

Research Reports - Research Article

Multimodal Wearable Sensors to Measure Gait and Voice

Dimitrios Psaltos^a Kara Chappie^a Fikret Isik Karahanoglu^a
Rachel Chasse^a Charmaine Demanuele^a Amey Kelekar^a
Hao Zhang^a Vanessa Marquez^a Tairmae Kangarloo^a Shyamal Patel^a
Matthew Czech^a David Caouette^a Xuemei Cai^{a, b}

^aEarly Clinical Development, Pfizer Inc., Cambridge, MA, USA; ^bTufts Medical Center, Boston, MA, USA

Keywords

Wearable sensors · Gait · Speech · Voice · Movement disorders

Abstract

Background: Traditional measurement systems utilized in clinical trials are limited because they are episodic and thus cannot capture the day-to-day fluctuations and longitudinal changes that frequently affect patients across different therapeutic areas. **Objectives:** The aim of this study was to collect and evaluate data from multiple devices, including wearable sensors, and compare them to standard lab-based instruments across multiple domains of daily tasks. **Methods:** Healthy volunteers aged 18–65 years were recruited for a 1-h study to collect and assess data from wearable sensors. They performed walking tasks on a gait mat while instrumented with a watch, phone, and sensor insoles as well as several speech tasks on multiple recording devices. **Results:** Step count and temporal gait metrics derived from a single lumbar accelerometer are highly precise; spatial gait metrics are consistently 20% shorter than gait mat measurements. The insole's algorithm only captures about 72% of steps but does have precision in measuring temporal gait metrics. Mobile device voice recordings provide similar results to traditional recorders for average signal pitch and sufficient signal-to-noise ratio for analysis when hand-held. Lossless compression techniques are advised for signal processing. **Conclusions:** Gait metrics from a single lumbar accelerometer sensor are in reasonable concordance with standard measurements, with some variation between devices and across individual metrics. Finally, participants in this study were familiar with mobile devices and had high acceptance of potential future continuous wear for clinical trials.

© 2019 The Author(s)
Published by S. Karger AG, Basel

Xuemei Cai, MD
Early Clinical Development, Pfizer Inc.
610 Main Street
Cambridge, MA 02139 (USA)
E-Mail xuemei.cai@pfizer.com

Introduction

Traditional clinical care and research is often restricted to patient and caregiver self-report and assessments completed in clinical visits. Due to the episodic nature of these visits, they may only capture a small proportion of objective measurements of continuous disease states. Technological advances in wearable sensors and remote monitoring may offer a more advantageous solution to the current episodic clinical paradigm and increase access to care for patients in remote locations [1]. Continuous sensor recording of patients over multiple days could yield objective, real-world data that may not be fully captured currently by patient or caregiver recall [2].

Wearable inertial sensors have been explored for managing and monitoring diseases, including sarcopenia, multiple sclerosis, Duchenne muscular dystrophy, and Parkinson's disease, in order to reduce clinical visits [3, 4]. Furthermore, smartphone-derived metrics have been correlated with clinical endpoints in posttraumatic stress disorder, depression, and cachexia [5]. Similarly, the multi-sensor components of smartphones, including accelerometer, gyroscope, and barometers, have already been utilized for monitoring physical movement [6]. Changes in voice features – which have fine motor components [7] – have been associated with a number of medical conditions, including Parkinson's disease [8], attention deficit/hyperactivity disorder [9], bipolar disorder [10], and functional dysphonia [11], offering additional remote monitoring opportunities to evaluate patient symptoms [12]. Here, we assess several consumer-grade wearable systems comprised of a smartwatch and smartphone worn in the lumbar position as well as pressure insoles against traditional gold-standard assessments in the performance of scripted movement and speech tasks. We compare assessments between metrics on different devices as well as evaluate wearability.

Materials and Methods

Experimental Protocol

Forty healthy participants aged 18–65 years (age: 34.8 ± 10.2 [20–58], gender: 21 female, 19 male; level of education: 11 college graduates, 27 postgraduates, 1 high school graduate, and 1 did not respond) performed a battery of scripted tasks while wearing inertial sensors during a single 60-min in-laboratory assessment. Demographics, including age, sex, weight, height, handedness, body mass index, and education level, were collected (online suppl. Fig. 1, see www.karger.com/doi/10.1159/000503282 for all online suppl. material). Each participant underwent assessment with the Edinburgh Handedness Scale (EHI) short form (3 patients were left handers, 34 right handers, and 3 ambidextrous [13, 14]) and completed the Mobile Device Proficiency Questionnaire (MDPQ) (score, 75.63 ± 4.69 [64–80]) [15]. Participants wore 1 Apple Watch Series 3™ (Watch iOS 4.3) on each wrist, 2 iPhonesSM 8 Plus (iOS 11) placed in an elastic sports band at approximately the 4th lumbar vertebrae level, and Moticon pressure insoles [9, 16] in both shoes and performed a 14-m walk (with 1 turn) captured on a 7-m GaitRite[®] gait mat [8] (online suppl. Fig. 2). The phone and watch ran a Sensor Data Capture (SDCapture) application (version 3.0.7) (Fig. 1). This application was designed to extract the raw inertial sensor data from the Apple Watch and iPhone. Participants repeated the walk without insoles while still wearing the watches and phones.

Participants were then given the iPhone and a Zoom mic H2n Field Recorder and were asked to record several voice tasks, including sustaining vowels /e/, /i/, and /o/, while seated. At the end of the visit, they completed a wearability questionnaire.

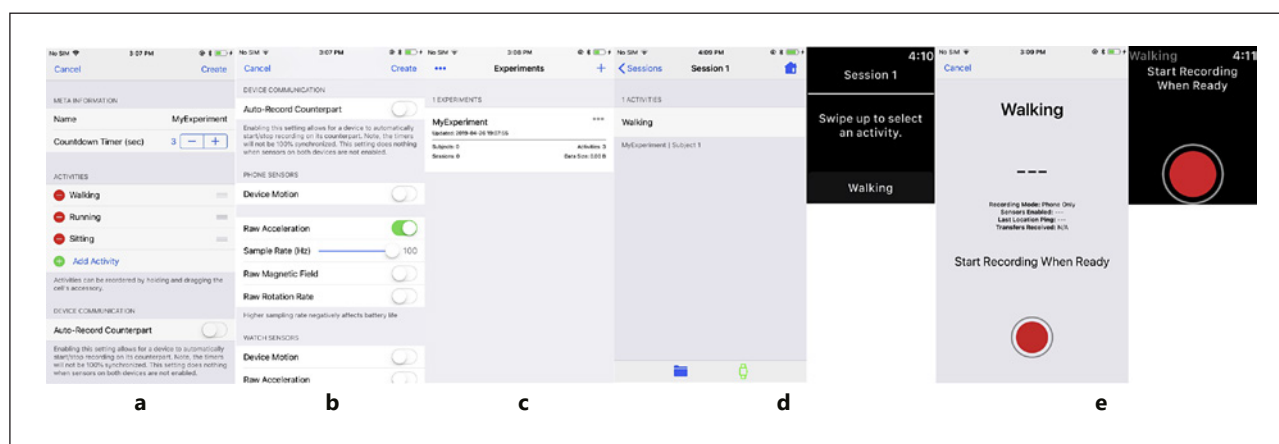


Fig. 1. Sensor Data Capture or “SDCapture” is an application that can be used to gain access to device sensors: Raw Accelerometer, Processed Device Motion, Raw Gyroscope (phone only), and Raw Magnetic Field (phone only) of the iPhone and Apple Watch. The app organizes data into “experiments” to allow for file export. Screen shots of the iPhone and Apple Watch screens show how these activities were selected during the study. **a** Create an experiment with a name, countdown, set of activities, sensors, and auto-recording. **b** Enable specific sensors and configure sampling rates. **c** After the experiment is created, the user is brought back to the list. **d** Send the session to the watch and choose an activity. The activity chosen must be the same on both the watch and phone. **e** Start and stop recording on each device. If auto-recording is enabled, pressing the recording controls on one device will also fire the same action on the other automatically.

Gait Analysis

Gait was analyzed by isolating individual gait task data segments, using task start and end times from the gait mat. GaitRite® Software (version 4.8.5) [8] and Moticon SCIENCE Software (version 01.11.00) [9, 16] provided gait metrics for each recognized footfall, including step time, step length, and stride velocity. Stride, swing, and stance time and double support were derived [17] (online suppl. Fig. 3).

Spatial and temporal gait characteristics were extracted from the iPhone’s single lumbar mounted tri-axial accelerometer using a wavelet-based inverted pendulum model algorithm [18]. Our algorithm was designed to utilize only the accelerometer, unlike other algorithms which also utilize a gyroscope [19], in order to reduce sensor size and preserve battery life. Using this method, heel strike and toe off events were determined for each step by Gaussian continuous wavelet-based transformations of the raw vertical accelerometer signal. Extraneous event detections outside physiologically realistic boundaries were filtered out via an optimization procedure. Temporal characteristics, including step, stride, swing, and stance times, were calculated from the remaining detected steps. Sensor height estimates and vertical displacement of the center of mass, extracted from vertical acceleration data, established spatial characteristics of gait [20]. The median of all metrics was computed for the walk task. The first 10 participants experienced technical data capture issues with the SDCapture app’s ability to communicate with the location services of the iPhone. Thirty participants had lumbar accelerometer data that was used in the final analyses.

The insole software provided individual footfall data [8] and temporal gait parameters for each insole. The left and right gait data were combined to find the median for the task. Temporal and spatial gait metrics were calculated by GaitRite® software for each gait mat-recognized step. The median of all metrics was taken for each individual gait task. These served as the gold standard comparator to the insoles and lumbar accelerometer results.

Voice Analysis

Voice audio files that were simultaneously recorded in a laboratory setting using the Voice Memo app of the iPhone with a lossy compressed audio quality setting, 48-kHz sampling rate, and variable bit rate, and recordings from the Zoom recorder, using a lossy compression technique with a 44.1-kHz sampling rate and 160 kbits/s bit rate, were compared in quality. Understanding the recording quality of smart phones is important to understanding their practical utility in capturing vocal changes in medical applications. To quantify any differences between the designated audio recording device and the multi-purpose smartphone, signal-to-noise ratio (SNR) was examined first. Vocal features (harmonicity and pitch) that have been used to quantify recording quality in past studies [21], and which have been studied with disease conditions [22], were then extracted from the co-recordings using the phonetics program Praat [13] and compared.

Thirty-nine participants (1 participant was removed due to recording failure), sustaining vowels /e/, /i/, and /o/ for as long as possible, were analyzed using Praat. The 3 vowel sounds were manually segmented in the time domain and programmatically trimmed 0.5 s on each end, so that the steady state portion of the vowel was analyzed. One noise sample for each participant was manually extracted at the completion of the 3 vowel tasks. SNR was computed using the average signal intensity level and the median noise level for each vowel, where the median noise level was determined from the 10th and 90th intensity percentiles of the noise sample. An adjustment to the iPhone's signal intensity was made using an inverse square loss model as the iPhone and Zoom Recorder were approximately 0.61 m apart.

Statistical Analysis

Statistical analysis was performed in R version 3.4.1 [23]. Shapiro-Wilk test was used to assess normality of the data, and parametric (such as paired *t* test) or nonparametric (such as Wilcoxon rank sum tests) statistics were used accordingly. All statistical tests were performed using Stats R package, unless stated otherwise.

Gait Analysis

For each gait task and for each algorithm used to investigate gait, the median value of gait metrics across all steps was computed and used in the analyses detailed below.

Bland-Altman plots with 95% lower and upper limits of agreement (mean difference \pm 1.96 standard deviation of differences) [24] and Pearson's correlation analysis were used to visualize and quantify agreement between gait features derived using the gait mat's algorithm and those derived from the lumbar sensor and the insoles. Bland-Altman analyses were computed using BlandAltmanLeh R package.

In order to determine whether wearing insoles affected the participants' gait, paired Wilcoxon rank sum tests were used to compare derived gait features from the gait mat, with and without insoles.

Voice Analysis

Bland-Altman plots with 95% lower and upper limits of agreement, Pearson's correlation analysis, and paired *t* tests were used to visualize and quantify agreement in pitch, harmonicity, and SNR of each vowel derived from the Zoom and iPhone recordings. Bland-Altman analyses were computed using BlandAltmanLeh R package.

Questionnaires

Questionnaire responses were summarized using descriptive statistics (number of participants, mean, median, standard deviation, minimum and maximum) for continuous (or near continuous) variables, and frequency and percentages for discrete variables.

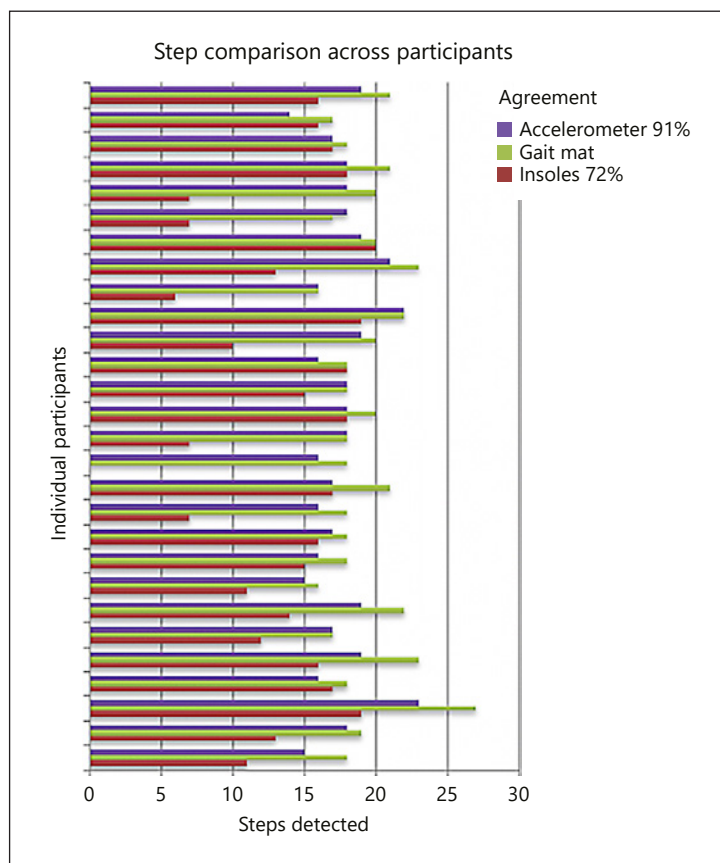


Fig. 2. Footfalls detected by gait mat across individual participants compared to footfalls detected by lumbar accelerometer and insoles.

The relationship between age and the total Mobile Device Proficiency Questionnaire (MDPQ) score was defined using Pearson's correlation analysis. Variations of the total MDPQ score with gender and with level of education were assessed separately using Wilcoxon rank sum test. Note that since the majority of participants were either college or postgraduate educated, level of education was treated as a factor with 2 levels.

We fit a linear regression model to the total MDPQ score using age, gender, and level of education and their interaction as main effects using the MASS R package. ANOVA findings were reported as *F* values and their corresponding *p* values using Type III sum of squares (statistics were derived using the car R package). Fisher's exact tests were used to quantify the relationship between wearability questionnaire responses and gender. Post hoc tests were used to investigate pairwise comparisons following significant findings using the fisher.multcomp function (RVAideMemoire R package). Kruskal-Wallis rank sum test was used to investigate the relationship between questionnaire responses and age.

Results

Gait Analysis

Accelerometer and Insoles versus Gait Mat

To inspect the ability of lumbar accelerometer and insole sensors to recognize gait, the numbers of steps recognized by each device were compared to those on the gait mat. Fewer steps were recognized by the wearable sensors compared to the gait mat. The lumbar accelerometer had a 91% agreement compared to 72% for insoles (Fig. 2).

Gait metric	Gait mat: accel						Gait mat: Insoles					
	Mean bias (Gait mat – accel)	Normalized bias, %	95% LLOA	95% ULOA	Pearson's <i>r</i>	<i>p</i> value (\leq)	Mean bias (Gait mat – insoles)	Normalized bias, %	95% LLOA	95% ULOA	Pearson's <i>r</i>	<i>p</i> value (\leq)
Stance time, s	0.01	1.46	−0.04	0.06	0.98	1E-15	−0.03	−3.65	−0.07	0.02	0.96	1E-12
Swing time, s	−0.03	−6.47	−0.06	0.01	0.91	1E-08	−0.01	−1.81	−0.09	0.07	0.61	1E-02
Step time, s	0.00	−0.79	−0.02	0.01	0.99	1E-15	−0.02	−3.99	−0.09	0.05	0.83	1E-05
Double support, s	0.06	14.83	−0.03	0.14	0.85	1E-06	0.17	47.80	0.07	0.26	0.61	1E-02
Step length, m	0.13	20.40	0.03	0.23	0.68	1E-03						
Stride velocity, m/s	0.23	20.67	0.04	0.42	0.79	1E-05						
Stride time, s	−0.01	−1.11	−0.04	0.02	0.99	1E-15						
Stride length, m	0.26	20.04	0.07	0.45	0.70	1E-03						

Fig. 3. Bland-Altman analysis of the upper/lower 95% limits of agreement (ULOA/LLOA), mean bias and normalized bias of lumbar accelerometer (accel) versus gait mat, and insoles versus gait mat, and Pearson correlation for the temporal and spatial gait metrics calculated from lumbar accelerometer and insoles as compared to gait mat.

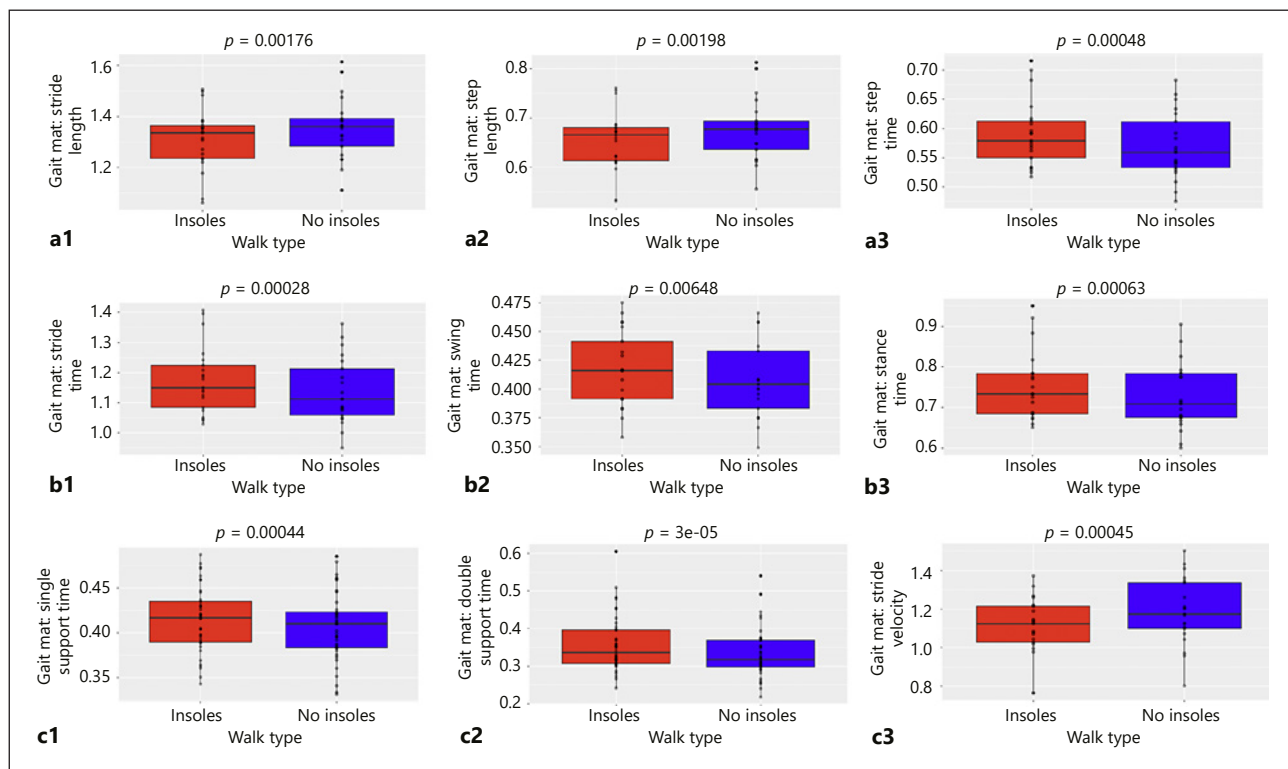


Fig. 4. Wilcoxon paired *t* tests comparing walking on gait mat with and without insoles.

Temporal Gait Parameters

The 4 overlapping temporal gait metrics obtained from both the insoles and lumbar accelerometer were stance time, swing time, step time, and double support time. The lumbar accelerometer and insoles' computed stance and step times were highly correlated with those of the gait mat (Pearson's $r > 80\%$, $p < 1.00e-5$) and had minimal normalized bias ($<5\%$) (Fig. 3) (online suppl. Fig. 4). The lumbar accelerometer-computed swing and double support times were also highly correlated (Pearson's $r > 85\%$, $p < 1.00e-5$), although the double support time had a high level of normalized bias ($\sim 15\%$). The insoles were less corre-

	Zoom: iPhone					
	Mean Bias (Zoom – iPhone)	Normalized Bias, % Zoom	95% LLoA	95% ULoA	Pearson's <i>r</i>	<i>p</i> value (\leq)
Average Pitch, Hz Vowel /e/	4.11	2.57	–30.82	39.04	0.92	1E-15
Average Pitch, Hz Vowel /i/	0.49	–0.10	–26.47	27.43	0.95	1E-15
Average Pitch, Hz Vowel /o/	6.07	3.20	–36.75	48.90	0.88	1E-12
Average Harmonicity, dB Vowel /e/	–1.37	–11.52	–7.34	4.61	0.62	1E-4
Average Harmonicity, dB Vowel /i/	–2.96	–22.60	–10.43	4.51	0.46	1E-2
Average Harmonicity, dB Vowel /o/	–1.53	–11.02	–8.27	5.21	0.7	1E-6
SNR, dB Vowel /e/	7.32	37.72	–1.87	16.50	0.58	1E-4
SNR, dB Vowel /i/	6.43	37.19	–2.97	15.84	0.51	1E-3
SNR, dB Vowel /o/	7.36	36.47	–1.69	16.41	0.67	1E-5

Fig. 5. Pearson correlation and Bland-Altman analysis of the upper/lower 95% limits of agreement (ULoA/LLoA), mean bias and normalized bias between Zoom and iPhone recorders.

lated in calculating the swing time and double support time (Pearson's $r = 0.61$, $p < 0.01$) and produced a double support time metric which had a high normalized bias ($\sim 48\%$).

Parameters not provided by the insoles included stride time, step length, stride length, and stride velocity. The lumbar accelerometer method had a highly correlated stride time (Pearson's $r = 0.99$, $p < 1.00\text{e-}15$) with little normalized bias ($\sim 1\%$) (Fig. 3) (online suppl. Fig. 4). Stride velocity was also correlated with that of the gait mat (Pearson's $r = 0.79$, $p < 1.00\text{e-}5$) with a higher normalized bias ($\sim 21\%$). The lumbar accelerometer was less correlated when calculating the spatial parameters of stride length and step length (Pearson's $r < 0.71\%$, $p < 1.00\text{e-}3$) and both had a higher normalized bias ($\sim 20\%$).

Effect of Wearing Insoles

The effect of wearing the insoles was investigated using gait mat metrics under wear/non-wear conditions. Gait mat-calculated metrics were analyzed using a Wilcoxon signed-rank test to inspect any changes due to insoles (Fig. 4). All gait metrics showed a significant change ($p < 0.05$) after removing insoles. After removing insoles, spatial gait metrics and stride velocity increased, while temporal gait metrics decreased.

Speech Analysis

SNR is widely used as a measure of signal quality. Knowing whether the iPhone and Zoom Recorder record with equivalent SNRs is of interest. Because vowels are periodic signals, harmonicity, which measures the proportion of energy in the periodic component of a signal to the energy in the nonperiodic component, offers a direct, single-step measure of SNR, assuming there are no periodic components in the noise. The resulting paired *t* test of equiv-

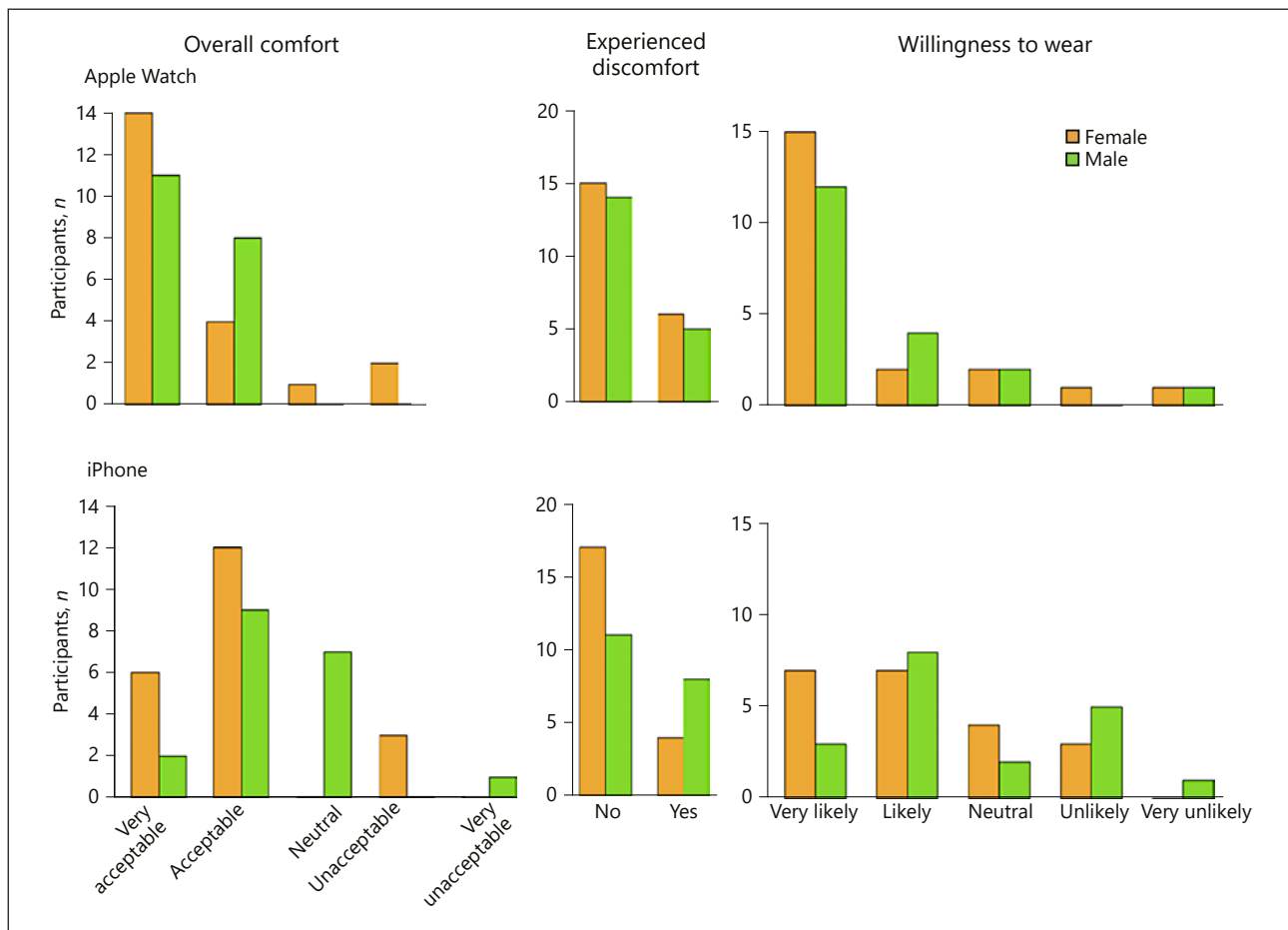


Fig. 6. Wearability results.

alent device harmonicity measure for the vowels /e/, /i/, and /o/ provided $p = 9.00\text{e-}3$, $p < 1.00\text{e-}4$, and $p < 0.01$, respectively. Bland-Altman analyses additionally indicated mean harmonicity bias between the devices for the respective 3 vowels, where the iPhone had a consistently larger average harmonicity (Fig. 5) (online suppl. Fig. 5). Pearson's correlation analysis revealed a statistically significant, moderately linear relationship between device measurements for /e/, /i/, and /o/. Note that this calculation did not yet co-locate the devices, and it was physically reasonable that the averaged iPhone harmonicity always exceeded the averaged Zoom harmonicity for all vowels as the iPhone microphone was closer to the subject during recording time.

Using harmonicity to describe SNR is valid providing the noise associated with the signal of interest is nonperiodic. Upon examining the noise signals associated with the iPhone and Zoom recordings, differing amounts of periodic instances ($p < 1.00\text{e-}4$) were observed. For this reason, SNR was computed traditionally, using signal and noise intensity levels. Results showed the iPhone SNR to be larger than the Zoom SNR, consistent with the harmonicity result. However, when the iPhone was co-located with the Zoom device, computed SNR levels indicated mean bias that the iPhone SNR was always less than the Zoom SNR (Fig. 5) (online suppl. Fig. 5).

Pitch is a measure of the perceived frequency of speech. Because a voice recorded on different devices generally sounds the same, it is expected that simultaneously recorded

sound samples analyzed from the iPhone and the Zoom Recorder will have the same pitch measurements. To test this hypothesis, an average pitch measurement was obtained for each vowel for both the iPhone and the Zoom Recorder with results shown in Figure 5.

Harmonicity results were obtained with the iPhone held closer to the participant and with Zoom in a fixed location on the table. The SNR result was with the iPhone and Zoom co-located at the same distance.

Wearability and MDPQ Analyses

The Total MDPQ score for participants included a minimum of 64, a median of 78, a mean of 75.63, and a maximum of 80. There was no relationship between MDPQ scores and age (Pearson's $r = -0.13$, $p = 0.45$), gender ($p = 0.10$), nor level of education ($p = 0.87$).

The linear regression model for MDPQ score revealed a significant main effect of education level ($p = 0.04$) and age ($p = 0.02$) and a significant interaction between age and education level ($F = 5.39$, $p = 0.03$), indicating that the relationship between total score and age differed by the level of education (a slope of -0.62 for college educated and a slope of -0.03 for post-graduates).

Participants were asked to rank the overall comfort of the watch and phone. Both were generally rated as “very acceptable” or “acceptable.” There was no significant difference in the watch overall comfort levels between women and men as most reported no discomfort (Fisher's exact test, $p = 1.0$). Participants were asked whether they were “very likely, likely, neutral, unlikely, or very unlikely” to wear the watch continuously at home over multiple days. Fifteen women and 12 men responded that they were “very likely” to wear the watch. There was no significant difference between genders ($p = 0.83$) (Fig. 6). There were some comfort issues raised for the 2 phones worn in an elastic pouch at the waist. Seven men reported the comfort as “neutral” and 3 women reported the comfort as “unacceptable” (post hoc tests, $p = 0.03$). In the remaining responses, 29 participants listed the phone in the elastic waist band to be either “very acceptable” or “acceptable.” When participants were asked if they were “very likely, likely, neutral, unlikely, or very unlikely” to wear the phone around their waist continuously at home for multiple days, 25 participants said they were “likely” or “very likely” to do so with 9 responding either “unlikely” or “very unlikely.”

None of the wearability questionnaire responses were found to vary significantly with age (Kruskal-Wallis rank sum test, $0.12 < p < 0.64$).

Discussion/Conclusion

Gait

We investigated several spatial and temporal gait metrics derived from a single lumbar accelerometer and insoles. The lumbar accelerometer was precise in calculating steps, recognizing 91% of the steps recognized from the gait mat. Temporal metric estimations (step time, stance time, and stride time) were highly precise, within 1–2% bias, except the double support (15%). Spatial metrics were consistently shorter than the gait mat results, suggesting that while we can confidently rely on the temporal metrics derived from accelerometer (except double support), spatial metrics are less reliable, and the underestimation should be noted when using this method. The insoles' algorithm identified approximately 72% of the steps recognized by the gait mat but was relatively precise in temporal metrics (2–4%), except double support time (48%).

Underestimation of gait metrics from the accelerometer compared to the gait mat could be due to multiple factors. The accelerometer-based gait parameters were calculated using an inverted pendulum model, which utilized the pendulum length (sensor height from the

ground) and change in height of the center of mass in order to derive spatial metrics of gait. Spatial estimation relies on the degree to which the model represents human walking, and discrepancies could have resulted from errors in estimating key parameters, such as the sensor height (l) and the sensor's vertical change in height (h). A possible source of error comes from the use of an elastic sports band to instrument the 2 iPhones in the lumbar position. The weight of the sensors and elasticity of the band may increase the variability of l and h parameters. In contrast, the gait mat calculated gait spatial metrics directly from foot-falls on the mat.

To compare metrics derived from devices, the effect of insoles on gait was examined. Figure 4 shows the Wilcoxon tests comparing gait metrics with and without insoles. All metrics had significant change after removing insoles ($p < 0.05$). Figure 4 (a1, a2, a3, and c3) shows that stride length, step length, and stride velocity increased while step time decreased.

Speech

Due to periodic noise differences in the device recordings, SNR based on signal and noise intensity levels was computed, showing that the iPhone SNR is always less than the Zoom SNR when devices are co-located. However, the iPhone is found to have sufficient SNR for signal processing when participants hold the phone in speaker mode.

Harmonicity evaluation showed that the iPhone and Zoom Recorder noise signal contained differing amounts of periodic instances, with more instances and more averaged periodic noise energy in the iPhone. Because noise is omnidirectional, observed differences in periodic noise indicated that the 2 devices either handle environmental noise differently or have different internal noise profiles.

Periodic vowel phonemes recorded simultaneously on the iPhone and Zoom Recorder measured pitch equivalently in regions of constant pitch. Because both devices recorded using lossy compression with different settings, there was no definitive way to know which device may have captured truer voice dynamics when pitch variations occurred. For the higher-frequency vowels /e/ and /o/, the iPhone had a greater average percentage of nonaligned pitch segments (18.9 and 7.01%, respectively) compared to total signal duration affecting the average pitch difference between the 2 devices. For the lower-frequency vowel /i/, the Zoom had the greater average with 3.5%. Because structural voice changes, such as pitch variations, can be indicative of disease [25], it is recommended that nonlossy compression techniques be used when the voice signal is analyzed for detecting and monitoring disease.

Limitations and Future Considerations

Limitations of this device comparison study include factors associated with subject backgrounds, number of devices used, and executed voice exercises. Because the tested population was highly educated based on self-reported education with high MDPQ scores, feedback regarding wearability and willingness to participate in future digital medicine studies may not be generalizable. Additionally, the number of devices tested in this study was limited, and follow-on work to be published includes additional devices with test-retest comparisons in at-home and in-clinic settings, and task randomization. Lastly, the highly studied vowel, /a/, was not included in the phoneme suite. A follow-on study to be published in the future includes this phoneme.

In addition to the research opportunities noted with the limitations, future work includes investigating a different size device for the lumbar sensor, as well as additional algorithm explorations for gait and evaluations of additional voice features. Because some study participants had negative feedback regarding wearing the plus model of the iPhone ($6.24 \times 3.07 \times 0.30$ in) weighing 202 g in the lumbar position in an elastic pouch, it may be prudent in future studies to consider using a smaller sensor at the lumbar position and giving participants

larger-screen devices to perform patient-reported outcome questionnaires or cognitive tasks. Because the insoles' algorithm missed a significant portion of steps but generated precise temporal gait metrics, future work includes analyses of data from raw pressure readings in insoles, which has shown an improvement in step recognition. Future work will also include processing the watch data from the wrist, with the goal of further improving the lumbar accelerometer gait detection algorithm. Lastly, additional comparative analyses for voice features, including jitter, shimmer, voice onset time, pause durations, and Mel Frequency Cepstral Coefficients, are intended.

In conclusion, across the highly educated healthy volunteer population, participants expressed high wearability for mobile devices. Mobile device voice recordings provided similar results to traditional recorders for average signal pitch, and sufficient SNR for analysis when hand-held. Mobile device recordings also showed strong agreement in calculating gait metrics compared to standard clinical measures.

Acknowledgement

The authors would like to thank Andrew Messere of Pfizer for his help during the conduct of the study for administrative support and project management.

Statement of Ethics

This research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki. Participants have given their written informed consent. The study protocol was approved by an independent institutional review board.

Disclosure Statement

The authors are employees of Pfizer, Inc., except for K.C., an employee of Atrium Staffing who was a paid contractor to Pfizer in the development of this manuscript and statistical analyses, and T.K., a former Pfizer employee.

Funding Sources

This study was sponsored by Pfizer, Inc.

Author Contributions

D.P. participated in the organization and execution of this study as well as manuscript preparation and review. K.C. participated in the statistical analysis execution, manuscript writing and preparation. F.I.K. participated in the statistical analysis execution, manuscript writing and preparation. R.C. participated in the organization and execution of this study as well as manuscript preparation and review. C.D. participated in the research project conception, statistical analysis design and execution, and manuscript writing. A.K. participated in the research project organization and execution, manuscript writing and preparation. H.Z. participated in the research project concept, organization and execution, manuscript writing and preparation. V.M.

participated in the research project organization and manuscript writing. T.K. participated in the research project conception, organization, and execution. S.P. participated in the statistical design and analysis and manuscript preparation. M.C. participated in the statistical analysis execution and manuscript preparation. D.C. participated in the research project conception, organization, and manuscript preparation. X.C. participated in the conception, organization, and execution of this study as well as manuscript preparation and review.

References

- 1 Dorsey ER, Glidden AM, Holloway MR, Birbeck GL, Schwamm LH. Teleneurology and mobile technologies: the future of neurological care. *Nat Rev Neurol*. 2018 May;14(5):285–97.
- 2 Izmailova ES, Wagner JA, Perakslis ED. Wearable Devices in Clinical Trials: hype and Hypothesis. *Clin Pharmacol Ther*. 2018 Jul;104(1):42–52.
- 3 Mirelman A, Bonato P, Camicioli R, Ellis TD, Giladi N, Hamilton JL, et al. Gait impairments in Parkinson's disease. *Lancet Neurol*. 2019 Jul;18(7):697–708.
- 4 Heldman DA, Harris DA, Felong T, Andrzejewski KL, Dorsey ER, Giuffrida JP, et al. Telehealth Management of Parkinson's Disease Using Wearable Sensors: An Exploratory Study. *Digit Biomark*. 2017 Sep;1(1):43–51.
- 5 Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders. *J Med Internet Res*. 2017 Mar;19(3):e75.
- 6 del Rosario MB, Redmond SJ, Lovell NH. Tracking the Evolution of Smartphone Sensing for Monitoring Human Movement. *Sensors (Basel)*. 2015 Jul;15(8):18901–33.
- 7 Zhang Z. Mechanics of human voice production and control. *J Acoust Soc Am*. 2016 Oct;140(4):2614–35.
- 8 Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J Acoust Soc Am*. 2011 Jan;129(1):350–67.
- 9 Garcia-Real T, Diaz-Roman TM, Garcia-Martinez V, Vieiro-Iglesias P. Clinical and acoustic vocal profile in children with attention deficit hyperactivity disorder. *J Voice*. 2013 Nov;27(6):787.e11–8.
- 10 Karam ZN, Provost EM, Singh S, Montgomery J, Archer C, Harrington G, et al. Ecologically Valid Long-Term Mood Monitoring of Individuals with Bipolar Disorder Using Speech. *Proc IEEE Int Conf Acoust Speech Signal Process*. 2014 May;2014:4858–62.
- 11 Roy N, Leeper HA. Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: perceptual and acoustic measures. *J Voice*. 1993 Sep;7(3):242–9.
- 12 Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng*. 2010 Apr;57(4):884–93.
- 13 Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971 Mar;9(1):97–113.
- 14 Veale JF. Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality*. 2014;19(2):164–77.
- 15 Roque NA, Boot WR. A New Tool for Assessing Mobile Device Proficiency in Older Adults: The Mobile Device Proficiency Questionnaire. *J Appl Gerontol*. 2018 Feb;37(2):131–56.
- 16 Braun BJ, Veith NT, Hell R, Döbele S, Roland M, Rollmann M, et al. Validation and reliability testing of a new, fully integrated gait analysis insole. *J Foot Ankle Res*. 2015 Sep;8(1):54.
- 17 Hollman JH, McDade EM, Petersen RC. Normative spatiotemporal gait parameters in older adults. *Gait Posture*. 2011 May;34(1):111–8.
- 18 Del Din S, Godfrey A, Galna B, Lord S, Rochester L. Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length. *J Neuroeng Rehabil*. 2016 May;13(1):46.
- 19 Trojaniello D, Ravaschio A, Hausdorff JM, Cereatti A. Comparative assessment of different methods for the estimation of gait temporal parameters using a single inertial sensor: application to elderly, post-stroke, Parkinson's disease and Huntington's disease subjects. *Gait Posture*. 2015 Sep;42(3):310–6.
- 20 Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. *Gait Posture*. 2003 Oct;18(2):1–10.
- 21 Uloza V, Padervinskis E, Vegiene A, Pribuisiene R, Saferis V, Vaiciukynas E, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *Eur Arch Otorhinolaryngol*. 2015 Nov;272(11):3391–9.
- 22 Vogel AP, Maruff P. Comparison of voice acquisition methodologies in speech research. *Behav Res Methods*. 2008 Nov;40(4):982–7.
- 23 R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
- 24 Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015 Jun;25(2):141–51.
- 25 Fabbri M, Guimarães I, Cardoso R, Coelho M, Guedes LC, Rosa MM, et al. Speech and Voice Response to a Levodopa Challenge in Late-Stage Parkinson's Disease. *Front Neurol*. 2017 Aug;8:432.