III (lower score indicates greater motor impairment); “H & Y” is the modified Hoehn & Yahr stage assessment [73] (higher score indicates more advanced PD stage); “MMSE” is the Mini Mental State Exam [74] (lower score indicates greater cognitive impairment); and “FOGQ” is the Freezing of Gait Questionnaire [75] (higher score indicates increased freezing severity).

doi:10.1371/journal.pone.0141694.t002

assessments, only the MMSE was administered to HE subjects. All subjects had a MMSE ≥ 24, indicating an absence of cognitive impairment.

Our initial target sample size for this study was 15 PD patients and 15 HE subjects. However, several issues during data collection resulted in incomplete data sets from one or more RAC conditions for 3 PD patients and 3 HE subjects, due to: (1) an insufficiently tight connection between SmartMOVE and the chest (i.e., a tighter elastic band was required than was available at the time); (2) data lost from SmartMOVE during file transfer; (3) poor quality contact between the footswitch sensor and the heel, yielding irregular heel strike timing data; or (4) insufficient data (particularly in GAITRite, which yields the fewest events per trial) due to patient fatigue. Because a complete set of data across all experimental conditions is required for ANOVA (i.e., no missing values are permitted), two options were available: (1) imputation of missing values or (2) casewise deletion of any subject with a missing condition. Because any form of data imputation would introduce its own new set of assumptions (e.g., as reviewed in [76]), we chose the second option for the sake of parsimony and clarity.

Thus, a final sample of 12 PD patients (5 female) and 12 HE subjects (4 female) were analyzed in this study. Table 2 presents key demographics and scores on standard clinical ratings scales for each subject. A two-sample t-test revealed that the PD and HE samples did not differ in age (p = .577) or MMSE score (p = .496).

2. Walking evaluation

Fig 1 illustrates key features of the walking evaluation. The SmartMOVE app runs on an Apple iPod Touch (running iOS 6.1) and was secured at the subject’s navel using an elastic strap (Fig 1A), and positioned with the screen facing towards the subject and the audio jack facing up (so that the device’s loudspeaker was unobstructed). SmartMOVE records 6 channels of IMU data, all at 100 Hz, in the iOS device xyz coordinate system: tri-axial acceleration (ax, ay, and az) and
tri-axial gyroscopic rotation rate ($\omega_x$, $\omega_y$, and $\omega_z$). For simplicity, directions are defined with respect to the subject: anterior–posterior (AP) acceleration as $a_{AP} = -a_z$ (positive values for anterior acceleration), up–down (UD) acceleration as $a_{UD} = -a_y$ (positive values for upward acceleration), and left–right (LR) acceleration as $a_{LR} = a_x$ (positive values for leftward acceleration). Similarly, we define the rotation rate around the AP axis (i.e., roll) as $\omega_{AP} = -\omega_z$, around the UD axis (i.e., yaw) as $\omega_{UD} = -\omega_y$, and around the LR axis (i.e., pitch) as $\omega_{LR} = \omega_x$. (These labels were incorrectly stated in Section 5.1 of Zhu et al. [68]; an erratum [77] and corrected manuscript [78] are available.)

A footswitch sensor was affixed to each heel pad before fitting subjects with a layer of rubber-toed socks (to protect the sensor and provide traction). A single trial consisted of a 26-meter path, including 7-m on the GAITRite, and a single turn at the half-way point (Fig 1C). Trials were repeated until a minimum of 40 steps were recorded on the GAITRite itself, in three sequential conditions: (1) "Self-paced" (no external cue), (2) "100% RAC" (metronome tempo set at the average self-paced cadence, as determined by GAITRite), and (3) "110% RAC" (metronome tempo set 10% faster than 100% RAC). 100% and 110% relative tempos are common in the RAC literature [15,16,21,22,40,79–81], and are easily configured using SmartMOVE’s menu settings (Fig 1B). Any trial in which a subject experienced gait freezing was discarded and performed again.
3. Gait analysis

A nine-part preprocessing pipeline was used to translate raw SmartMOVE accelerometry data into a series of inter-step times and inter-step lengths, and is detailed in S1 Text. For a given experimental condition, “Δ-series” of inter-step events (times and lengths) was derived by concatenating inter-step events across individual trials in that condition as in Lord et al. [17,82]). Two outcome measures were obtained for each Δ-series: the mean (ΔM) and the coefficient of variation (ΔCV, defined as 100 × ΔSD / ΔM, where ΔSD is the standard deviation of Δ-values). As noted in the Introduction, these statistics are widely used in the PD literature, including investigations of RAC [15,16,21–23,26,30], and have high test–retest reliability [83].

4. Quantifying device-related measurement error: ANOVA and η²

In seeking to validate a novel measurement device against a standard measurement device, an analysis of the magnitude of measurement error is critical. Just as critical, however, is understanding the context in which that inter-device error emerges; that is, the magnitude of measurement error relative to the magnitude of the target experimental effects; for example, inter-group (PD vs. HE) and inter-task (self-paced vs. metronome-cued) effects. ANOVA provides a convenient way of capturing all these sources of variance simultaneously (see S2 Text for further justification).

For each outcome measure (step time and step length ΔM; step time and step length ΔCV), three ANOVAs were performed using Statistica. Each of these analyses represents a potentially “self-sufficient” experimental question, and thus a distinct ANOVA design:

1. A Group (2 levels: PD and HE) × RAC (3 levels: self-paced, 100%, and 110%) × Device (2 levels: either SmartMOVE Biometrics footswitch sensors for step time outcome measures, or SmartMOVE and GAITRite sensor walkway for step length outcome measures).

2. A Group × Device ANOVA during self-paced walking to determine whether differences between PD and HE were significantly different when measured by SmartMOVE versus the novel versus the standard device.

3. An RAC × Device ANOVA for PD patient group in isolation.

4. An RAC × Device ANOVA for HE subject group in isolation.

For each ANOVA component (main effect or interaction), eta-squared (η²) effect sizes were computed [84, 85]. η² quantifies the proportion of total variance (from 0 to 1) that is captured by a particular ANOVA component. By convention, an η² ≈ .02 is considered a “small” effect, an η² ≈ .13 a “medium” effect, and an η² ≈ .26 a “large” effect [86, 87]. Thus, the primary purpose of these analyses is to identify those instances (if any) in which a significant main effect for Device or interaction with Device emerged, and to quantify the magnitude of Device-related effects (i.e., η² values) relative to Group- or RAC-related effects. Instances in which “expected” sources of variance (i.e., main effects for Group and/or RAC) had substantially larger effect sizes than main effects or interactions with Device would indicate situations in which SmartMOVE could serve as a viable alternative to conventional gait analysis systems.

Results

Two important “precursor” results which extend our previous report [68] should first be noted. First, accelerometer waveform peaks were confirmed to be the most temporally stable waveform analogue of actual heel strikes both PD and HE datasets, as detailed in Step 4 of S1 Text. Second, the machine learning algorithms for waveform peak identification, step length...
calculation, and left-versus-right foot identification were confirmed to perform with very high accuracy in the full dataset comprising PD and HE subjects, as detailed Step 5 of S1 Text.

Next, the key results of the present study—the accuracy of SmartMOVE-derived outcome measures relative to heel contact–derived outcome measures—are presented. Fig 2 plots group-level means and standard errors for each outcome measure, as a function of group (PD in red; HE in blue), RAC condition levels (x-axis), and device (separate panels). \( \eta^2 \) values for three simple effects (PD versus HE during self-paced walking; self-paced versus 110% RAC in PD; self-paced versus 110% RAC in HE) are highlighted for each outcome measure on Fig 2. S1 Table and S2 Table provide the group means and error bars for step time and step length data, and S1 Dataset and S2 Dataset contain the individual subject outcome measure data used to perform the step time and step length ANOVAs. The four ANOVAs described in Section 4 of the Methods are presented in succession.

1. **Group \( \times \) RAC \( \times \) Device ANOVA**

The three-factor ANOVA enables a comparison of inter-group, inter-condition, and inter-device variance for each outcome measure. Table 3 and Table 4 present the step time and step length ANOVA statistics, respectively. As expected, significant group differences emerged in all four outcome measures: relative to HE subjects, PD patients walked with slower steps (step time \( \Delta M \)), shorter steps (step length \( \Delta M \)), and increased step time and step length variability (\( \Delta CV \)). Additionally, relative to self-paced walking, metronome-cued walking set to 110% of self-paced cadence yielded significantly faster walking with longer steps and less step-to-step variability. Effect sizes (\( \eta^2 \) values) for **Group** were typically much larger than for **RAC**, indicating a greater separation of outcome measure values. Both \( \Delta CV \) measures showed a significant main effect for **Device**, with SmartMOVE showing inflated outcome measure values relative to heel contact–based devices. The corresponding effect sizes, however, were small (\( \eta^2 < .02 \)), as were effect sizes for all other interactions with **Device**.

2. **Group \( \times \) Device ANOVA during self-paced walking**

This ANOVA quantified the impact of RAC in PD patients and whether that effect was significantly influenced by measurement device. ANOVA statistics are presented in Table 5 and Table 6. \( \eta^2 \) values for **Group** were larger for \( \Delta M \) statistics (> .41) than \( \Delta CV \) statistics (< .22). Statistically significant main effects for **Device** were present in step time \( \Delta CV \), step length \( \Delta M \), and step length \( \Delta CV \). However, the corresponding \( \eta^2 \) values for these significant effects were all small (all \( \eta^2 < .02 \)), as were \( \eta^2 \) values for all **Group \( \times \) Device** interactions.

3. **RAC \( \times \) Device ANOVA for PD patients**

This ANOVA quantified the impact of RAC on gait in PD, and whether RAC effects were significantly influenced by measurement device. ANOVA statistics are presented in Table 7 and Table 8. Significant main effects for RAC were present in step time \( \Delta M \), step time \( \Delta CV \), and step length \( \Delta M \). Only step time \( \Delta M \) showed a large (\( \eta^2 = .3067 \)) effect size. All main effects and interactions with **Device** had negligible effect sizes (all \( \eta^2 < .01 \)).

4. **RAC \( \times \) Device for HE subjects**

For sake of completeness, a **RAC \( \times \) Device** ANOVA was performed for HE subject data; ANOVA statistics are presented in Table 9 and Table 10. Main effects for RAC were significant
Fig 2. Outcome measure results. Results of the Group (separate lines: PD vs. HE) × Condition (x-axis: self-paced, 100% RAC, 110% RAC) × Device (separate panels: SmartMOVE vs. heel contact–based) ANOVAs for step time (a.) and step length (b.) outcome measures.

doi:10.1371/journal.pone.0141694.g002
for both step time $\Delta_M$ and step length $\Delta_M$, with a substantially larger effect size for step time $\Delta_M (\eta^2 = .3586)$. Unlike the previous ANOVAs, Device had a strikingly large effect size for step time $\Delta_M (\eta^2 = .5099)$, and a small-to-medium effect size for step length $\Delta_M (\eta^2 = .0848)$. In both cases, Device effect sizes were larger than RAC effect sizes: outcome measures in HE subjects were more affected by device-induced measurement error than they were by the target RAC manipulation.

### Discussion

The current paper presents the first systematic validation of the accuracy of smartphone-based gait analysis in both Parkinson’s disease (PD) patients and age-matched healthy elderly (HE) subjects. Concurrent validity was obtained for the novel smartphone app (“SmartMOVE”) by simultaneously recording walking patterns using two heel contact–based measurement devices (footswitches to quantify step durations, and a GAITRite sensor walkway to quantify step displacements). Analysis of variance (ANOVA) and eta-squared ($\eta^2$) effect sizes were used to quantify the magnitude of device-related measurement error associated with quantifying differences between PD and HE during self-paced walking, differences between self-paced and metronome-cued walking in PD, and differences between self-paced and metronome-cued walking in HE.

The present study captured two experimental effects frequently noted in the PD literature (see Fig 2). First, during self-paced walking, relative to HE subjects, PD patients walked with

### Table 3. Statistics associated with the Group $\times$ RAC $\times$ Device ANOVA for step time outcome measures.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\Delta_M$</th>
<th>$\Delta_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group $\times$ Device</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAC $\times$ Device</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“SS” is the portioned sums of squares for each ANOVA term. Significant $p$-values are highlighted in bold.

doi:10.1371/journal.pone.0141694.t003

### Table 4. Statistics associated with the Group $\times$ RAC $\times$ Device ANOVA for step length outcome measures.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\Delta_M$</th>
<th>$\Delta_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group $\times$ Device</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAC $\times$ Device</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“SS” is the portioned sums of squares for each ANOVA term. Significant $p$-values are highlighted in bold.

doi:10.1371/journal.pone.0141694.t004

The ability to replicate inter-group differences between PD and HE and inter-condition differences between self-paced and externally cued walking serves as a useful manipulation check, allowing the key unknown source of variance—the magnitude of device-related measurement error—to be examined within the context of “typical” experimental results.

1. Device-related measurement error in context: Summary

Three summary statements may be offered with respect to patterns η² values among the three target factors: Group, RAC, and Device (cf. Table 3 to Table 10).

The first pattern is defined as a is medium-to-large effect size (η² ≈ .20 or greater) for a target experimental effect (Group or RAC), and a small effect size (η² < .02) for Device. This pattern was observed several times: (1) in all four outcome measures when assessing differences between PD and HE during self-paced walking (Table 5 and Table 6); (2) in step time ΔM when assessing the influence of RAC on gait in PD patients (Table 7); and (3) in step time ΔM when assessing the influence of RAC on gait in HE subjects (Table 9). In these situations, the target experimental manipulation yields a pronounced change that dwarfs device-related measurement error, and indicates a potentially good opportunity for SmartMOVE in the clinic.

The second pattern is defined as a small-to-medium effect size (η² ≈ .08) for RAC, and a very small (or perhaps “negligible”) effect size (η² < .005) for Device. This pattern was observed twice: for step time ΔCV in PD (Table 7), and for step length ΔM in HE (Table 10). Here, the RAC manipulation (self-paced vs. metronome-cued) yielded a less pronounced (though still statistically significant) group-level change, likely due to increased heterogeneity in the way individual subjects responded (as is suggested by the relatively larger error bars). This more complex finding indicates a possible opportunity for SmartMOVE in the clinic, with the knowledge that individual differences may contribute more strongly in these cases.
The third pattern is defined as an effect size for RAC that is either too similar to Device (step length $\Delta CV$ in PD; Table 8) or smaller than Device (step time $\Delta CV$ and step length $\Delta CV$ in HE; Table 9 and Table 10). Such a pattern indicates a poor opportunity for SmartMOVE in the clinic; that is, conventional gait analysis systems should be used.

Why does $\Delta CV$ exhibit greater discrepancies between SmartMOVE and conventional gait analysis, particularly when quantified in HE subjects? A possible explanation for this may be offered.

Step time events were obtained directly from the timestamps of acceleration waveform peaks. By contrast, step length events are derived from machine learning regression models; as such, individual step lengths are associated with stochastic measurement error. Step length estimation involves double integration over the entire waveform segment between detected heel strike analogues (i.e., waveform peaks). These segments contain some degree of noise, due to limited device resolution (i.e., 100-Hz sampling rate) and unintended device movement during walking. Stochastic measurement error on individual steps can, in turn, translate into inflated estimates of step length variability. This effect is particularly noticeable in the case of HE, as those subjects have lower GV to begin with. Measuring RAC-mediated changes in GV in HE subjects (as opposed to PD patients), however, is likely to be of lower clinical interest. As a result, the increased measurement error associated with SmartMOVE in this experimental condition should carry less weight when considering the overall utility of SmartMOVE.

The use of effect sizes as an aid to determine “good opportunities” or “possible opportunities” for SmartMOVE in clinical research is just one interpretation of the evidence presented here. Individual clinicians will have unique requirements with respect to the level of measurement accuracy required to make informed decisions about the care and treatment of individual patients. For a clinician or researcher without access (or with limited access) to a conventional gait analysis system, however, SmartMOVE provides the ability to perform quantitative analysis of gait and gait variability or to assess the potential efficacy of an RAC paradigm where no opportunity was previously available.

---

### Table 7. Statistics associated with the RAC $\times$ Device ANOVA for step time outcome measures in PD patients.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\Delta_M$</th>
<th>$\Delta_M$</th>
<th>$\Delta_V$</th>
<th>$\Delta_V$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>F</td>
<td>p</td>
<td>$\eta^2$</td>
</tr>
<tr>
<td>Device</td>
<td>$3.54 \times 10^{-7}$</td>
<td>5.50</td>
<td>.039</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RAC</td>
<td>$3.34 \times 10^{-2}$</td>
<td>40.12</td>
<td>&lt;.001</td>
<td>.3067</td>
</tr>
<tr>
<td>RAC $\times$ Device</td>
<td>$3.53 \times 10^{-8}$</td>
<td>.13</td>
<td>.883</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

“SS” is the portioned sums of squares for each ANOVA term. Significant $p$-values are highlighted in bold.

doi:10.1371/journal.pone.0141694.t007

---

### Table 8. Statistics associated with the RAC $\times$ Device ANOVA for step length outcome measures in PD patients.

<table>
<thead>
<tr>
<th>Effect</th>
<th>$\Delta_M$</th>
<th>$\Delta_M$</th>
<th>$\Delta_V$</th>
<th>$\Delta_V$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>F</td>
<td>p</td>
<td>$\eta^2$</td>
</tr>
<tr>
<td>Device</td>
<td>$1.96 \times 10^{-5}$</td>
<td>.34</td>
<td>.571</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RAC</td>
<td>$2.67 \times 10^{-2}$</td>
<td>18.32</td>
<td>&lt;.001</td>
<td>.0517</td>
</tr>
<tr>
<td>RAC $\times$ Device</td>
<td>$2.38 \times 10^{-5}$</td>
<td>.84</td>
<td>.444</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

“SS” is the portioned sums of squares for each ANOVA term. Significant $p$-values are highlighted in bold.

doi:10.1371/journal.pone.0141694.t008
2. SmartMOVE-based gait analysis: Caveats and considerations

Three caveats should be noted with respect to choices made in the present experimental methodology, along with our rationale for those choices.

A first caveat relates to the size and clinical characteristics of the present sample. For safety reasons, the inclusion/exclusion criteria for the present study (see Section 1 of the Methods), precluded patients with severe gait dysfunction; most patients in the present sample would be considered to have "moderately advanced" PD. Whether SmartMOVE would perform as well in the case of severe gait dysfunction (e.g., shuffling steps or frequent gait freezing episodes) is thus unknown. Quantifying differences between gait or GV in severe PD versus HE, however, would likely have little diagnostic value. Furthermore, patients with severe gait dysfunction are, most likely, poor candidates for home-based, long-term interventions using cueing strategies such as RAC, and thus outside the “target demographic” which would get the most benefit from a tool like SmartMOVE. Nevertheless, a larger and wider sampling across the PD spectrum—within the limits of patient safety—is a valuable future step for this project, both from the perspective of clinical inference and with respect to improving the accuracy and robustness of the machine learning algorithms. Note that the present study’s attrition rate (i.e., yielding samples of 12 PD and 12 HE instead of the target 15 PD and 15 HE) was due to a combination of factors necessitated by stringent methodological requirements (see Section 1 of the Methods), not because smartphone-based gait analysis is inherently more challenging than conventional gait analysis.

A second caveat relates to the placement of SmartMOVE on the body (i.e., affixed at the navel). As reviewed in Table 1, previous investigations of smartphone-based gait analysis have positioned the device in a variety of locations; most commonly, over the third lumbar vertebra, as is common in traditional (i.e., non-smartphone) investigations of accelerometry [58,59]. The decision to affix the smartphone on the ventral surface of the body was made with an eye towards two future applications. First, to enhance usability, so that a user (patient) could interact with the device without assistance (e.g., by slipping it in and out of a chest harness). Second, to take advantage of possible camera or video recording applications, which would require that the smartphone’s camera face forwards.

Table 9. Statistics associated with the RAC × Device ANOVA for step time outcome measures in HE subjects.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Δ_M</th>
<th>Δ_CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>6.24 × 10^-8</td>
<td>5.71 × 10^0</td>
</tr>
<tr>
<td>RAC</td>
<td>2.63 × 10^-2</td>
<td>3.55 × 10^0</td>
</tr>
<tr>
<td>RAC × Device</td>
<td>1.86 × 10^-6</td>
<td>1.52 × 10^-2</td>
</tr>
</tbody>
</table>

*Δ_M* is the portioned sums of squares for each ANOVA term. Significant p-values are highlighted in bold.

Table 10. Statistics associated with the RAC × Device ANOVA for step length outcome measures in HE subjects.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Δ_M</th>
<th>Δ_CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>5.53 × 10^-4</td>
<td>1.00 × 10^1</td>
</tr>
<tr>
<td>RAC</td>
<td>2.23 × 10^-2</td>
<td>6.55 × 10^0</td>
</tr>
<tr>
<td>RAC × Device</td>
<td>7.02 × 10^-5</td>
<td>7.18 × 10^-1</td>
</tr>
</tbody>
</table>

*Δ_M* is the portioned sums of squares for each ANOVA term. Significant p-values are highlighted in bold.
A third caveat relates to the outcome measures calculated by SmartMOVE. In the current paper, we focused on two mathematically straightforward and outcome measures derived from the first-order difference of inter-step times and inter step lengths: $\Delta M$ and $\Delta CV$, which are widely reported in the clinical literature (e.g., [8–10]). Of course, there are many other outcome measures of gait, which can be summarized in three groups. “Group 1” measures are derived exclusively from a heel strike event series for calculation, such as $\Delta M$ and $\Delta CV$, as well as detrended fluctuation analysis (e.g., [6,93]). “Group 2” measures require both heel-strike and toe-off events for calculation, and include statistics such as stance or swing time means, SDs, or CVs, or asymmetry values (e.g., [13,94]); and the phase coordination index [95]. “Group 3” measures are derived from frequency-domain analysis of a continuous acceleration signal (as in [61,64,66,67]).

Logically, the optimal outcome measures of GV to investigate are those which best differentiate gait in PD versus gait in HE, or differences in gait between self-paced and metronome-cued walking. A meta-analysis across prior investigations would provide such an answer; unfortunately, no meta-analysis of GV in PD (or PD versus HE) exists. Thus, our choice of step time and step length $\Delta M$ and $\Delta CV$ (which are all Group 1 measures) was jointly motivated by their prevalence in the clinical literature, as well as by three restrictions imposed by the present experimental design.

First, the practical limitations being able to collect no more than 10 to 12 consecutive steps on a single walk across the 7-m GAITRite mat precluded the exploration of detrended fluctuation analysis, as there continues to be debate in the literature concerning (1) the degree to which "stitching together" multiple short segments of data violates the assumptions of long-range correlations [96], and (2) the minimum number of gait events (steps or strides) required to reliably differentiate normal versus pathological gait—a number which may be in the hundreds [97,98]. Second, the lack of a clear acceleration waveform analogue to a toe-off event precluded the calculation of Group 2 outcome measures, including the phase coordination index. Third, $\Delta M$ and $\Delta CV$ measures have been carefully evaluated in terms of their test–retest reliability (as reviewed in [8]), clinical effectiveness (via meta-analysis, as in [26]), and construct validity (via factor analysis, as in [11]; or correlation analysis, as in [83]). No Group 3 measure has been similarly scrutinized.

A final comment related to outcome measures. In the present analyses, care was taken to ensure the same number of gait events across outcome measures (i.e., step time and step length series had the same number of elements per trial) and across devices (i.e., same number of events collected by SmartMOVE and heel contact–based measurement device). We did not, however, control the number of events across subjects, other than to set a maximum event count (= 50) per RAC condition. Although there was no systematic difference in number of gait events between groups or across RAC conditions (see S1 Text), increasing the number of events per condition (and setting it as a constant across all subjects) may yield more stable outcome measures, thereby reducing both inter-subject differences and inter-device measurement error.

3. SmartMOVE-based gait analysis: Future aims

The present set of results suggests that SmartMOVE offers moderate-to-high accuracy in characterizing differences in first- second-outcome measures of gait between PD patients and HE subjects, and also in characterizing changes in first-moment outcome measures of gait during an RAC paradigm. Three important future research aims emerge directly from these findings.

A first future aim is to improve the stability of device placement on the torso, with an eye towards minimizing errors in step timing and step length estimation. Currently, SmartMOVE
is optimized for clinic-based use; i.e., it requires the active participation of a well-trained user to ensure that the quality of collected data is high. Extending the utility and validity of SmartMOVE outside the clinic is an important goal.

A second future aim is to explore other situations in which SmartMOVE could be incorporated into the clinic (e.g., characterizing differences between On versus Off medication states, or freezers versus non-freezers), as well as a more formal assessment of the classification accuracy of SmartMOVE relative to heel contact based gait analysis using machine learning techniques [99–101]. A smartphone-based tool with the ability to identify individuals with atypical gait or GV characteristics (e.g., relative to a large sample of age-matched walkers who have previously been tested using smartphone-based gait analysis) would be a highly useful tool in a clinician’s or a physician’s toolbox. In conjunction with this, further work to improve the user interface experience with SmartMOVE (e.g., numeric or graphical representations of data) will be performed.

A third future aim is to evaluate the efficacy of SmartMOVE-enhanced home-based gait training using RAC. RAC has been the focus of numerous previous investigations, systematic reviews [22–24], and meta-analyses [25,26]. For this reason, a detailed exploration of RAC efficacy was not a key feature of the present design, but rather, an illustration of the feasibility of implementing an RAC paradigm on a smartphone. Several previous methods for portable RAC delivery combined with gait analysis have been proposed [50,56,102–104]; all, however, have been designed around the use of footswitches or shoe-mounted IMUs to collect the gait data itself. SmartMOVE captured significant changes in step time $\Delta M$ and step time $\Delta CV$ in PD with very low device-related measurement error (see Table 5), indicating that these outcome measures in particular could be used to monitor patients’ progress during home-based RAC.

An additional advantage of smartphone-based implementation, in line with the above aim, is the ability to link with commercial streaming music services such as Deezer [105], Rdio [106], or Spotify [107]. Such links would enable the delivery of an RAC paradigm that is personalized to an individual patient’s needs (i.e., optimal tempo) and preferences (i.e., favorite musical style or genre). Along these lines, related work from our lab [108,109] has developed a search engine specifically geared towards identifying and retrieving commercial music recordings that maintain a steady tempo—a prerequisite for rhythmic exercise and/or rehabilitation applications. Such personalization is integral to what Boonstra et al. [110] define as the “take home” message with respect to the effective use of cueing paradigms for gait training in PD: that it “should not be prescribed as a ‘one size fits all’ treatment”. An explicit validation of the efficacy of SmartMOVE-based gait training using RAC (relative to “conventional” methods of RAC delivery) is a key future aim.

Conclusion

As the burden of Parkinson’s disease continues to expand, new strategies and tools—particularly those rooted in telemedicine—will be required to meet it. Here, we describe the foundations of a smartphone-based application, SmartMOVE, that provides clinicians and researchers with a new tool for performing gait analysis: a “hybrid” between the ease and convenience of stopwatch-based assessments and the high accuracy and detail of component-based (sensor plus software plus display) gait analysis systems. In addition, SmartMOVE provides a means to close the loop between the quantification of widely used outcome measures and the delivery of personalized rhythmic auditory cueing—paving the way towards establishing “RAC 2.0”. We hope that technologies like SmartMOVE may one day serve as an accessible tool for the detection of gait dysfunction and an effective nonpharmacological adjuvant for its treatment.
Supporting Information

S1 Dataset. Subject-level outcome measure values for the step time ANOVAs.
(XLSX)

S2 Dataset. Subject-level outcome measure values for the step length ANOVAs.
(XLSX)

S1 Table. Group-level means and standard deviations associated with step outcome measures.
(DOCX)

S2 Table. Group-level means and standard deviations associated with step length outcome measures.
(DOCX)

S1 Text. Gait analysis preprocessing steps.
(PDF)

S2 Text. Inter-group, inter-task, and inter-device variance.
(PDF)

Acknowledgments

A number of individuals have provided valuable assistance, insights, and feedback during the long gestation of this project; at Beth Israel Deaconess Medical Center: Kathrin Czarnecki, Scott Kaplan, and Daniel Tarsy; at Singapore General Hospital: Jacqueline Chen, Christal Chiang, Lee Meng Ho, and Patsy Tan; and at National University of Singapore: Hugh Anderson, Sam Fang, and Cynthia Hsu. This project was partially funded by the National Research Foundation (NRF) under grant R-253-300-001-490 and managed through the multi-agency Interactive & Digital Media Programme Office (IDMPO) hosted by the Media Development Authority of Singapore (MDA) under Centre of Social Media Innovations for Communities (COSMIC). Additional funding was provided by the Agency for Science, Technology and Research (A*STAR) under grant R-252-000-510-305, and by the Singapore Ministry of Education (MOE) under grant R-252-000-516-112.

Author Contributions

Conceived and designed the experiments: RJE YSN GS YW. Performed the experiments: RJE YSN SZ DMT. Analyzed the data: RJE SZ BA. Wrote the paper: RJE YSN SZ DMT BA GS YW.

References


25. de Dreu MJ, van der Wilk ASD, Poppe E, Kwakkel G, van Wegen EEH. Rehabilitation, exercise therapy and music in patients with Parkinson’s disease: a meta-analysis of the effects of music-based...


