Validating an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease

*This is the corrected version of the paper originally published in ACM MM'14

Shenggao Zhu[1], Robert J Ellis[2], Gottfried Schlaug[3], Yee Sien Ng[4], and Ye Wang[1,2]
[1]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
[2]School of Computing, National University of Singapore, Singapore
[3]Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, USA
[4]Department of Rehabilitation Medicine, Singapore General Hospital, Singapore
shenggaozhu@nus.edu.sg; {ellis, wangye}@comp.nus.edu.sg;
gschlaug@bidmc.harvard.edu; ng.yee.sien@sgh.com.sg

ABSTRACT

Movement disorders such as Parkinson's disease (PD) will affect a rapidly growing segment of the population as society continues to age. Rhythmic Auditory Cueing (RAC) is a well-supported evidence-based intervention for the treatment of gait impairments in PD. RAC interventions have not been widely adopted, however, due to limitations in access to personnel, technological, and financial resources. To help "scale up" RAC for wider distribution, we have developed an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) mobile application to deliver RAC and assess motor performance in PD patients. The touchscreen of the mobile device is used to assess motor timing during index finger tapping, and the device's built-in tri-axial accelerometer and gyroscope to assess step time and step length during walking. Novel machine learning-based gait analysis algorithms have been developed for iRACE, including heel strike detection, step length quantification, and left-versus-right foot identification. The concurrent validity of iRACE was assessed using a clinic-standard instrumented walking mat and a pair of force-sensing resistor sensors. Results from 10 PD patients reveal that iRACE has low error rates ($<\pm 1.0\%$) across a set of four clinically relevant outcome measures, indicating a potentially useful clinical tool.

Categories and Subject Descriptors

K.4.2 [**Computing Milieux**]: Computers and Society—*Social Issues*; H.1.2 [**Information Systems**]: Models and Principles—*User/Machine Systems*; J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms

Design, Experimentation, Measurement

Keywords

Rhythmic Auditory Cueing (RAC); Mobile app; Motor performance; Gait analysis; Tapping; Concurrent validity

MM'14, November 3–7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2647868.2654952.

1. INTRODUCTION

A 2007 review of published prevalence studies projects that the number of individuals over the age of 50 with Parkinson's disease (PD) living in 15 of the world's most populous nations will double between 2005 (roughly 4.4 million) and 2030 (roughly 9.0 million) [11]. Although administration of carbidopa/levodopa remains the "gold-standard" treatment for motor impairments in PD [32], gait parameters such as cadence, stride length, and velocity remain significantly reduced in PD patients relative to age-matched healthy controls, even when patients are tested during the "peak" effect of medication [25]. Together, these concerns motivate the search for additional strategies or therapies to help maintain motor functions in PD patients.

1.1 Rhythmic Auditory Cueing for PD

The use of physical therapy for the treatment of gait impairments in PD (e.g., bradykinesia, freezing, falling) has been the subject of systematic reviews and "best practice" treatment recommendations for therapy delivery [20]. A specific, evidence-based treatment recommendation was the use of *rhythmic sensory cueing* (RSC)—in particular, rhythmic auditory cueing (RAC) (for reviews, see [22, 36]). RAC is the use of an auditory pacing stimulus (either a simple metronome, or music with a steady beat) to which patients attempt to synchronize while walking.

The beneficial effects of RAC on gait in PD have been noted for several decades. Single-session RAC leads to improvements along multiple gait parameters (e.g., velocity, stride length, and stepping rate [22]), and a handful of multi-week interventions have found sustained improvements in gait parameters during a post-intervention follow-up (e.g., [7, 33]). Perhaps most importantly, RAC leads to a reduction in *motor timing variability* (MTV), quantified as stride-to-stride timing fluctuations during walking (for a detailed discussion, see [15]). PD patients have significantly higher MTV during gait than healthy controls, even under normal medication regimens [23]. Furthermore, MTV is both prospectively and retrospectively associated with fall risk [14]. The incidence of falls in PD is high: an estimated two-thirds of patients fall at least once a year, and half experience multiple falls per year [13]. Therefore, reduced MTV by RAC means less falls (and thus less cost of falls) in PD.

1.2 Assessment of PD Motor Performance

Although the efficacy of RAC for PD (i.e., statistically significant improvements in gait parameters) is well sup-

ported (e.g., [8, 22, 36]), RAC has not been widely adopted due to the lack of affordable and convenient *clinimetric tools* that would enable a medical professional (e.g., neurologist, physical therapist) to accurately assess a patient's motor performance in the short term, which could provide insight into whether that patient would respond to RAC in the longer term [7, 33].

With the rapid growth of mobile devices (e.g., smartphones) and the emergence of mobile cloud computing technology [10], mobile devices are becoming a ubiquitous computing platform in our daily life and healthcare. Compared to traditional devices in human gait assessment, such as foot-mounted force sensors [16], sensor-embedded walkways [27] and motion-capture cameras [30], mobile devices have clear advantages in portability, convenience, cost efficiency, and comprehensiveness of platform features (e.g., on-device data collection, analysis and visualization).

Here, we describe such a tool to facilitate and "scale up" RAC interventions: an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) mobile application ("app"). iRACE operates on an iPhone or iPod touch, and connects to a cloud-based website. iRACE records activities during walking (via the on-board accelerometer and gyroscope) and index finger tapping (via the touchscreen). Motor production is assessed both with and without RAC, and the auditory cues can be highly customized by each user. The recorded walking and tapping data is analyzed to obtain the basic motor parameters (step time/length, tapping intervals, etc.), as well as a set of four clinically relevant outcome measures. These results may then be visualized on the iRACE app as well as the website.

We have developed novel machine learning-based algorithms for iRACE gait analysis, including heel strike detection, step length quantification, and left-vs.-right foot identification. These algorithms were validated against ground truth measures obtained from a clinic-standard instrumented walking mat and force-sensing resistor sensors. Results from 10 PD patients revealed that iRACE had less than $\pm 1.0\%$ error rates (averaged across subjects) across four outcome measures, indicating a potentially useful clinical tool.

Using iRACE, medical professionals can easily assess whether RAC is effective to an individual patient, and if so, use iRACE to deliver long-term RAC-based therapy and monitor the patient's performance over time, benefiting the patient physically, mentally, and financially.

The rest of the paper is structured as follows. Section 2 reviews related literature on smartphone-based gait analysis. Section 3 describes the iRACE system. The clinical validity study design of iRACE is outlined in Section 4. iRACE-based gait analysis, clinical outcome measures, and tapping analysis are detailed in Sections 5, 6 and 7, respectively. Experiment results are presented in Section 8. Section 9 concludes our work and explores future directions.

2. RELATED WORK

A number of recent studies have investigated the viability of smartphone's built-in inertial measurement unit (IMU; accelerometer and/or gyroscope) to measure and quantify human gait [6, 18, 21, 26, 31, 34, 42, 43], with key details summarized in Table 1. A careful reading of this group of studies, however, reveals several concerns.

First, in some cases [6, 18, 21], only a single healthy subject's data was examined, precluding the ability to infer whether the specific algorithms or analyses that were used would generalize across individuals.

Table 1: Smartphone-based gait investigations

Ref.	Subject	N	Device	Location	Fs
[6]	HC	1	iOS	Left pocket	100
[18]	HC	1	Android	Front waist	60
[21]	HC	1	iOS	Left ankle	100
[26]	HC	49	Android	L3	50
[31]	HC	30	Android	L3	33
[34]	HC	49	Android	L3	50
[42]	HC/RA	20/39	Android	L3	33
[43]	HC	13	Android	L3	100

Notes: N is the number of healthy controls (HC) or patients with rheumatoid arthritis (RA) tested. "L3" indicates device placed in the vicinity of the third lumbar vertebrae on the lower back. Fs is the sampling frequency of the device's IMU in Hz.

Second, in some cases [21, 42], no concurrent ground-truth gait analysis was utilized, precluding any assessment of the sensitivity, specificity, and temporal precision of smartphone-derived calculations. In other cases [26, 31, 43], larger samples of healthy adults were studied, and a conventional waist-mounted tri-axial accelerometer was employed as ground truth. Such a design, however, is suboptimal, in that it would reveal only whether the smartphone's accelerometer functions as well as a conventional accelerometer, rather than whether the smartphone *actually* (and accurately) captures the spatio-temporal dynamics of human gait, as measured by "gold standard" devices: foot-mounted force-sensing resistors [16], sensor-embedded walkways [27], ground reaction force plates [39], or motion-capture cameras [30].

Third, some studies [21, 31, 42, 43] did not utilize traditional step-level spatio-temporal analysis (i.e., heel strike detection and step time/length quantification), which is standard in classic IMU-based gait analysis [46, 47]. Instead, these studies used statistics which quantify properties of the raw accelerometer waveform, such as its peak frequency after a fast Fourier transform, or the period at which successive peaks appear in the waveform's autocorrelation. Although such statistics are mathematically *valid*, they lack clinical *utility*, as there are no large-scale studies which provide normative or relative (e.g., PD patients versus age-matched controls) values for these statistics, unlike traditional outcome measures (for reviews, see [17, 24]; cf. Section 6 below).

Finally—and most relevant to the current project—no previous study has evaluated the validity of smartphone-based gait analysis in PD patients.

3. SYSTEM DESCRIPTION

iRACE consists of three main components: an iOS-based mobile device such as an iPhone or iPod touch (the "iRACE device"), a mobile application (the "iRACE app"), and a back-end cloud computing service (the "iRACE website"). An iPhone or iPod touch is equipped with an inertial measurement unit (IMU) comprising a microelectromechanical tri-axial linear accelerometer and Coriolis vibratory gyroscope, each with a maximum nominal sampling frequency of 100 Hz. The device's capacitive touchscreen has a maximum refresh rate of 60 Hz. The IMU data and touchscreen tap events are recorded during walking and tapping, respectively, which are then uploaded to the iRACE website.

Figure 1: iRACE app screenshots. (a.) Home screen. (b.) Settings for evaluation parameters and playlists. (c.) RAC bimanual tapping. (d.) Rapid tapping. (e.) Tapping analysis and visualization.

3.1 iRACE App

3.1.1 Overview

iRACE administers three widely used tests of motor performance of both walking and finger tapping movements: (1) self-paced bilateral performance, (2) synchronization–continuation bilateral performance, and (3) rapid unilateral and bilateral performance. Launching iRACE (Figure 1a) provides the option to conduct a Walking Evaluation (using tests 1–2) or a Tapping Evaluation (using tests 1–3).

iRACE has a number of flexible settings (Figure 1b), including the number of iterations of each test and the desired playlist of RAC stimuli. The playlist can be pre-defined and fixed across subjects, or *dynamically* generated such that its range of tempos can automatically "bracket" each subject's self-paced walking/tapping tempo.

iRACE guides the user through all tests in succession. In a self-paced test, no RAC is provided; the subject will walk or tap at a rate deemed most comfortable. In a synchronization–continuation test, the RAC stimulus is delivered during the first part of the test (synchronization), during which the subject attempts to walk or tap with the stimulus beat. After a pre-specified number of events (steps or taps), the stimulus fades to silence over a few seconds, whereupon the subject attempts to continue the same pattern of movement at the same rate (continuation).

3.1.2 Walking and Tapping Evaluations

Quantitative evaluations of self-paced [23] and RAC-based synchronization (for reviews, see [22, 36]) walking are widely reported in the PD literature. During an iRACE Walking Evaluation, the iRACE device is secured to the ventral surface of the lower trunk (at the navel) using an elastic sleeve and waist band (Figure 2a). Although several previous studies have positioned the accelerometer on the dorsal surface of the body (e.g., over the third lumbar vertebrae; cf. Table 1), securing the device at the navel enables the user to more easily interact with the iRACE device (with an eye towards eventual home-based use). Both self-paced walking and synchronization–continuation walking tests are performed in this manner.

Index finger tapping is a widely used motor timing assessment in PD, both as a "fast-as-possible" test [12] and as a synchronization–continuation test [40]. To our knowledge, there exist no smartphone-implemented evaluations of finger tapping in either healthy controls or patients with movement disorders. During all tapping tests (self-paced, synchronization-continuation, and rapid; cf. Figure 1c–d), the iRACE device is positioned on a desk or table such that the subject's arms and wrists are comfortably supported, permitting smooth index finger movements. iRACE can process and visualize the tapping data in-app immediately after a test (Figure 1e). (In-app gait analysis will be implemented in a future version of iRACE.)

3.2 iRACE Website

A cloud-based service has been established to facilitate iRACE data management and visualization. Each iRACE user can create an account to login the iRACE app and website. The website has the following main features: (1) To build user profiles and manage user roles and relationships (patients, therapists, etc.); (2) To receive and analyze data files uploaded from iRACE app; (3) To visualize walking and tapping analysis results in tables and graphs; (4) To review evaluation history and track user progress over time.

4. CLINICAL STUDY DESIGN

This section describes the novel fusing of clinic-standard ground truth measures with iRACE, and the testing protocol into which they were incorporated.

4.1 Ground Truth Gait Measures

As noted in Section 2, no previous investigation has validated smartphone-based gait analysis using ground truth data that directly samples heel strikes in the temporal and/or spatial dimensions. In our study, we utilized both: a GAITRite electronic walkway[1], and a force-sensing resistor assembly[2] and telemetry unit (DataLOG[3]).

GAITRite is a 61 cm × 610 cm roll-up mat embedded with pressure sensors arranged with 1.27-cm spacing, with a maximum sampling frequency of 240 Hz. GAITRite computes various spatio-temporal gait parameters for each walk across the mat, such as cadence, step time and step length. Its accuracy has been validated by concurrent analyses using insole footswitch sensors [2] and motion-capture cameras [41]. GAITRite has been the primary data collection

[1]http://www.gaitrite.com
[2]http://www.biometricsltd.com/forcesensors.htm
[3]http://www.biometricsltd.com/datalog.htm

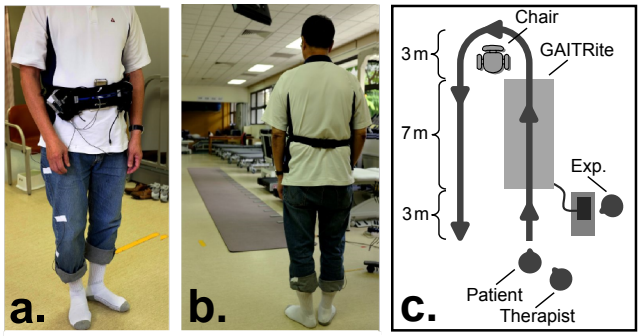

Figure 2: Clinical study settings. (a.) iRACE device and DataLOG secured to a belt on patient's lower trunk. (b.) Straight line walking segment on GAITRite. (c.) Illustration of the walking path.

device in numerous gait investigations in the elderly, including PD patients (e.g., [8, 24]).

In the present study, a force-sensing resistor (FSR) sensor was attached to the heel pad of each foot to record the moment of each heel strike. Both FSR sensors were wired to a telemetry unit (DataLOG) and sampled at 1000 Hz as binary values (0 if the pressure on FSR exceeds a certain threshold, and 1 otherwise). FSR sensors are widely used in gait analysis, including gait analysis in PD (for reviews, see [14, 15]). When a heel strikes the ground, the pressure placed on the FSR toggles its reading from 1 to 0; that toggling timestamp was used as the "ground truth" heel strike time. DataLOG also received an analog audio signal directly from the iRACE app, which was used to synchronize iRACE and DataLOG clocks (detailed in Section 5.2).

As DataLOG has a higher sampling frequency (i.e., with millisecond-level temporal resolution) than GAITRite, FSR and DataLOG recorded heel strike timestamps were used as the ground truth to validate the temporal precision of iRACE (cf. Section 5.3), and GAITRite was used to evaluate the step length algorithm of iRACE (cf. Section 5.5).

4.2 Subjects and Testing Protocol

Ten patients (4 female) diagnosed with idiopathic Parkinson's disease were recruited from Singapore General Hospital. The average age of the patients was 66.3 (range: 58 to 81), and average duration of PD was 5.7 years (range: 2 to 14). The patients had a Hoehn and Yahr stage of 2.0 ($N = 1$), 2.5 ($N = 5$), 3.0 ($N = 3$) and 4.0 ($N = 1$), where N is the number of patients. None of the patients had a coexisting dementia (Mini-Mental State Examination score > 24 points) or other diagnosed neurological impairments. All patients were under stable medication regimens for the preceding four weeks, and were tested at least 30 minutes after taking morning medications.

All testing procedures were approved by the institutional review board (IRB) at Singapore General Hospital. The iRACE "belt" and FSR sensors (Figure 2a) were attached as described above (cf. Section 3.1.2 and 4.1). The experimenter then demonstrated the target walking path (Figure 2c), including a 3-meter acceleration/deceleration phase to allow for steady-state walking while on the GAITRite mat.

Each subject then completed four or more walking trials (i.e., enough to record at least 40 steps on GAITRite) in each of three sequential RAC conditions: (1) self-paced (no RAC), (2) RAC with metronome tempo set to 100% of mean self-paced cadence, (3) RAC with metronome tempo set to 110% of mean self-paced cadence; these two tempos are common in RAC [36]. (The mean self-paced cadence was computed using GAITRite, as the data was collected *before* the algorithms detailed in Section 5 were developed.)

The iRACE Tapping Evaluation (cf. Section 3.1.2) was conducted after the completion of the Walking Evaluation.

5. WALKING ANALYSIS

This section describes the processing pipeline used to transform the raw IMU data recorded by iRACE into temporal and spatial gait parameters (step time, step length, etc.).

5.1 Preprocessing

During a Walking Evaluation, the iRACE device was oriented with its screen facing towards the subject and its audio jack facing up. iRACE simultaneously records 6 channels of IMU data in the iOS device xyz coordinate system: tri-axial acceleration (a_x, a_y and a_z) and tri-axial rotation rate (ω_x, ω_y and ω_z). For simplicity, from the subject's perspective, we define anterior–posterior (A–P) acceleration as $a_{AP} = -a_z$ (positive values for anterior acceleration), up–down (U–D) acceleration as $a_{UD} = -a_y$ (positive for upward acceleration), and left–right (L–R) acceleration as $a_{LR} = a_x$ (positive for leftward acceleration). Similarly, we define the rotation rate around the A–P axis (i.e., roll) as $\omega_{AP} = -\omega_z$, around the U–D axis (i.e., yaw) as $\omega_{UD} = -\omega_y$, and around the L–R axis (i.e., pitch) as $\omega_{LR} = \omega_x$.

Four preprocessing steps were applied: (1) IMU waveform interpolation; (2) straight-line walking identification; (3) device tilt compensation; and (4) low-pass filtering.

First, since the time interval between successive IMU samples varies stochastically, raw IMU data is usually interpolated and resampled at a fixed frequency [18, 26, 31]. In the present study, all IMU data was cubic-spline interpolated and resampled at 1000 Hz, which as a result could refine the peak and trough locations (i.e., improve temporal resolution) in the IMU waveforms (e.g., Figure 3a).

Second, an algorithm was implemented to isolate straight-line walking segments from turn segments (cf. Figure 2c), stationary segments, or other non-walking segments (e.g., freezing). This was achieved by analyzing the acceleration amplitude of a_{AP} (to identify actual walking) and rotation rate ω_{UD} (to identify turning). For simplicity, only straight-line segments will be analyzed further.

Third, because the iRACE device usually pitches (around the L–R axis) as the subject walks (thus affecting IMU waveforms and the accuracy of subsequent heel strike detection), the A–P and U–D waveforms were adjusted (using a trigonometric algorithm) according to the estimated device tilt angle (based on rotation rate ω_{LR} and average gravity in A–P and U–D axes; derived from [29]).

Fourth, all IMU data was low-pass filtered with a 4th-order zero-lag Butterworth filter with a cut-off frequency of 20 Hz, as suggested by [46, 47].

5.2 iRACE & Ground Truth Synchronization

Since the internal clocks in the iOS device and DataLOG are independent and not synchronized, a novel synchronization method was developed via an acoustic signal. Specifically, iRACE generated a 2-second square-wave audio (using the iOS Audio Units API). iRACE toggled the volume

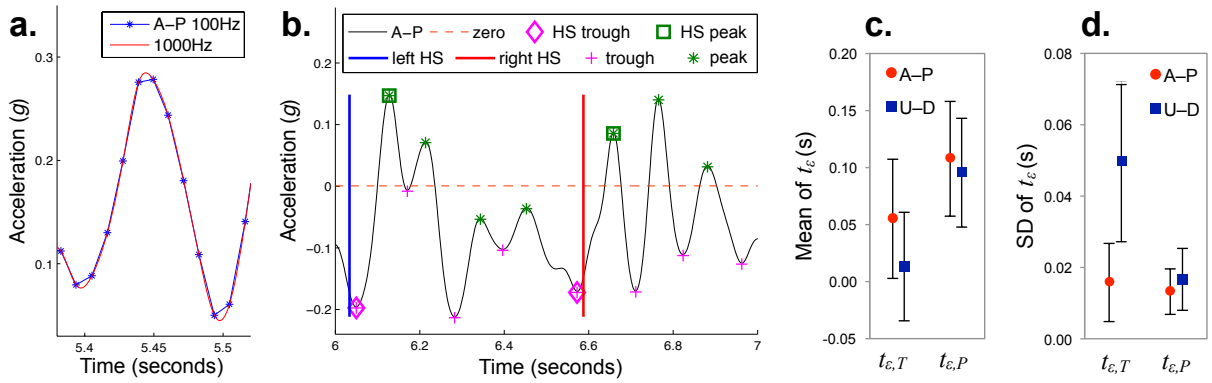

Figure 3: (a.) Interpolation and resampling of iRACE acceleration data. (b.) Annotation of heel strike points (HS peak and HS trough) in the A–P waveform. (c.) and (d.) Group mean and standard deviation (SD) of the temporal error (t_ε) between HS timestamps and annotated waveform peaks ($t_{\varepsilon,P}$) and troughs ($t_{\varepsilon,T}$).

of this audio repeatedly (approximately every 30 ms) between silence and normal volume, and saved the toggling timestamps (in iRACE's clock). DataLOG recorded this audio (at 2 kHz), which was then analyzed to locate these volume toggling timestamps (in DataLOG's clock), i.e., the rising and falling edges of the audio volume contour. The temporal offset between the two clocks was found by linearly regressing the DataLOG timestamps onto the iRACE timestamps; the y-axis intercept was the "best fitting" offset in the least-squares sense, with sub-millisecond level accuracy.

GAITRite records the session start time to the millisecond level, so that the steps measured by GAITRite can be correctly matched to those measured by DataLOG and iRACE.

5.3 Waveform Annotation & Timing Analysis

Previous accelerometry-based gait analysis methods have used waveform peaks (i.e., local maxima) in the A–P channel [26, 47] and/or U–D channel [28] as analogues of physical heel strike (HS) events (for a review, see [19]). As noted in Section 2 (cf. Table 1), however, the above papers were primarily focused on gait analysis in healthy individuals, and have not been extended to gait analysis in PD patients. Even a brief glance at the waveform excerpt from a PD patient in Figure 3b reveals that the widely used convention of "an A–P peak preceding a zero-crossing" as a HS analogue [44, 46, 47] is problematic, as *multiple* zero-crossings may be present between successive ground truth HS events (i.e., as recorded by DataLOG and FSR sensors).

To find out which accelerometry waveform event was the most temporally stable analogue of a true HS event, a more systematic approach was taken in the present study. For each ground truth HS timestamp (t_{HS}), the timestamps of a peak (t_P) and its preceding trough (t_T, i.e., local minimum) that were most likely to correspond to t_{HS} were manually annotated on the A–P and U–D accelerometry waveforms. Typically, the target waveform event (t_P or t_T) was closest to t_{HS} among all peaks (or troughs), but exceptions existed due to irregular waveform fluctuations (i.e., irregular gait). For convenience, manually annotated peaks and troughs will be referred to as "HS peaks" and "HS troughs", respectively (cf. Figure 3b).

After annotation, the signed temporal errors (t_ε) of four waveform event timestamps (i.e., $t_{P,AP}$, $t_{P,UD}$, $t_{T,AP}$, $t_{T,UD}$) relative to the corresponding t_{HS} were calculated for each t_{HS}, and the mean (M_{t_ε}) and standard deviation (SD_{t_ε}) of t_ε were computed for each subject. Figure 3c–d present the group-level mean and standard deviation (as error bars) of M_{t_ε} and SD_{t_ε} across all subjects, respectively. Logically, temporal *variability* (i.e., SD_{t_ε}) rather than temporal *delay* (i.e., M_{t_ε}) is the more critical statistic, as a physical delay is inherent between a HS event and the corresponding waveform peak/trough recorded in a waist-mounted accelerometer. A 2 (*Channel*: A–P or U–D) × 2 (*Event*: SD of $t_{\varepsilon,P}$ or $t_{\varepsilon,T}$) repeated-measures analysis of variance (ANOVA) was performed, revealing a significant effect for *Channel* ($F_{1,9} = 26.86$, $p = .0006$; i.e., A–P had lower SDs than U–D), *Event* ($F_{1,9} = 32.16$, $p = .0003$; i.e., HS peaks had lower SDs than HS troughs), and their interaction ($F_{1,9} = 15.96$, $p = .003$). Because A–P peaks ($t_{P,AP}$) had the smallest group-level mean and variance of SD_{t_ε}, they may be considered the most temporally stable analogue of HS events.

5.4 Heel Strike Detection

As highlighted in Figure 3b, conventional rule-based algorithms using hand-crafted rules or hard-coded thresholds (e.g., a simple zero-crossing rule to identify HS analogues) may be imprecise in PD patients, due to individual differences in disease staging, bradykinesia, freezing, and so forth.

To address this limitation, a machine learning-based HS detection algorithm was proposed, which formulates HS detection as a binary classification problem: whether a given accelerometry A–P peak was a HS analogue (i.e., a HS peak) or not. (A–P peaks are the most temporally stable waveform events relative to true HS events; cf. Section 5.3.)

Firstly, two sets of features were extracted for each A–P peak: (1) Features from the A–P peak (including the peak amplitude, the difference in time or amplitude between a peak and its adjacent troughs, etc.); and (2) Features from all accelerometer and gyroscope waveforms. The average step time (T_s; i.e., the duration of a step) was previously derived from the A–P waveform autocorrelation [43]. A segment (about $0.2\times T_s$) centered on the peak timepoint was selected in each waveform, and features (mean, standard deviation, etc.) from these segments were extracted and concatenated. These two sets of features together formed a feature vector, denoted as $\mathbf{f}_i$ for the i-th peak.

Next, a set of features based on a peak sequence (i.e., combinations of adjacent peaks) was constructed for the i-th peak, including $\mathbf{f}_i - \mathbf{f}_{i-1}$, $\mathbf{f}_i - \mathbf{f}_{i+1}$, $\mathbf{f}_i + 0.5\mathbf{f}_{i-1} - 0.5\mathbf{f}_{i+1}$, etc. After that, a set of the most statistically significant features

are selected among all the extracted features using the `rankfeatures`[4] function in MATLAB (with a two-sample t-test used as the class separability criteria).

A two-layer neural network classifier [3] then took the selected features as input. The ground truth of whether each peak was a HS peak was derived from the waveform annotation. Different numbers of selected features and units in the hidden layer of the classifier were iterated and tested; the optimal parameter combination was obtained, which yielded the highest classification accuracy based on 10-fold cross-validation (for more details about parameter selection, see [3]). In the present study, 100 selected features and 20 units in the hidden layer were used.

For each peak, a number (continuously valued from 0 to 1) was output by the classifier, indicating the probability of that peak being a HS peak. This number was then thresholded at 0.5, giving an initial prediction of the identity of a HS peak or a non-HS peak. Next, this initial prediction was refined, as follows. Based on the previously derived average step time (T_s), a pair of rules was applied: (1) If any two adjacent HS peaks were separated by less than $0.7\times T_s$, the one with the lower probability was adjusted to a non-HS peak; (2) If no HS peak existed in the window [$0.7\times T_s$, $1.3\times T_s$] before or after the timestamp of a given HS peak, the peak with highest probability in that window was selected as a HS peak.

5.5 Step Length Calculation

After HS detection, the spatial displacement (i.e., step length) between successive HS peaks was calculated. Algorithms for IMU-based step length calculation can be generally grouped into two categories: double integration of acceleration [5, 45], and machine learning-based regression methods [37]. For trunk-mounted smartphones, the main difficulties of the double integration method lie in the integration drift over time and the estimation of initial velocity at the beginning of a step (i.e., the HS peak time), since the trunk velocity at that time is typically non-zero. For example, in the present PD patients' data, if the initial velocity were set to zero (as assumed in [5]), the resultant step length based on double integration usually drifted to a negative value.

To overcome these difficulties, a hybrid machine learning algorithm was proposed, which took the double integration results together with other low-level waveform features as the input, and uses neural network regression [3] to calculate step length. Specifically, for each step, features were extracted from segments of tri-axial acceleration and the "energy waveform" (i.e., $\sqrt{a_{AP}^2 + a_{LR}^2 + (a_{UD}+1)^2}$; e.g., [6]) including single and double integration, time duration, standard deviation, and so on. These segments include the exact step segment (between successive HS peaks) and its shifted versions (by half a step duration before or after it). Similar features were also extracted from tri-axial gyroscope waveforms. Next, Principal Component Analysis [3] was used to select the most significant features, which were then input to a two-layer neural network regression model [3]. Model parameters were trained using 10-fold cross-validation.

5.6 Left-versus-Right Foot Identification

Left versus right foot identification is important for clinical gait measures, but presents some challenges when the measurement device is attached centrally at the trunk. Some studies use L–R acceleration to distinguish left and right foot, based on the sign of L–R acceleration, or the sign of its single or double integration [5, 46, 47]. These methods assume that the direction of trunk displacement is different for different feet. However, the validity of these assumptions has not been tested using smartphones, or in PD patients. We tested these methods in our PD patient data, and confirmed that they worked for a majority of patients, but some failure cases still existed due to weak and irregular L–R acceleration. At the same time, we observed that the gyroscope data (especially rotation rate around U–D axis) usually also showed different patterns for different feet, which relates to the different rotation directions of the trunk.

To fully utilize all accelerometer and gyroscope data, a machine learning-based method was proposed for left-versus-right foot classification. For each step between successive HS peaks, two sets of features were extracted. First, step-based features (single/double integration, maximum, minimum, etc.) from segments of all accelerometer and gyroscope waveforms. The durations of these segments include the exact step duration (between successive HS peaks) and its shifted versions (by half a step duration before or after it). These features are denoted as a vector $\mathbf{f}_i$ for the i-th step. Second, features from a step sequence, including $\mathbf{f}_i - \mathbf{f}_{i-1}$, $\mathbf{f}_i - \mathbf{f}_{i+1}$, $\mathbf{f}_i - 0.5\mathbf{f}_{i-1} - 0.5\mathbf{f}_{i+1}$, etc. The ground truth of left-vs.-right foot was obtained from DataLOG left and right foot sensors. As in the HS detection method (cf. Section 5.4), feature selection was performed, and the neural network classifier was evaluated using 10-fold cross-validation.

6. CLINICAL OUTCOME MEASURES

iRACE calculates four clinically relevant outcome measures that quantify specific statistical properties of an event series (ES) $\{e_1, e_2, \ldots, e_N\}$ of N successive finger tap or heel strike timestamps (in elapsed seconds, relative to an arbitrary "0" timepoint) or N successive heel strike displacements (in meters, relative to an arbitrary starting position). An inter-event series (IES) refers to the first-order difference (i.e., intervals) of an event series, with $N-1$ elements.

6.1 Outlier Flagging

Outliers in an ES come in two forms: false positives (i.e., a "double tap" of the same button during bimanual tapping, or a "double peak" in an A–P acceleration waveform; cf. Figure 3b), or false negatives (i.e., "missing" events due to inadequate heel strike pressure or touchscreen contact, or by *a priori* exclusion of turns during a walk; cf. Section 5.1).

For all walking and tapping trials other than Rapid Tapping, a simple "successive percentage change" (S_{SPC}) transformation was used to flag outliers in the original ES. For all adjacent elements of an IES, $S_{SPC,i}$ is defined as:

$$S_{SPC,i} = 100\frac{|IES_{i+1} - IES_i|}{IES_i}. \quad (1)$$

S_{SPC} will have $N-2$ elements relative to the original ES.

A single threshold percentage value (P) was then set; here, $P = 75$. For any $S_{SPC,i} \geq P$, the associated ES element e_{i+2} was flagged as an outlier. Additionally, any repeated event on the same side of the body (i.e., the second of two successive left or successive right taps or heel strikes) was flagged as an outlier. The first-order difference of a series of ≥ 6 consecutive ES elements without an outlier flag will be

[4] http://www.mathworks.com/help/bioinfo/ref/rankfeatures.html

labeled a "ΔES". (For rapid tapping trials, a simpler logic was applied: an outlier was identified as any IES_{i+1} which followed IES_i by less than 50 ms. This threshold was established to eliminate moments of near-synchronous tapping that are more likely to occur during alternate bimanual tapping, particularly at fast rates.)

6.2 Outcome Measures

Two statistics were calculated from a ΔES. First, the mean of a ΔES (Δ_M) is straightforward, and can be taken for either a timestamp ΔES ($\Delta_{M,T}$) or a displacement ΔES ($\Delta_{M,D}$). Two transformations yield the commonly reported statistics of *cadence* (= $60/\Delta_{M,T}$, in steps-per-minute) and *velocity* (= $\Delta_{M,D}/\Delta_{M,T}$, in meters-per-second). Second, the coefficient of variation of a ΔES (Δ_{CV}) is widely used in analyses of motor timing in PD [15]:

$$\Delta_{CV} = 100\frac{\text{std}(\Delta ES)}{\text{mean}(\Delta ES)}, \quad (2)$$

where std(·) is the sample standard deviation.

Two other statistics were derived from a transformation of an outlier-free ES proposed by [35] to quantify motor timing during walking. For each string of N ($N \geq 3$) right (e_R) and N left (e_L) events without an outlier flag, $\{e_{R,1}, e_{L,1}, e_{R,2}, e_{L,2}, \ldots, e_{R,N}, e_{L,N}\}$, two relative phase series were defined using the right (Φ_R) or left (Φ_L) foot as the "reference" foot (the following formulas assume the first R event precedes the first L event):

$$\Phi_{R,i} = 360\frac{e_{L,i} - e_{R,i}}{e_{R,i+1} - e_{R,i}}, \quad \Phi_{L,i} = 360\frac{e_{R,i+1} - e_{L,i}}{e_{L,i+1} - e_{L,i}}. \quad (3)$$

For example, a $\Phi_{R,i} = 180$ indicates that the left event $e_{L,i}$ precisely bisects the time (or displacement) between successive right events $e_{R,i}$ and $e_{R,i+1}$.

The mean absolute percentage deviation from 180 (Φ_M) indicates the presence of an overall phase asymmetry (e.g., right steps do not evenly divide left strides, or vice versa):

$$\Phi_{M,R} = 100\frac{\text{mean}(|\Phi_R - 180|)}{180}, \quad (4)$$

$$\Phi_{M,L} = 100\frac{\text{mean}(|\Phi_L - 180|)}{180}. \quad (5)$$

The coefficient of variation of Φ (Φ_{CV}) was calculated in an analogous fashion to Δ_{CV}:

$$\Phi_{CV,R} = 100\frac{\text{std}(\Phi_R)}{\text{mean}(\Phi_R)}, \quad \Phi_{CV,L} = 100\frac{\text{std}(\Phi_L)}{\text{mean}(\Phi_L)}. \quad (6)$$

The nature of the Φ transform means that Φ_{CV} reflects *local* (i.e., event-to-event) variability relative to the mean, while Δ_{CV} reflects *global* variability relative to the mean.

For parsimony, the maximum of $\Phi_{M,R}$ and $\Phi_{M,L}$ (referred to simply as Φ_M), and the maximum of $\Phi_{CV,R}$ and $\Phi_{CV,L}$ (referred to as Φ_{CV}) will be retained for a given test. Thus, any reduction in the magnitude of Δ_{CV}, Φ_M, or Φ_{CV} would reflect an improvement on that particular aspect of gait stability, which may be caused by the positive facilitation of RAC (relative to without RAC).

7. TAPPING ANALYSIS

Two limitations that occur when recording iOS touchscreen events have the potential to affect the outcome measures of iRACE tapping. First, this touchscreen has a relatively slow maximum sampling frequency (60 Hz) compared to the iOS IMU (100 Hz). Second, the touchscreen datastream is effectively binary (0 = no contact; 1 = contact) rather than continuously valued like the IMU, precluding it from interpolation (and the resulting refinement of temporal resolution; cf. Figure 3a). Together, these limitations suggest that iRACE estimates of inter-tap interval variability (i.e., Δ_{CV} and Φ_{CV}) may be inflated.

For example, as an IES becomes faster (i.e., with a smaller Δ_M) or more temporally stable (i.e., a smaller Δ_{CV}, Φ_M, or Φ_{CV}), the impact of 60-Hz sampling error becomes more prominent; the potential error in each timestamp is a larger proportion of the underlying period. This is highlighted in Figure 4a, which visualizes an inter-tap interval series with a true $\Delta_M = 0.25$ s and $\Delta_{CV} = 2.0\%$ (blue trace). However, after downsampling this series at 60 Hz (by advancing each timestamp in the original series forward to the next possible 60-Hz sample point, yielding the red trace), the observed Δ_M is accurate (= 0.25 s) but the observed Δ_{CV} is inflated (= 3.8%). By contrast, in Figure 4b, a series with a true Δ_M = 1.0 s and $\Delta_{CV} = 8.0\%$ has both an accurate observed Δ_M (= 1.0 s) and Δ_{CV} (= 8.1%) after downsampling.

To systematically quantify the error induced by 60-Hz sampling on the four temporal outcome measures (Δ_M, Δ_{CV}, Φ_M, and Φ_{CV}), the following simulation was conducted. A random "pull" of 30 values (which forms an IES) was drawn from the normal distribution ($\mu = 0$, $\sigma = 1$), and shifted to have a mean of M (where $M = 0.125$ to 1.0 in ten geometric steps). The corresponding ES series was generated by taking the cumulative sum of the IES. Next, the value of Δ_{CV}, Φ_M, or Φ_{CV} was rescaled (over separate manipulations) to have a value of exactly P% (where $P = 2.0$, 4.0, or 8.0). The rescaled ES was then downsampled to 60 Hz. Finally, the four outcome measures were calculated from the downsampled timestamps. In total 10 (M) × 3 (P) × 2000 iterations were performed.

The simulation results are presented in Figure 4c–f, plotting the error induced by 60-Hz sampling (y-axis) at different values of M (x-axis) and P (different colors) for each outcome measure (separate panels). The median (heavy line) and central 95% (lighter lines) of all iterations are shown for each P value. Several conclusions can be made. First, Δ_M is accurately captured at all levels of P; this finding is perhaps expected, as even with a fast M (e.g., Figure 4a), the numerous underestimated and overestimated elements (relative to the simulated ground truth) tend to cancel each other out. Second, as predicted, the accuracy of all outcome measures increases (i.e., the y-axis width between each pair of 95% lines decreases) as M increases. Most importantly, for the range of "plausible" walking and tapping rates (i.e., $\Delta_M \geq 0.5$ s, or 120 events-per-minute or slower), outcome measures differ from ground truth by less than ±1.0%.

8. RESULTS

This section presents the concurrent validation results, including the accuracy of iRACE-quantified gait parameters, outcome measures, and measures of RAC facilitation.

8.1 Gait Measurement Accuracy

A heel strike dataset of 3898 HS events was combined across all Walking Evaluations of all ten subjects. Each HS event has both a ground truth HS timestamp from the DataLOG FSR sensor, and an annotated HS peak in iRACE A–P waveform. A total of 16295 A–P peaks (including HS

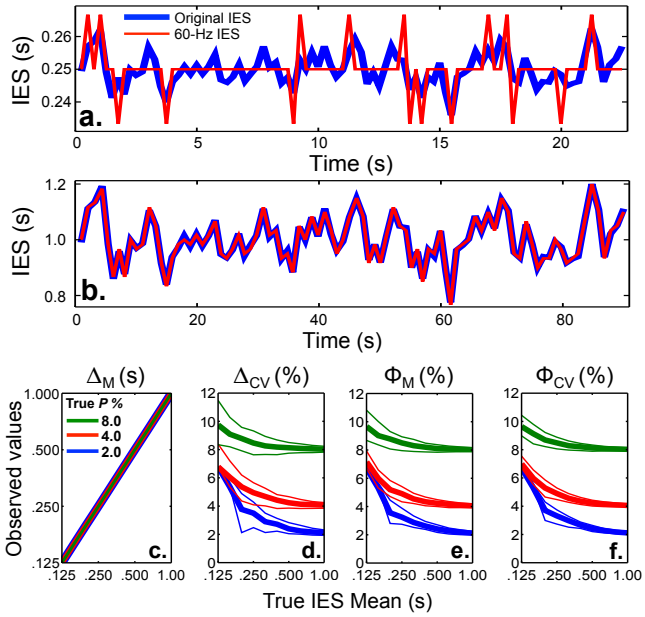

Figure 4: Tapping simulation. (a.) and (b.) 60-Hz downsampled inter-event interval (IES) series with a mean of 0.25 s and 1.0 s. (c.–f.) Outcome measures.

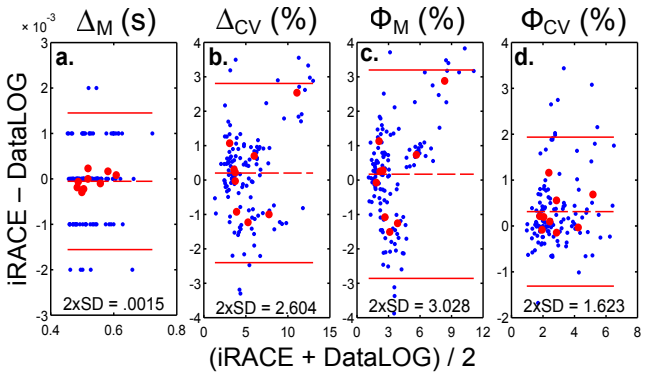

Figure 5: Bland–Altman step time analysis.

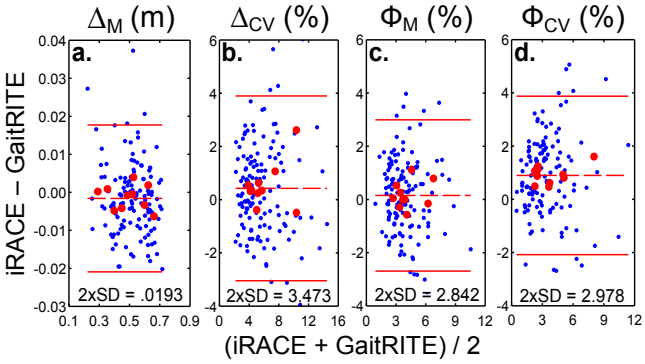

Figure 6: Bland–Altman step length analysis.

peaks and non-HS peaks) were identified. Based on 10-fold cross-validation, our HS detection algorithm (cf. Section 5.4) achieved 100% precision and recall. If the binary output was directly obtained from the classifier using a decision boundary of 0.5 (i.e., no further rule-based refinement was applied), the precision and recall were 99.98% and 99.87%, respectively (F_1 score = 99.93%).

On the same HS dataset, our left-vs.-right foot identification algorithm (cf. Section 5.6) also achieved 100% accuracy.

A step length dataset of 1668 steps was combined across all walks of all subjects with ground truth step lengths collected from GAITRite. The root mean square error (RMSE) of our step length calculation algorithm was 3.22 cm based on 10-fold cross-validation, or 6.95% when expressed as a CV value (i.e., RMSE divided by mean step length).

8.2 Outcome Measure Accuracy

A repeated-measures Bland–Altman (B–A) analysis [4] is presented for Step Time (Figure 5) and Step Length (Figure 6). As used here, a B–A analysis summarizes the accuracy of an experimental device (iRACE) against a control device (ground truth device; DataLOG or GAITRite). Outcome measures from individual trials (blue points) are averaged to create a single value per subject (red points). The key statistics in a B–A plot are the y-axis mean (dashed red line) and 2×SD error bars (solid red lines), which indicate the expected mismatch between the two devices at the population level (with the error term adjusted for multiple trials per subject, per [4]). Similar to the tapping simulation (cf. Figure 4c), Δ_M was very accurately estimated for both Step Time (Figure 5a) and Step Length (Figure 6a). Although the amount of y-axis error in iRACE-quantified variability measures (Δ_{CV}, Φ_M, Φ_{CV}) was greater than that of Δ_M, this error was less than ±1.0% in most subjects (i.e., averaged across individual walks). This indicates that iRACE would be able to distinguish parkinsonian gait from normal gait based on their percentage-based variability measures, which can differ by as much as 5% for stride time variability [9] and 2.5% for stride length variability [1].

8.3 Accuracy of RAC Facilitation

Finally, to quantify whether any effect of RAC facilitation was significantly different between iRACE and the relevant ground truth, a 2 (*Device*: iRACE vs. ground truth) × 3 (*RAC* conditions: Self-paced, 100%, 110%; cf. Figure 7) repeated-measures ANOVA was performed separately on each outcome measure. A significant effect for *RAC* was found for Step Time Δ_M ($F_{2,16} = 7.17$, $p = .006$), Step Time Δ_{CV} ($F_{2,16} = 4.05$, $p = .038$), and Step Length Δ_M ($F_{2,16} = 8.43$, $p = .003$), indicating the expected influence of RAC: decreasing inter-step durations and increasing inter-step lengths as tempo increased. Importantly, no significant *Device* × *RAC* interactions were present in any of the ANOVAs (each $p > .10$, as shown in Figure 7), indicating no systematic difference between iRACE and ground truth outcome measures in the ability to detect RAC facilitation when it actually occurred.

9. DISCUSSION AND CONCLUSION

The current paper presents the first-ever systematic ground truth validation of the accuracy of smartphone-based gait analysis in PD, and describes an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) mobile application to facilitate RAC efficacy diagnosis, via an assessment of upper motor (finger tapping) and lower motor (walking) perfor-

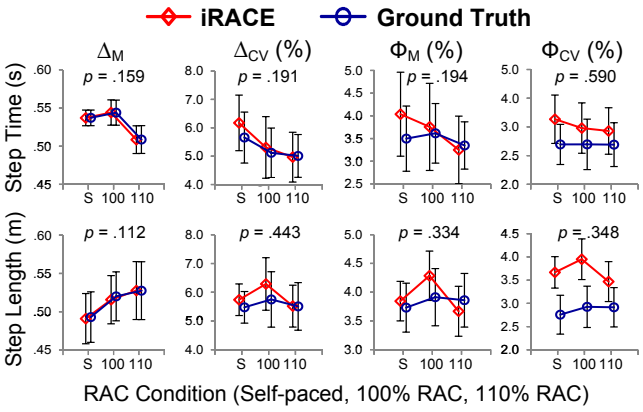

Figure 7: Outcome measures of step time and step length under 3 RAC conditions: self-paced ("S"), and cued with metronomes set to 100% ("100") and 110% ("110") of self-paced walking cadence.

mance. Novel machine learning-based gait analysis algorithms have been developed for iRACE, and the present pilot study reveals that iRACE has low error rates relative to clinic-standard ground truth measures, indicating a potentially useful clinical tool.

It is important to note that the present paper is concerned with evaluating the *accuracy* of smartphone-based gait measurement, not the *efficacy* of RAC (which has been the focus of prior studies [22, 23, 36, 38]). RAC has been widely noted for its facilitative effects on gait [12, 40]. iRACE enables medical professionals to quickly and accurately quantify the *degree* of RAC facilitation (cf. Section 8.3), enabling a decision of whether RAC might be useful as a "neurophysiological adjuvant" for individual PD patients.

The flexibility, portability, and validity of iRACE suggest both prognostic and analytic applications. First, because it is smartphone implemented, an entire world of RAC stimuli can be made available, ranging from the user's own music collection to commercially available streaming music (e.g., Deezer[5], Google Play[6], Spotify[7]), matched to the user's music preference and the requirements of RAC therapy (e.g., steady tempo and strong beats). Second, iRACE could be used to track the improvement of motor performance over time (e.g., in a longitudinal clinical trial) or in conjunction with a behavioral, pharmacological, or neurostimulatory intervention. Third, iRACE would enable clinicians to evaluate correlations between (1) upper motor and lower motor timing variability, and (2) step time and step length variability during gait, both of which remain largely unexplored in the clinical literature. Fourth, the portability of iRACE would enable it to be used by neurologists and physical therapists around the globe. The resultant surge in collected data (archived and managed on the iRACE website; cf. Section 3.2) could then be used to explore large-sample correlations with other disease-relevant variables, revealing potentially novel insights into how PD affects the motor system. Together, this future work may lead to an improved ability to characterize—and, ultimately, treat—motor dysfunction in Parkinson's disease.

[5]http://www.deezer.com
[6]https://play.google.com/music
[7]https://www.spotify.com

ACKNOWLEDGEMENTS

We thank Haotian "Sam" Fang, Hang Guo, Dr. Dawn Tan, Dr. Patsy Tan, Jacqueline Chen, Christal Chiang, and Lee Meng Ho for their help with clinical data collection, and Boyd Anderson for helpful comments. This project was partially funded by the National Research Foundation (NRF) 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), and partially funded by the Agency for Science, Technology and Research (A*STAR) under grant R-252-000-510-305.

10. REFERENCES

[1] P. Arias and J. Cudeiro. Effects of rhythmic sensory stimulation (auditory, visual) on gait in Parkinson's disease patients. *Exp. Brain Res.*, 186(4):589–601, 2008.

[2] B. Bilney, M. Morris, and K. Webster. Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait. *Gait & Posture*, 17(1):68–74, 2003.

[3] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[4] J. M. Bland and D. G. Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2):135–160, 1999.

[5] F. Bugané, M. Benedetti, et al. Estimation of spatial-temporal gait parameters in level walking based on a single accelerometer. *Comp Meth and Prog in Biomed*, 108(1):129–137, 2012.

[6] H. K. Chan, H. Zheng, et al. Feasibility study on iPhone accelerometer for gait detection. In *Proc. PervasiveHealth*, pages 184–187. IEEE, 2011.

[7] N. de Bruin, J. B. Doan, et al. Walking with music is a safe and viable tool for gait training in Parkinson's disease: the effect of a 13-week feasibility study on single and dual task walking. *Parkinson's Disease*, 2010, 2010.

[8] N. de Bruin, J. B. Doan, G. Turnbull, et al. Walking with music is a safe and viable tool for gait training in Parkinson's disease. *Parkinson's Disease*, 2010, 2010.

[9] M. F. del Olmo and J. Cudeiro. Temporal variability of gait in Parkinson disease: Effects of a rehabilitation programme based on rhythmic sound cues. *Parkinsonism & Related Disorders*, 11(1):25–33, 2005.

[10] H. T. Dinh, C. Lee, D. Niyato, and P. Wang. A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless Communications and Mobile Computing*, 13(18):1587–1611, 2013.

[11] E. Dorsey, R. Constantinescu, J. Thompson, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, 68(5):384–386, 2007.

[12] C. G. Goetz, S. Fahn, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Movement Disorders*, 22(1):41–47, 2007.

[13] Y. A. Grimbergen, M. Munneke, and B. R. Bloem. Falls in Parkinson's disease. *Current Opinion in Neurology*, 17(4):405–415, 2004.

[14] J. M. Hausdorff. Gait dynamics, fractals and falls: finding meaning in the stride-to-stride fluctuations of human walking. *Human Movement Science*, 26(4):555–589, 2007.

[15] J. M. Hausdorff. Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos: Interdiscipl J Nonlinear Sci*, 19(2):026113, 2009.

[16] J. M. Hausdorff, Z. Ladin, and J. Y. Wei. Footswitch system for measurement of the temporal parameters of gait. *J Biomechanics*, 28(3):347–351, 1995.

[17] J. H. Hollman, E. M. McDade, and R. C. Petersen. Normative spatiotemporal gait parameters in older adults. *Gait & Posture*, 34(1):111–118, 2011.

[18] T.-V. How, J. Chee, E. Wan, and A. Mihailidis. Mywalk: a mobile app for gait asymmetry rehabilitation in the community. In *Proc. PervasiveHealth*, pages 73–76. ICST, 2013.

[19] J. J. Kavanagh and H. B. Menz. Accelerometry: a technique for quantifying movement patterns during walking. *Gait & Posture*, 28(1):1–15, 2008.

[20] S. H. Keus, M. Munneke, et al. Physical therapy in Parkinson's disease: evolution and future challenges. *Movement Disorders*, 24(1):1–14, 2009.

[21] R. LeMoyne, T. Mastroianni, and W. Grundfest. Wireless accelerometer iPod application for quantifying gait characteristics. In *Proc. EMBC*, pages 7904–7907. IEEE, 2011.

[22] I. v. Lim, E. Van Wegen, et al. Effects of external rhythmical cueing on gait in patients with Parkinson's disease: a systematic review. *Clin Rehab*, 19(7):695–713, 2005.

[23] S. Lord, K. Baker, et al. Gait variability in Parkinson's disease: an indicator of non-dopaminergic contributors to gait dysfunction? *J Neurol*, 258(4):566–572, 2011.

[24] S. Lord, T. Howe, et al. Gait variability in older adults: a structured review of testing protocol and clinimetric properties. *Gait & Posture*, 34(4):443–450, 2011.

[25] M. E. McNeely, R. P. Duncan, and G. M. Earhart. Medication improves balance and complex gait performance in Parkinson disease. *Gait & Posture*, 36(1):144–148, 2012.

[26] S. Mellone, C. Tacconi, and L. Chiari. Validity of a smartphone-based instrumented Timed Up and Go. *Gait & Posture*, 36(1):163–165, 2012.

[27] H. B. Menz, M. D. Latt, et al. Reliability of the GAITRite walkway system for the quantification of temporo-spatial parameters of gait in young and older people. *Gait & Posture*, 20(1):20–25, 2004.

[28] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick. Acceleration patterns of the head and pelvis when walking are associated with risk of falling in community-dwelling older people. *J Gerontology Series A: Bio Sci and Med Sci*, 58(5):M446–M452, 2003.

[29] R. Moe-Nilssen. A new method for evaluating motor control in gait under real-life environmental conditions. part 1: The instrument. *Clin Biomech*, 13(4):320–327, 1998.

[30] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2):90–126, 2006.

[31] S. Nishiguchi, M. Yamada, et al. Reliability and validity of gait analysis by Android-based smartphone. *Telemedicine and e-Health*, 18(4):292–296, 2012.

[32] C. W. Olanow, Y. Agid, et al. Levodopa in the treatment of Parkinson's disease: current controversies. *Movement Disorders*, 19(9):997–1005, 2004.

[33] C. Pacchetti, F. Mancini, et al. Active music therapy in Parkinson's disease: an integrative method for motor and emotional rehabilitation. *Psychosomatic Medicine*, 62(3):386–393, 2000.

[34] L. Palmerini, S. Mellone, L. Rocchi, and L. Chiari. Dimensionality reduction for the quantitative evaluation of a smartphone-based Timed Up and Go test. In *Proc. EMBC*, pages 7179–7182. IEEE, 2011.

[35] M. Plotnik, N. Giladi, and J. M. Hausdorff. A new measure for quantifying the bilateral coordination of human gait. *Exp. Brain Res*, 181(4):561–570, 2007.

[36] T. C. Rubinstein, N. Giladi, and J. M. Hausdorff. The power of cueing to circumvent dopamine deficits: a review of physical therapy treatment of gait disturbances in Parkinson's disease. *Movement Disorders*, 17(6):1148–1160, 2002.

[37] Y. Song, S. Shin, S. Kim, D. Lee, and K. Lee. Speed estimation from a tri-axial accelerometer using neural networks. In *Proc. EMBC*, pages 3224–3227, Aug 2007.

[38] S. J. Spaulding, B. Barber, et al. Cueing and gait improvement among people with Parkinson's disease: A meta-analysis. *Archives Phys Med and Rehab*, 94(3):562–570, 2013.

[39] D. H. Sutherland. The evolution of clinical gait analysis part III–kinetics and energy assessment. *Gait & Posture*, 21(4):447–461, 2005.

[40] A. L. Taylor Tavares, G. S. Jefferis, 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. *Movement Disorders*, 20(10):1286–1298, 2005.

[41] K. E. Webster, J. E. Wittwer, and J. A. Feller. Validity of the GAITRite walkway system for the measurement of averaged and individual step parameters of gait. *Gait & Posture*, 22(4):317–321, 2005.

[42] M. Yamada, T. Aoyama, et al. Using a smartphone while walking: a measure of dual-tasking ability as a falls risk assessment tool. *Age and Ageing*, 40(4):516–519, 2011.

[43] M. Yang, H. Zheng, et al. Assessing the utility of smart mobile phones in gait pattern analysis. *Health and Technology*, 2(1):81–88, 2012.

[44] M. Yang, H. Zheng, et al. igait: An interactive accelerometer based gait analysis system. *Comp Meth and Prog in Biomed*, 108(2):715–723, 2012.

[45] S. Zhu, H. Anderson, and Y. Wang. A real-time on-chip algorithm for IMU-based gait measurement. In *Advances in Multimedia Information Processing–PCM 2012*, pages 93–104. Springer, 2012.

[46] W. Zijlstra. Assessment of spatio-temporal parameters during unconstrained walking. *European J Applied Physiology*, 92(1-2):39–44, 2004.

[47] W. Zijlstra and A. L. Hof. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait & Posture*, 18(2):1–10, 2003.