

# Comparison of Acoustic Voice Features Derived From Mobile Devices and Studio Microphone Recordings

\*<sup>†</sup>Vitória S. Fahed, \*<sup>†</sup>Emer P. Doheny, \*<sup>†</sup>Monica Busse, §,<sup>||</sup>Jennifer Hoblyn, and \*<sup>†</sup>Madeleine M. Lowery,

\*<sup>†</sup>§<sup>||</sup>Dublin, Ireland, and ‡Cardiff, UK

**Summary: Objectives/Hypothesis.** Improvements in mobile device technology offer new opportunities for remote monitoring of voice for home and clinical assessment. However, there is a need to establish equivalence between features derived from signals recorded from mobile devices and gold standard microphone-preamplifiers. In this study acoustic voice features from android smartphone, tablet, and microphone-preamplifier recordings were compared.

**Methods.** Data were recorded from 37 volunteers (20 female) with no history of speech disorder and six volunteers with Huntington's disease (HD) during sustained vowel (SV) phonation, reading passage (RP), and five syllable repetition (SR) tasks. The following features were estimated: fundamental frequency median and standard deviation (F0 and SD F0), harmonics-to-noise ratio (HNR), local jitter, relative average perturbation of jitter (RAP), five-point period perturbation quotient (PPQ5), difference of differences of amplitude and periods (DDA and DDP), shimmer, and amplitude perturbation quotients (APQ3, APQ5, and APQ11).

**Results.** Bland-Altman analysis revealed good agreement between microphone and mobile devices for fundamental frequency, jitter, RAP, PPQ5, and DDP during all tasks and a bias for HNR, shimmer and its variants (APQ3, APQ5, APQ11, and DDA). Significant differences were observed between devices for HNR, shimmer, and its variants for all tasks. High correlation was observed between devices for all features, except SD F0 for RP. Similar results were observed in the HD group for SV and SR task. Biological sex had a significant effect on F0 and HNR during all tests, and for jitter, RAP, PPQ5, DDP, and shimmer for RP and SR. No significant effect of age was observed.

**Conclusions.** Mobile devices provided good agreement with state of the art, high-quality microphones during structured speech tasks for features derived from frequency components of the audio recordings. Caution should be taken when estimating HNR, shimmer and its variants from recordings made with mobile devices.

**Key Words:** Mobile devices—Acoustic voice features—Huntington's disease—Microphone.

## INTRODUCTION

Mobile devices are widely available, with the majority of the global population having a least one device, such as a smartphone or tablet. Portable, easy-to-use technology has been making its way into health research in recent years, though the use of notifications reminders,<sup>1</sup> passively monitoring health,<sup>2–4</sup> or actively engaging with participants.<sup>5–7</sup> The recent COVID-19 pandemic has highlighted the need to further harness the potential of mobile devices, to enable research and healthcare monitoring to be conducted from a distance when necessary. Although there is still some

resistance to the feasibility of real-world applications for mobile devices,<sup>8</sup> the evidence indicates that healthcare workers<sup>9</sup> and patients<sup>10</sup> are ready to take advantage of this opportunity to improve care and communication in hospitals, clinics, and homes.

The quality of microphones embedded in mobile devices has improved with advances in technology, offering a low-cost, and accessible alternative to studio microphones traditionally used for speech analysis. Changes in acoustic voice features can be used to quantify speech impairment in neurodegenerative disorders including Parkinson's disease (PD) and Huntington's disease (HD),<sup>11–13</sup> and offer the potential for quantitative clinical evaluation of symptoms and disease progression. However, there is a need to ensure the equivalence of acoustic voice features derived from recordings made from microphones in mobile devices before they can be deployed in healthcare applications.

The use of mobile devices to record structured and free speech has been previously investigated, mainly in healthy individuals<sup>14–18</sup> and with synthetic voice samples,<sup>19,20</sup> but also in populations with speech disorders, such as glottic cancer, vocal fold paralysis, and dysphonia.<sup>21–24</sup> No significant differences were reported between recordings of the open vowel [a:] from an iOS smartphone allied with an iRig when compared with the gold standard computer/preamplifier configuration in healthy individuals.<sup>14</sup> A subsequent study in an English-speaking cohort, during sustained vowel

Accepted for publication October 10, 2022.

Funding: This work was funded through the EU joint program for Neurodegenerative Disease Research as part of the JPND funding call in to Health and Social Care (2019) with funding from Alzheimer's Society, Secretary of State for Health and Social Care, Health and Care Research Wales, Public Health Agency Northern Ireland, Jacques and Gloria Gossweiler Foundation, Bundesministerium für Bildung und Forschung, Narodowe Centrum Badań i Rozwoju, Swiss National Science Foundation (SNF), 32ND30\_185548 and Health Research Board (JPND-HSC-2018-003).

From the \*School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland; †Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland; ‡Centre for Trials Research, Cardiff University, Cardiff, UK; §School of Medicine, Trinity College Dublin, Dublin, Ireland; and the ||Bloomfield Health Services, Dublin, Ireland.

Address correspondence and reprint requests to Vitória S. Fahed, School of Electrical and Electronic Engineering, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: [vitoria.fahed@ucdconnect.ie](mailto:vitoria.fahed@ucdconnect.ie)

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■–■■■  
0892-1997

© 2022 The Authors. Published by Elsevier Inc. on behalf of The Voice Foundation. This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

<https://doi.org/10.1016/j.jvoice.2022.10.006>

(SV) and reading passage (RP), compared acoustic features from audio recorded with both studio microphone and iPhone.<sup>16</sup> Acceptable agreement in the mean fundamental frequency across devices was observed, however the random error for jitter and shimmer were deemed too high for practical applications. Another study in healthy participants, which analyzed SV and RP, showed correlation between features derived from mobile devices, landline phone, and head-mounted microphone, but only the fundamental frequency and cepstral peak prominence had an error below 10% when compared with microphone recordings.<sup>15</sup> Significant bias has also been reported for all acoustic voice features analyzed during SV and RP for a range of Android and iOS devices simultaneously recording voice from 22 speakers.<sup>18</sup> Shimmer and HNR, but not fundamental frequency, were reported to differ across devices in a Flemish Dutch-speaking cohort of both healthy and patients with a variety of speech disorders, when assessing SV and RP.<sup>22</sup> Considering the available evidence, while various approaches have been used to compare the performance of mobile devices and studio microphones, their findings are not always in agreement. This may be due to differences in the language spoken, devices, and methods of speech collection or analysis. Additionally, the majority of validation studies presented to-date have focused on analyzing acoustic voice features in speech tests comprised of SV and RP tasks, few have included syllable repetition (SR) in the analysis.<sup>25,26</sup> Consequently, it remains unclear whether acoustic voice features recorded using mobile devices provide sufficient agreement with gold standard microphone to be used in remote monitoring and clinical assessment.

Several studies have analyzed speech in PD using mobile technology,<sup>25,27–29</sup> including a mobile device validation study.<sup>30</sup> However, though quantitative analysis of speech in HD is receiving increasing attention, the reliability of mobile devices for monitoring speech in HD has not been examined to-date. Characterizing changes and patterns of speech in those with HD can add to our understanding of this disorder. Such differences, measured reliably with mobile technology, may be used in clinical evaluations of disease progression or person-centered outcomes in future clinical trials of interventions for those with HD. Lessons learnt from studying those with HD may have implications for those with other less well defined dementias, now and into the future.

The objective of this study was to compare acoustic features extracted from voice recordings using mobile devices (smartphone and tablet) and a high-quality studio microphone allied with an audio interface/preamplifier in a group of control participants and participants with HD. Three speech tests were examined, SV, RP, and SR, and acoustic voice features were extracted, focusing primarily on features related to voice quality including jitter and shimmer. The results provide guidance on the use of mobile devices for recording speech for the analysis of acoustic voice features during vocal tasks typically used during motor speech assessment.

## MATERIAL AND METHODS

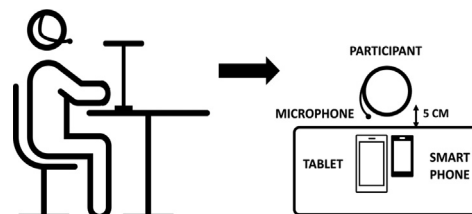
### Data collection

Thirty-seven healthy adults (aged  $34.16 \pm 12.75$  years; 20 female) and six participants with genetically confirmed HD (aged  $53.5 \pm 16.78$  years; three female and three males, all with manifest stages of HD) gave their written consent to participate in the study which was approved by the University College Dublin Human Research Ethics Committee, in collaboration with Bloomfield Health Services, Dublin. Voice was recorded simultaneously with a smartphone device (Google Pixel 4), a tablet (Samsung Tab S6 Lite) and a headset omnidirectional microphone (6066, DPA, Denmark) with a sound interface (Mix Pre3 II, Sound Devices, USA) connected to a laptop. The recordings took place in the university research laboratory for the healthy participants (control group) and in the clinical setting for the HD group, with environmental noise measured with a sound level meter, always below 45 dBA. Participants were seated wearing the headset microphone, and facing the tablet and the smartphone, which were both placed on tripods, at the height of the participants' mouth, approximately 5 to 10 cm away, [Figure 1](#). The equipment and protocol for data collection were selected according to the recommendations for instrumental assessment of voice from the American Speech-Language-Hearing Association.<sup>31</sup> Acoustic data were recorded during SV phonation ([a:]),<sup>32,33</sup> reading of the Rainbow Passage<sup>32,33</sup> and SR ([pa], [ta], [ka], [pataka], and [pati]) tasks, commonly known as sequential and alternating motor rate tasks.<sup>33</sup> Participants were asked to sustain the vowel "ah" for as long as they could during one breath. They were instructed to read the Rainbow Passage at their own pace, as they would naturally read out loud. Finally, participants were asked to repeat each syllable as fast and as clearly as they could for 5 seconds. Participants repeated the tasks two or three times. Audio data were sampled at 24-bit, 44.1 kHz or 48 kHz, and saved in an uncompressed .wav format.

### Signal analysis

#### Preprocessing

Speech signals were filtered with a fourth order Butterworth band-pass filter between the 10 Hz and 5 kHz, and down-sampled to 44.1 kHz for consistency across devices. The mean was removed from each signal, 0.5 seconds of data at the start and end of the signal was discarded to avoid edge errors, and the signal was padded with zeros for 2 seconds



**FIGURE 1.** Schematic diagram of experimental set-up for speech recording with three devices simultaneously.

at the start and end. Finally, each signal was normalized with respect to its maximum amplitude.

### Voiced/unvoiced detector

A voiced/unvoiced detector was developed to ensure that the estimated features corresponded to voiced frames. The Teager-Kaiser Energy operator (TKO)<sup>34</sup> was applied to all recordings, followed by a three-sample maximum rolling window, and a five-sample mean rolling window. Finally, the recordings were low-passed filtered at 10 Hz with a second order Butterworth filter, to extract the signal envelope. The methods for voiced/unvoiced detector for the recordings for each speech task (SV, SR, PR) differed after this step.

For the SV task, a threshold was implemented to detect candidates for the onset and offset times, Figure 2. The threshold for the control group was set as 10% of the mean of the Teager-Kaiser envelope, and for the HD group, 20% of the mean of the envelope. The onset/offset times were defined for each phonation. If the signal contained more than one voice break during phonation, then the interval between each pair of detected voice onset/offset times was calculated and the pair with the maximum interval chosen.

The onset/offset detector was used also for the detection of the voice signal during RP, with a threshold of 70% of the mean of the Teager-Kaiser envelope for the control group and 80% of the mean of the envelope for the HD group. All the voiced parts during the passage were detected, Figure 2. The unvoiced parts were not included in the analysis.

For the SR, the onset/offset detector was deployed with the threshold for detection set as the mean of the Teager-Kaiser envelope for both the control and the HD group. Each pair of onset and offset times corresponded to a syllable, Figure 2. A minimum duration of the syllable phonation of 0.05 seconds was applied to avoid detection of noise with short duration and high amplitude. Each syllable repetition was analyzed individually without windowing. The

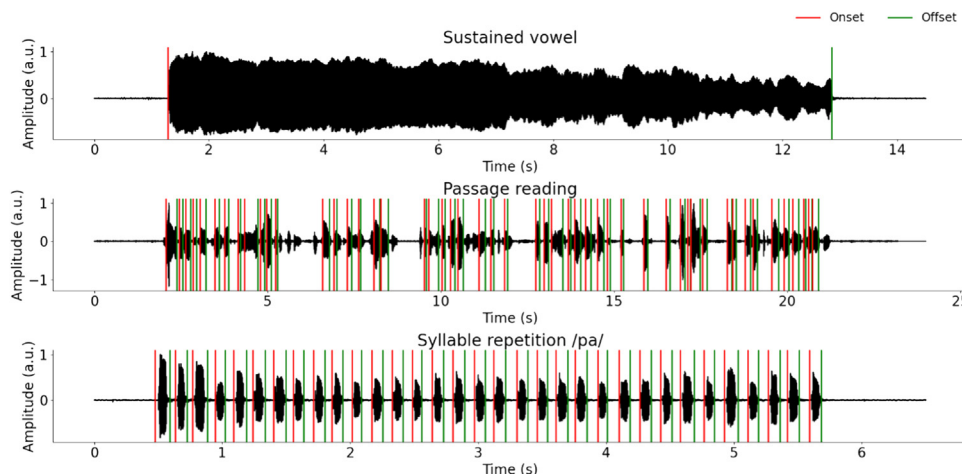
first and last syllables were removed to avoid detecting noise at the beginning and end of the recording, unless five or less syllable repetitions were detected in the trial.

### Feature estimation

For SV, a 0.5 seconds moving window with 50% overlap was applied to the detected voiced data for feature estimation. Acoustic voice features were estimated for each 0.5-second epoch and averaged across all epochs to obtain the final estimate for that feature. To avoid edge errors, the first and last window of each signal were discarded if the vowel was sustained for more than 5 seconds total. Otherwise, the entirety of the signal was analyzed.

For PR, each voiced interval was used as one epoch for feature calculation. Finally, acoustic voice features were estimated for each syllable for the SR task and the median value across one repetition of the task was taken as the value of that feature.

The features examined are commonly used in voice assessment in neurodegenerative diseases, such as PD<sup>35</sup> and HD,<sup>36</sup> and were compared across devices. The following frequency-based features were estimated for each recording: Median and standard deviation of the fundamental frequency (F0 and SD F0), harmonics-to-noise ratio (HNR), local jitter in percentage (jitter), and the jitter variants: relative average perturbation of jitter in percentage (RAP), five-point period perturbation quotient in percentage (PPQ5), difference of differences of periods in percentage (DDP). In addition, the following amplitude-based features were also estimated: shimmer in decibels (shimmer), and the shimmer variants: three-point amplitude perturbation quotient in percentage (APQ3), the five-point amplitude quotient in percentage (APQ5), the 11-point amplitude quotient in percentage (APQ11), the difference of differences of amplitude in percentage (DDA). The features were estimated using custom Python scripts as described below. Following publication, the control data and Python code for estimating acoustic voice features will be made publicly available.



**FIGURE 2.** Sample audio data recorded using a microphone from a representative control subject. The onset and offset times determined using the voiced/unvoiced detector for each speech task are indicated.

A number of different methodologies and algorithms have been used to estimate the fundamental frequency of audio signals.<sup>37–40</sup> Here, a 70:30 weighted linear combination was implemented to calculate the fundamental frequency,<sup>35</sup> with 70% being comprised of a modified Praat F0 calculation<sup>37</sup> and a Librosa<sup>41</sup> implementation of the YIN algorithm<sup>38</sup> accounted for the remaining 30%. The 70:30 weighting was selected to provide accurate estimates of F0 across epochs of varying length, including shorter epochs for which the YIN algorithm is susceptible to higher variability. The modified version of the Praat algorithm for estimating the fundamental frequency was implemented in Python without the path finder and with a preference for higher frequencies. For estimation of the fundamental frequency, each voiced epoch was divided into 0.04-second windows with 75% overlap. Within each window, candidates for F0 that fell within the frequency range of interest (75 Hz and 600 Hz) were identified and the candidate that represented the lowest lag selected as F0. The median value of F0 across all windows within each epoch was estimated.

HNR was also estimated based on the Praat algorithm,<sup>37</sup> which calculates the heights of the peaks of the autocorrelation function. Each epoch was divided into 0.08-second windows with 75% overlap. The highest amplitude peaks within the autocorrelation function were identified, the two with the greatest amplitude were chosen and their difference in amplitude calculated. If the absolute difference was less than 10% the algorithm defaulted to the candidate lying at the higher frequency. Otherwise, the algorithm chose the candidate with the highest amplitude in the autocorrelation.

The jitter and its variants were estimated from a period sequence formed by taking the inverse of the fundamental frequency for each window within an epoch as described above. Shimmer and its variants were calculated from the amplitude sequence formed by the normalized value of the amplitude of the signal at each instance of glottal opening and closing identified using the MATLAB toolbox Voicebox for each epoch.<sup>42</sup>

In addition to the effect of device, acoustic voice features estimated using the entire phonation of the vowel were compared with the values obtained by averaging across overlapping 0.5-second epochs.

### Statistical analysis

Bland Altman analysis<sup>43</sup> was used to examine the agreement between features estimated from acoustic data recorded using the different devices. Linear mixed models (LMM)<sup>44</sup> were implemented in R<sup>45</sup> for each speech task and cohort group separately, to investigate differences in the acoustic features across devices for each cohort. For all features, device, age, and biological sex were included as fixed effects and the intercept for participants was included as a random effect. For the SR task, all five syllables were included into the same model, and syllable was added as another random effect. *P* values were obtained by comparing the model with a null model without the effect of device using an

ANOVA.<sup>44</sup> The effect of device was investigated, along with the effect of age and sex as a secondary objective. If a significance difference was observed, the False Discovery Rate was used as a post hoc test within the devices group. Bonferroni correction was applied to account for multiple comparisons within each family of speech task (SV, RP, and SR), with the threshold for significance set at  $P < 0.004$ . Correlation analysis between microphone and mobile devices was performed using the Spearman rank-order correlation coefficient. Correlations were examined for the control group only, due to the small sample size of the HD group. Threshold levels of significance for correlation coefficients were also adjusted for multiple comparisons using Bonferroni correction.

## RESULTS

The results for each of the vocal tasks are presented separately in four sections: 3.1 Sustained vowel phonation, 3.2 Reading passage, 3.3 Syllable repetition, and 3.4 Effect of biological sex, age, and epoch size. Within the first three sections (3.1, 3.2, and 3.3), an analysis of the Bland-Altman plots is first presented, followed by the results of the linear mixed models and the correlation analysis between devices. Lastly, Section 3.4 presents the effect of biological sex and age on acoustic voice features across the devices is described. At the end of Section 3.4, the effect of the length of the time epoch chosen for analysis in the SV test is also reported. Exact *P* values for all statistical tests are presented in the Appendix.

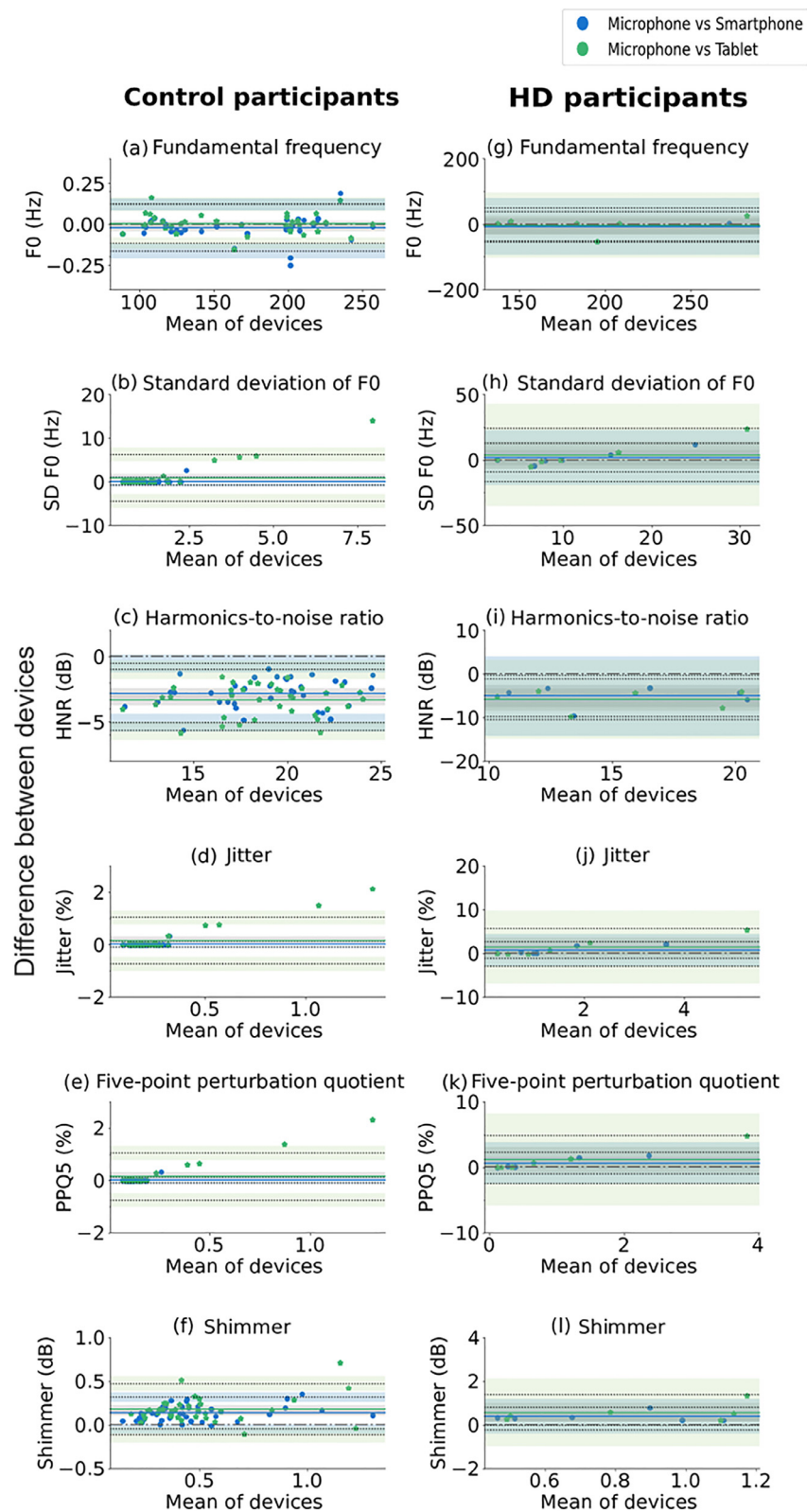
### Sustained vowel phonation

Bland-Altman analysis revealed a significant bias in HNR, and shimmer and its variants, with both mobile devices when compared with the microphone recording during SV phonation in control and HD groups, [Figure 3](#).

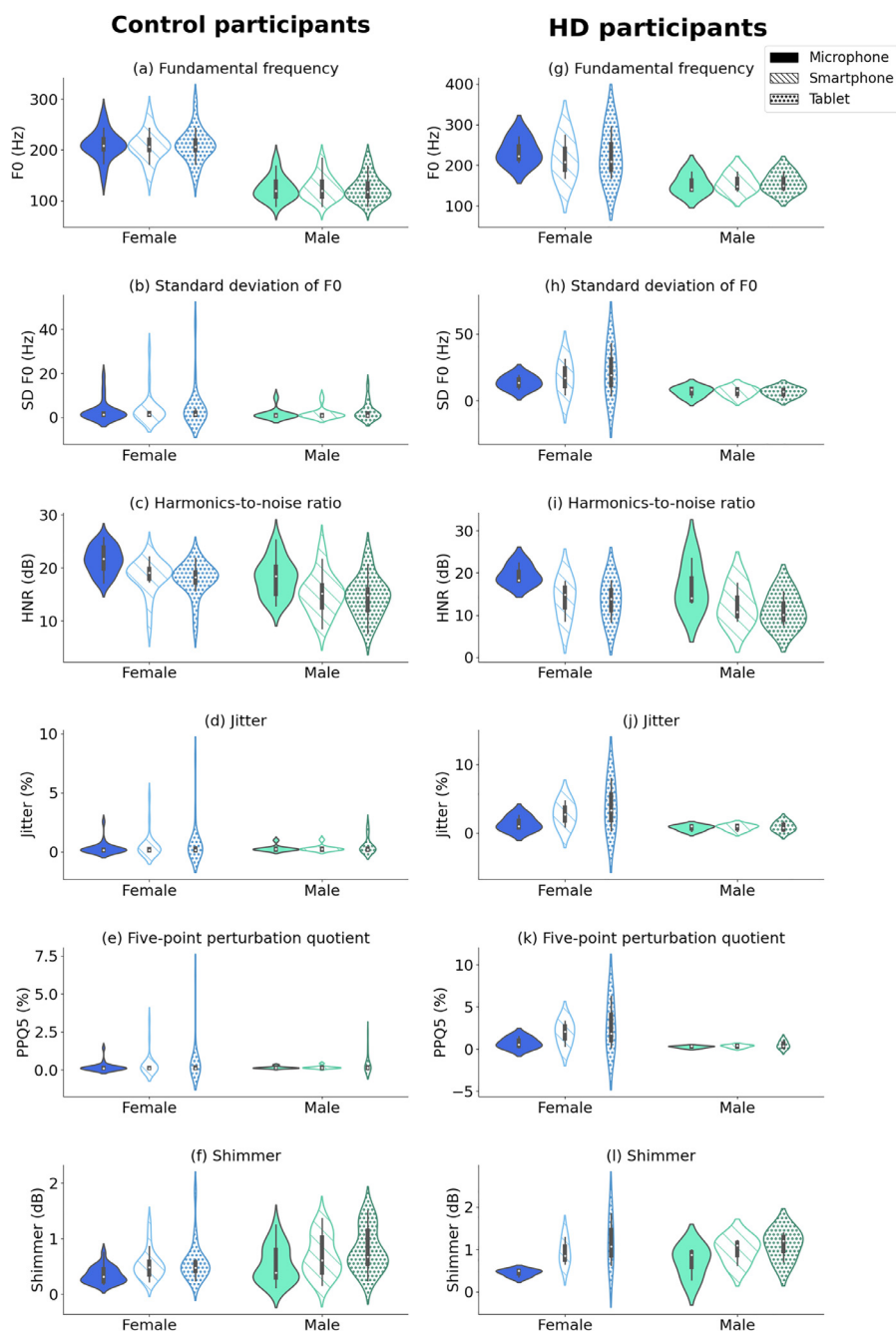
Linear mixed model analysis revealed significant differences in HNR, and shimmer and its variants, between microphone and mobile devices for both cohort groups, with lower HNR and higher shimmer values observed with the mobile devices, [Figure 4](#). Values for the mean and standard deviation of the estimated acoustic voice features along with the *P* values for the linear-mixed models are presented in the Appendix in [Tables A.1](#) and [A.2](#) for the control and HD groups, respectively. Post hoc analysis confirmed significant differences in these features between microphone and smartphone, and microphone and tablet, but not between smartphone and tablet for these features, [Table A.2](#). A strong correlation was observed between both mobile and the studio microphone recordings for all features examined for the control group, [Table 1](#).

### Reading passage

During passage-reading, Bland-Altman analysis revealed a significant bias in HNR, jitter, RAP, DDP, shimmer and its variants between microphone and mobile devices for the



**FIGURE 3.** Bland-Altman plots comparing the gold standard microphone with mobile devices during the SV for the control group: (a) fundamental frequency, (b) standard deviation of the fundamental frequency, (c) harmonics-to-noise ratio, (d) jitter, (e) 5 five-point perturbation quotient, and (f) shimmer; and for the HD group: (g) fundamental frequency, (h) standard deviation of the fundamental frequency, (i) harmonics-to-noise ratio, (j) jitter, (k) 5 five-point perturbation quotient, and (l) shimmer. Data points and confidence intervals for the comparison of microphone and smartphone are indicated in blue; data points and confidence intervals for the comparison of microphone and tablet are in green (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).



**FIGURE 4.** Violin plots illustrating the distribution of the estimated features during SV for the control group: (a) fundamental frequency, (b) standard deviation of the fundamental frequency, (c) harmonics-to-noise ratio, (d) jitter, (e) 5 five-point perturbation quotient, and (f) shimmer; and for the HD group: (g) fundamental frequency, (h) standard deviation of the fundamental frequency, (i) harmonics-to-noise ratio, (j) jitter, (k) 5 five-point perturbation quotient, and (l) shimmer.

control group. Additionally, significant bias was observed for SD F0 and PPQ5 between microphone and tablet, [Figure 5](#). Significant bias was observed for F0, HNR, shimmer and its variants when comparing microphone and mobile devices for recordings for the HD group. Additional bias was observed for RAP, PPQ5 and DDP when comparing microphone with smartphone for the HD group, [Figure 5](#).

Linear mixed models confirmed an effect of device on all features except F0 in the control group, [Figure 6](#) and

[Table A.3](#). Further post hoc analysis revealed differences between microphone and smartphone for HNR, shimmer and its variants. Differences between microphone and tablet were observed for all features that reported an effect of device. In the HD group, significant differences were observed in HNR, PPQ5, shimmer, APQ3, APQ5, and DDA. Post hoc analysis revealed differences between microphone and smartphone, and between microphone and tablet for HNR, shimmer, APQ3, APQ5, and DDA, [Table A.4](#). As for the RP task, correlation analysis revealed that all the

**TABLE 1.**  
**Correlation Coefficients Rho for All the Features Estimated From Each Task Recorded From the Control Group**

Feature	Sustained Vowel		Reading Passage		Syllable Repetition	
	Microphone Vs smartphone	Microphone Vs Tablet	Microphone Vs Smartphone	Microphone Vs Tablet	Microphone Vs Smartphone	Microphone Vs Tablet
F0	1.00 **	1.00 **	1.00 **	1.00 **	0.98 **	0.96 **
SD F0	0.97 **	0.82 **	0.73 **	0.55 *	0.74 **	0.72 **
HNR	0.97 **	0.94 **	0.99 **	0.98 **	0.97 **	0.97 **
Jitter	0.93 **	0.82 **	0.83 **	0.72 **	0.89 **	0.89 **
RAP	0.91 **	0.83 **	0.94 **	0.84 **	0.90 **	0.89 **
PPQ5	0.90 **	0.81 **	0.89 **	0.80 **	0.88 **	0.84 **
DDP	0.91 **	0.83 **	0.93 **	0.84 **	0.90 **	0.89 **
Shimmer	0.88 **	0.80 **	0.95 **	0.90 **	0.84 **	0.78 **
APQ3	0.83 **	0.85 **	0.93 **	0.88 **	0.83 **	0.74 **
APQ5	0.88 **	0.83 **	0.90 **	0.86 **	0.91 **	0.85 **
APQ11	0.88 **	0.77 **	0.94 **	0.88 **	0.95 **	0.95 **
DDA	0.83 **	0.85 **	0.92 **	0.86 **	0.81 **	0.76 **

\*  $P < 0.004$ .

\*\*  $P < 0.001$ .

features were highly correlated across devices except SD F0 for the control group, [Table 1](#).

### Syllable repetition

Similar to the sustained phonation task, a significant bias was observed for HNR, shimmer and its variants when comparing microphone with mobile devices for the control group, [Figure 7](#). A significant bias was also observed in the control group when comparing jitter, RAP and DDP from the tablet and microphone recordings. When analyzing features from the HD group, a significant bias with mobile devices was observed for HNR, PPQ5, and shimmer and its variants.

A significant effect of device was observed for HNR, shimmer and its variants, in both groups, [Figure 8](#), [Tables A.5](#), [A.6](#). Additionally, an effect of device was also observed for RAP in the control group. Post hoc analysis revealed significant differences between all three devices for HNR, shimmer, APQ3, APQ5, and DDA in the control group, and between microphone and tablet for RAP and APQ11. For the HD group, an effect of device was observed when comparing microphone and smartphone, and microphone and tablet, but not between smartphone and tablet. Correlation analysis confirmed a high correlation coefficient between microphone and smartphone, and microphone and tablet from the control group estimated features ( $P < 0.001$ ), [Table 1](#).

### Effect of biological sex, age, and epoch size

An effect of biological sex on the fundamental frequency and on the harmonics-to-noise ratio was observed for the control group ( $P < 0.004$ ) during the SV task. No significant difference with biological sex was observed for any feature in the HD group during the SV task, [Table A.7](#). An effect of biological sex on F0, HNR, shimmer and its variants was

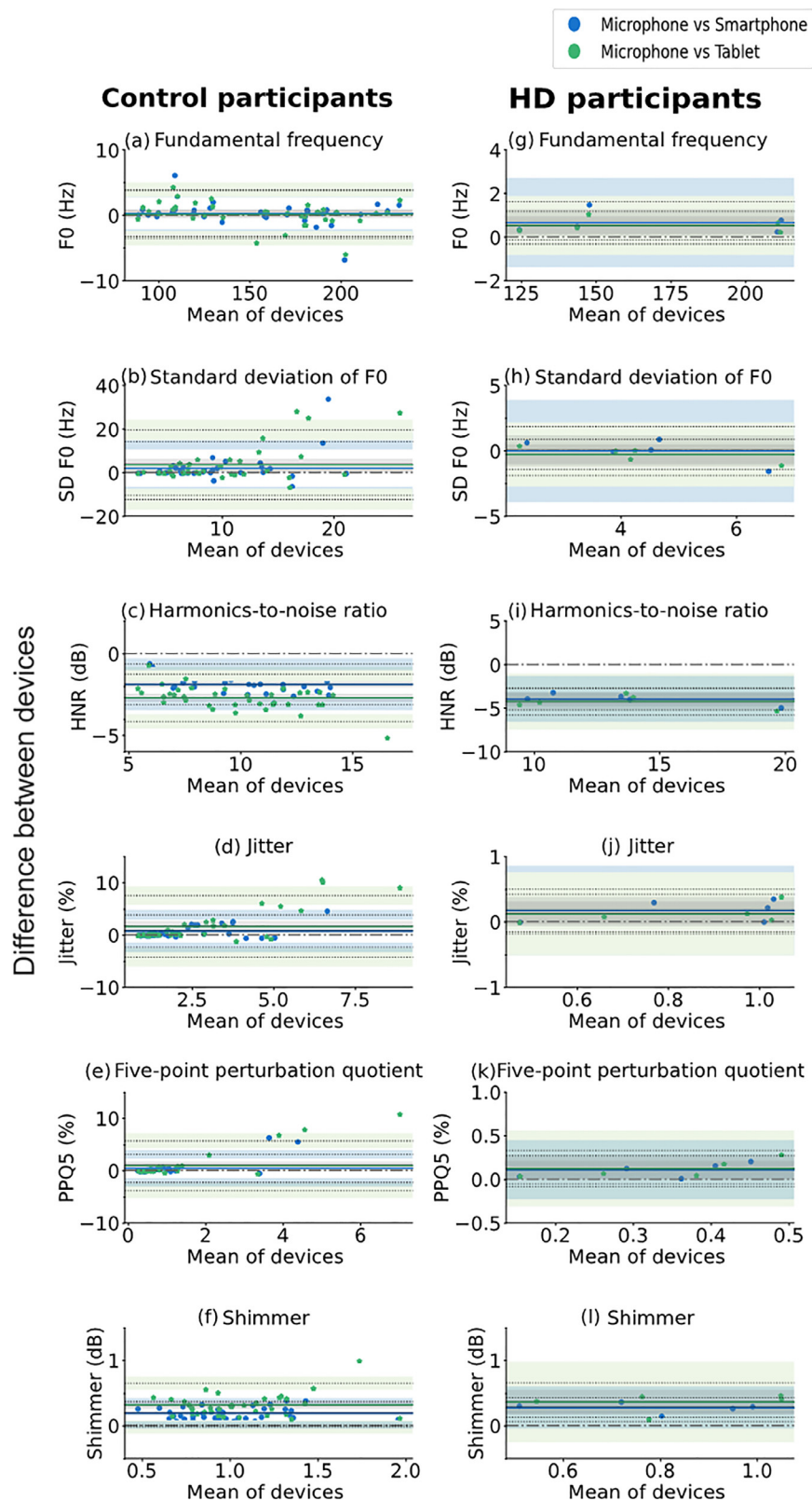
also observed during the RP task for the control group, with an effect of biological sex observed only for F0 for the HD group. For the SR task, an effect of biological sex was observed in all features with the exception of APQ5, APQ11, and DDA for the control group, [Table A.7](#). No significant effect of biological sex was observed in the HD group.

Age had no significant effect in any of the features estimated from all three speech tasks, SV, RP, and SR, [Table A.8](#). Finally, the effect of epoch length was investigated for the SV task. When the entire signal was analyzed the results were consistent with those obtained using a moving epoch, with an effect of device observed for HNR, shimmer, and its variants for both cohort groups, [Table A.9](#).

## DISCUSSION

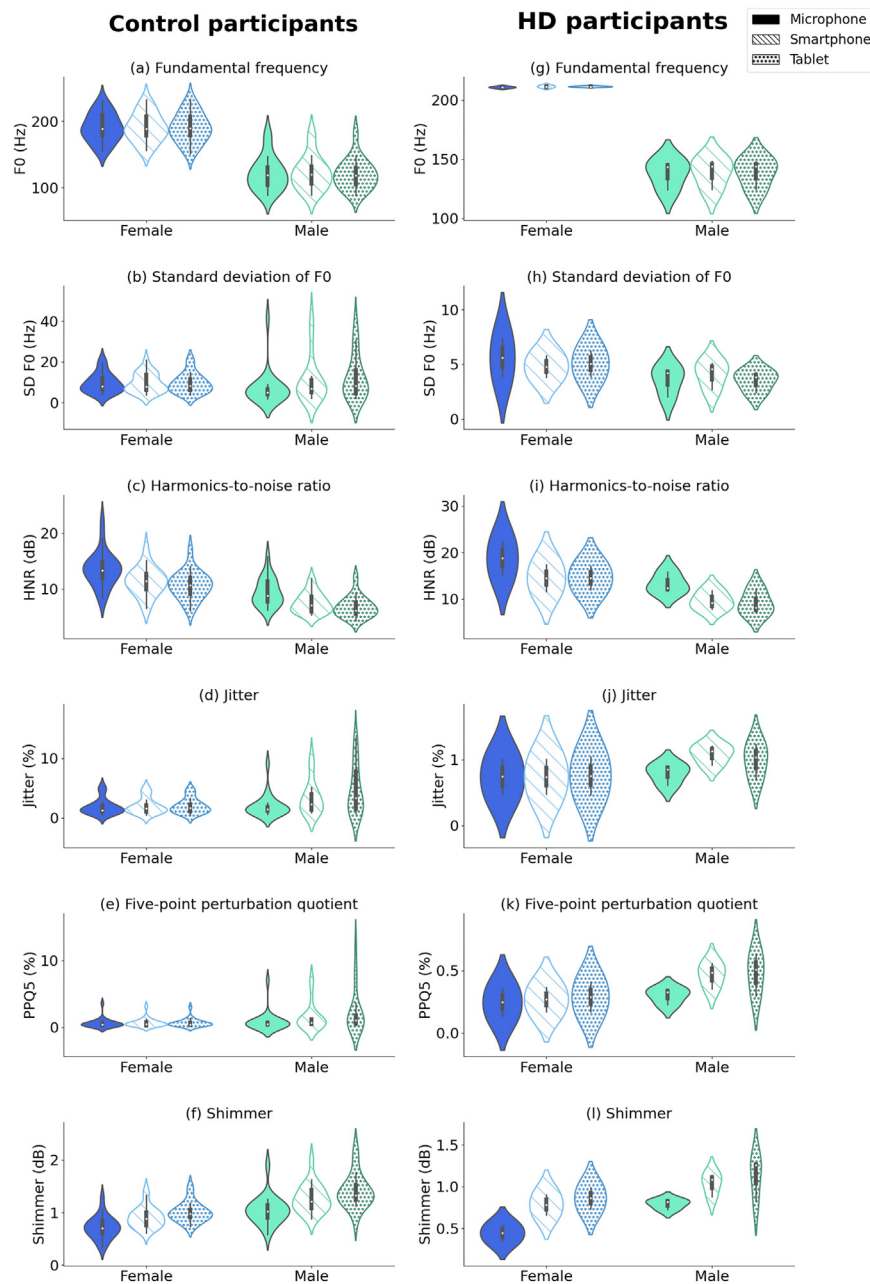
In this study, quantitative acoustic voice features extracted from speech recordings made using Android mobile devices were compared with those derived from recordings made using gold standard technology, an omnidirectional microphone allied with a preamplifier. Acoustic voice features were estimated from audio recordings during three speech tasks, sustained phonation of the vowel [a:], reading of the Rainbow Passage, and repetition of syllables in healthy control participants and in a small cohort of participants with HD.

The results indicate that fundamental frequency could be estimated with similar accuracy across all devices and speech tasks examined for both the control group in a laboratory setting and the HD group in a clinical setting, consistent with a number of previous studies.<sup>14,16,29,46</sup> No difference in F0 was observed across devices and tasks, with a significant bias observed only in the case of RP task for the HD group, [Figure 5](#). The observed bias, however, was small at 0.66 Hz between microphone and smartphone, and  $-0.53$  Hz between microphone and tablet for



**FIGURE 5.** Bland-Altman plots comparing the gold standard microphone with mobile devices during RP for the control group: (a) fundamental frequency, (b) standard deviation of the fundamental frequency, (c) harmonics-to-noise ratio, (d) jitter, (e) 5 five-point perturbation quotient, and (f) shimmer; and for the HD group: (g) fundamental frequency, (h) standard deviation of the fundamental frequency, (i) harmonics-to-noise ratio, (j) jitter, (k) 5 five-point perturbation quotient, and (l) shimmer. Blue markers and confidence intervals refer to the comparison of microphone and smartphone, and green markers and confidence intervals refer to the comparison of microphone and tablet (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).

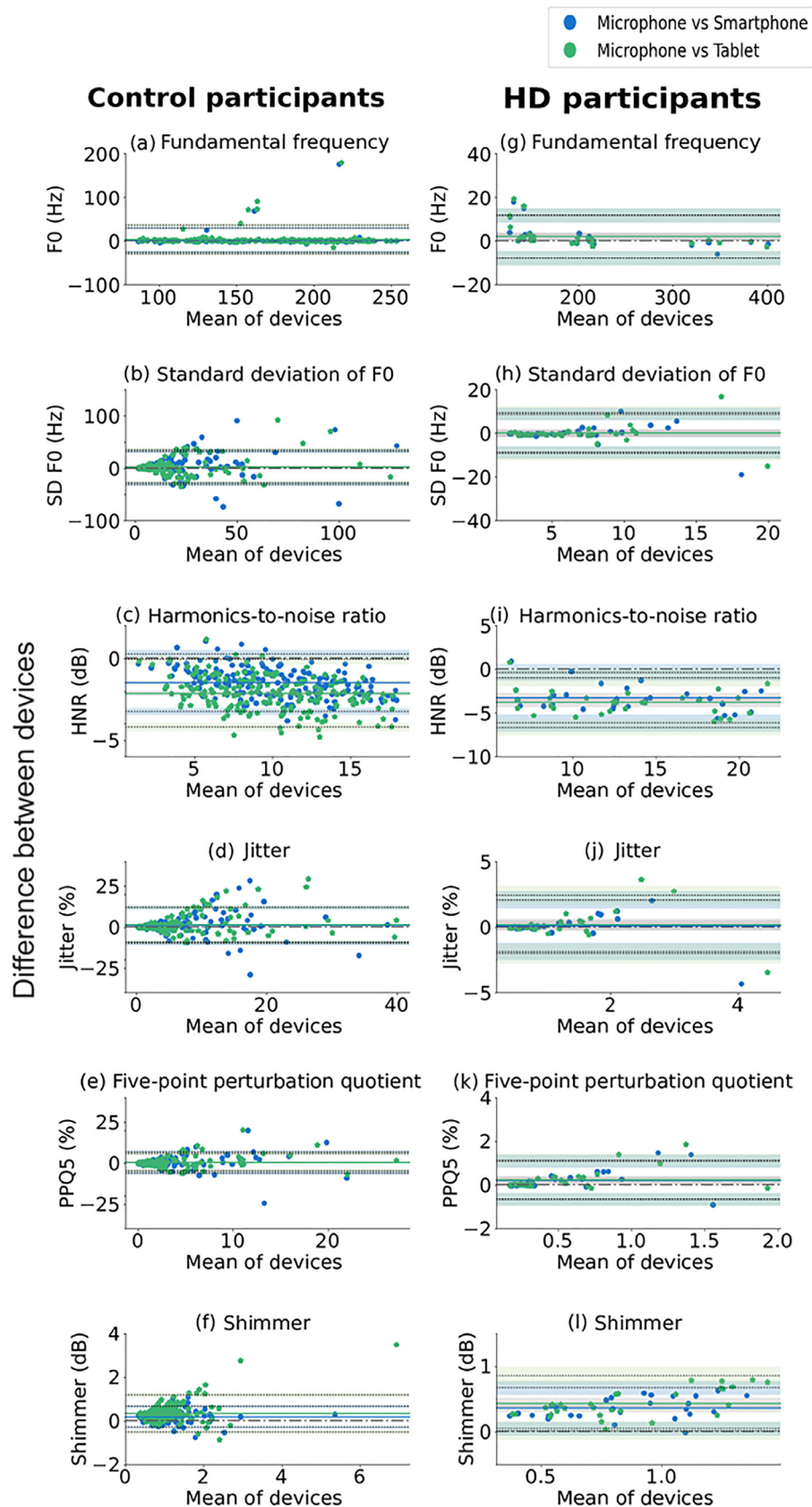




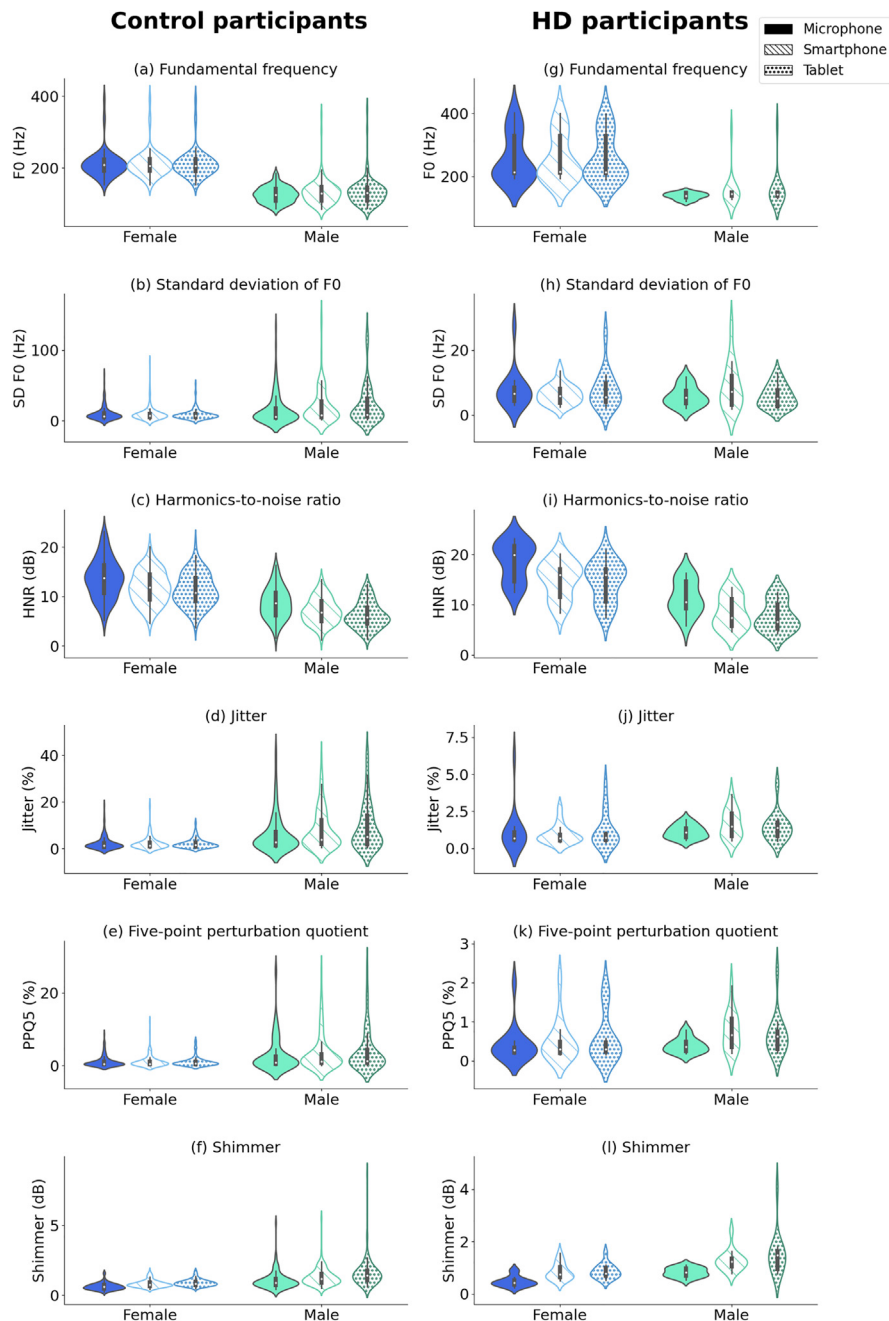
**FIGURE 6.** Violin plots illustrating the distribution of the estimated features during RP for the control group: (a) fundamental frequency, (b) standard deviation of the fundamental frequency, (c) harmonics-to-noise ratio, (d) jitter, (e) 5 five-point perturbation quotient, and (f) shimmer; and for the HD group: (g) fundamental frequency, (h) standard deviation of the fundamental frequency, (i) harmonics-to-noise ratio, (j) jitter, (k) 5 five-point perturbation quotient, and (l) shimmer.

mean values of 167.24 Hz for microphone, 167.90 Hz for smartphone, and 167.77 Hz for tablet, Table A.4. As the first harmonic in the frequency domain, the fundamental frequency can be clearly detected even in the presence of noise or interference. In contrast, HNR was consistently different across devices for all tasks in control and HD groups, Figures 3 to 8 and Tables A.1 to A.6. A bias in the HNR recorded using the microphone and smartphone was observed for all tasks in both groups. The bias in HNR was relatively large, at approximately  $-2$  dB in all comparisons, Figures 3, 5, and 7.

Good agreement across devices was observed for the other frequency-derived features, SD F0, jitter, and its variants, RAP, PPQ5, and DDP, Figures 3, 5, and 7, in partial agreement with previous studies.<sup>14,16,29</sup> The exception to this was during SR for the control group in which significant differences were observed between microphone and tablet for the jitter variant RAP. Though significant bias was present for some features during SR, Figure 7, unlike linear mixed models, Bland-Altman analysis does not account for repeated measures within subjects or multiple comparisons, thus the mean of repeated measures must be estimated



**FIGURE 7.** Bland-Altman plots comparing the gold standard microphone with mobile devices during the SR for the control group: (a) fundamental frequency, (b) standard deviation of the fundamental frequency, (c) harmonics-to-noise ratio, (d) jitter, (e) 5 five-point perturbation quotient, and (f) shimmer; and for the HD group: (g) fundamental frequency, (h) standard deviation of the fundamental frequency, (i) harmonics-to-noise ratio, (j) jitter, (k) 5 five-point perturbation quotient, and (l) shimmer. Blue markers and confidence intervals refer to the comparison of microphone and smartphone, and green markers and confidence interval refer to the comparison of microphone and tablet (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).



**FIGURE 8.** Violin plots illustrating the distribution of the estimated features during the SR. Data presented here represent control group ( $N = 37$ ) and participants with HD ( $N = 6$ ), repeating five different syllables. Control group: (a) fundamental frequency, (b) standard deviation of the fundamental frequency, (c) harmonics-to-noise ratio, (d) jitter, (e) 5 five-point perturbation quotient, and (f) shimmer; and for the HD group: (g) fundamental frequency, (h) standard deviation of the fundamental frequency, (i) harmonics-to-noise ratio, (j) jitter, (k) 5 five-point perturbation quotient, and (l) shimmer.

rather than using the individual data points.<sup>43</sup> Furthermore, the relative bias introduced by the tablet was relatively small (1.29%, 0.83%, and 1.96% for the jitter, RAP, and DDP, respectively).

The amplitude-based features, shimmer and its variants, APQ3, APQ5, APQ11, and DDA, were found to be significantly different between microphone and smartphone, and between microphone and tablet recordings for all tasks across both groups, Figures 4, 6, and 8 and Tables A.1 to

A.6. Significant bias was observed across all tasks and groups. This result is not unexpected since shimmer provides a measure of the variability in the amplitude of the signal.<sup>47</sup> The experimental setup used a smartphone and tablet placed at the same height as the participants mouth, however, the microphone was head mounted, so it was placed closer to the source, leading to a higher amplitude voice signal. Relatively higher levels of environmental noise were thus likely present in the mobile devices, even though noise

was kept below 45 dBA. Despite the differences observed between devices for the amplitude-based features, correlation analysis revealed a high correlation ( $\rho > 0.7$ ,  $P < 0.001$ ) for all features examined, apart from SD F0 for the RP task, Table 1. These differences in correlation across devices between RP and the other speech tasks may be due to the more dynamic nature of the task<sup>48</sup> which would lead to a greater variability in the acoustic features extracted from it, especially SD F0, which is the variance of the fundamental frequency.

Previous comparison studies between mobile devices and microphone for acoustic voice analysis of healthy and pathological voices have reported conflicting results. Our results are in agreement with previous findings of high correlation between F0, SD F0, and jitter from microphone and Android smartphone devices.<sup>49</sup> However, the analysis in that study was conducted on running speech using a custom-made application for Android. Another study which analyzed SV and RP in a Dutch-speaking population found that F0 and jitter were consistent across devices, and HNR and shimmer were significantly different when comparing a head-mounted microphone with smartphone and tablet,<sup>22</sup> which is in agreement with our findings.

Our results are in partial agreement also with previous studies which reported that F0 and jitter are reliable measures across devices during sustained phonation of vowel [a:], but unlike our study, reported that shimmer could also be reliably estimated.<sup>16,29</sup> Our results indicate that F0 and jitter can be reliably estimated across devices but caution should be taken when estimating shimmer. This is in partial agreement with earlier results comparing signals recorded using an AKG microphone and Samsung smartphone during SV [a:] which observed that F0 and shimmer were consistent across devices in a control group, though jitter differed, and all the three features were consistent across devices in a cohort with pathological voices.<sup>46</sup> Also, in partial agreement with our findings, a previous study found no differences between any of the acoustic voice features estimated from the SV [a:] with iOS iPad and iPhone allied with an iRig when compared with a studio microphone allied with a preamplifier.<sup>14</sup> This difference in results could be due higher performance and signal quality of the iRig which is an audio interface for musical instruments. In contrast to our findings, a previous study reported that HNR was consistent when measured with a head-mounted microphone and smartphone held by the participant.<sup>29</sup> This difference could be due to several factors, including differences in native language (Czech), or variability in the cohorts examined, which comprised of controls, PD, and sleep disorder participants.

The effect of biological sex is expected on the fundamental frequency, with males typically having a lower F0 than females.<sup>50</sup> This was confirmed by differences in biological sex observed for fundamental frequency during the SV and

SR tasks for the control group, Table A.7. However, they were not observed in SV and SR tasks performed by the participants with HD, Table A.7. The effect of sex might have been more pronounced in RP since it involves the phonation of different words that may have led to a greater sensitivity to biological sex, even with a small sample size which may not have been sufficient to distinguish between sex in highly structured tasks, such as SV and SR. In addition, while the effect of disease was not investigated in this study, it may have had an effect on the acoustic features estimated. An effect of biological sex was also present in HNR in the SV task, and in SD F0, HNR, jitter, RAP, PPQ5, DDP, and shimmer in the SR task. No effect of age was observed in any of the features across all tests and cohorts, Table A.8.

The linear mixed model,<sup>44</sup> together with Bland-Altman and correlation analysis, provides a more comprehensive evaluation of the effect of device on acoustic voice features than a single metric alone. The results of the linear mixed models were almost always in agreement with the Bland-Altman plots, confirming significant differences where a significant bias was observed. However, though linear mixed models can control for multiple factors and avoid pseudoreplication, it is essentially a comparison of means. The correlation analysis showed that all the features were highly correlated across devices even where significant bias was observed. This is in line with previous findings of high correlation between acoustic voice features from microphone and smartphones examined for control participants and synthesized voice data.<sup>19,51</sup> Collectively the results suggest that frequency-based features can be accurately derived from recordings with mobile devices, but that amplitude-based features and HNR, though highly correlated across devices, may be sensitive to device type and location.

## CONCLUSIONS

Acoustic voice features can be reliably estimated using mobile devices for sustain phonation of vowel [a:], reading of the Rainbow Passage and repeating syllables, [pa], [ta], [ka], [pataka], [pati], if consistency in device and device placement is ensured. Caution should be taken with HNR and amplitude-based features such as shimmer, APQ3, APQ5, APQ11, and DDA, as these features are sensitive to environmental noise and signal amplitude which is dependent on the position of the device and microphone quality.

## Acknowledgments

The authors would like to acknowledge the DOMINO-HD Consortium, Bloomfield Health Services, and Neuromuscular Systems Group, in particular Muthukumaran Thangarajanujam, Caitlin McDonald, and Jeremy Liegey for assisting with data collection.

## APPENDIX A

TABLE A.1.

Mean  $\pm$  Standard Deviation of Estimated Acoustic Voice Features Estimated From the Sustained Vowel Phonation Task, and Their Respective *P* Value From Linear Mixed Models (LMM) Comparison Between Microphone and Smartphone, and Microphone and Tablet for Features Estimated From Recordings From the Control Group (\* Indicates Significant Difference Following Bonferroni Correction)

Feature	Mean $\pm$ Standard Deviation			<i>P</i> Values (LMM)		
	Microphone	Smartphone	Tablet	Microphone Vs. Smartphone	Microphone Vs. Tablet	Smartphone Vs. Tablet
F0	166.38 $\pm$ 49.53	166.36 $\pm$ 49.54	167.10 $\pm$ 50.03	0.995	0.995	0.995
SD F0	1.11 $\pm$ 0.45	1.19 $\pm$ 0.63	1.96 $\pm$ 2.76	0.836	0.072	0.072
HNR	20.24 $\pm$ 3.25	17.44 $\pm$ 3.46	16.96 $\pm$ 3.32	<0.001 *	<0.001 *	0.028
Jitter	0.19 $\pm$ 0.06	0.20 $\pm$ 0.08	0.33 $\pm$ 0.48	0.930	0.053	0.053
RAP	0.08 $\pm$ 0.02	0.09 $\pm$ 0.04	0.17 $\pm$ 0.28	0.905	0.035	0.035
PPQ5	0.12 $\pm$ 0.03	0.13 $\pm$ 0.06	0.26 $\pm$ 0.47	0.915	0.056	0.056
DDP	0.24 $\pm$ 0.07	0.26 $\pm$ 0.12	0.52 $\pm$ 0.84	0.905	0.035	0.035
Shimmer	0.41 $\pm$ 0.25	0.54 $\pm$ 0.28	0.59 $\pm$ 0.29	<0.001 *	<0.001 *	0.108
APQ3	2.46 $\pm$ 1.54	3.21 $\pm$ 1.74	3.40 $\pm$ 1.83	<0.001 *	<0.001 *	0.286
APQ5	2.87 $\pm$ 1.75	3.77 $\pm$ 1.93	3.95 $\pm$ 1.84	<0.001 *	<0.001 *	0.312
APQ11	3.63 $\pm$ 2.03	4.78 $\pm$ 2.07	5.26 $\pm$ 2.07	<0.001 *	<0.001 *	0.017
DDA	7.37 $\pm$ 4.62	9.63 $\pm$ 5.23	10.21 $\pm$ 5.48	<0.001 *	<0.001 *	0.286

\* Significant difference.

TABLE A.2.

Mean  $\pm$  Standard Deviation of Estimated Acoustic Voice Features Estimated From the Sustained Vowel Phonation Task, and Their Respective *P* Value From Linear Mixed Models (LMM) Comparison Between Microphone and Smartphone, and Microphone and Tablet From Recordings From Participants with HD (\* Indicates Significant Difference Following Bonferroni Correction)

Feature	Mean $\pm$ Standard Deviation			<i>P</i> Values (LMM)		
	Microphone	Smartphone	Tablet	Microphone Vs Smartphone	Microphone Vs Tablet	Smartphone Vs Tablet
F0	164.76 $\pm$ 35.05	164.20 $\pm$ 33.97	164.37 $\pm$ 34.51	0.772	0.772	0.772
SD F0	12.27 $\pm$ 9.16	10.27 $\pm$ 5.41	11.89 $\pm$ 7.82	0.718	0.718	0.718
HNR	9.25 $\pm$ 2.56	10.29 $\pm$ 2.31	9.79 $\pm$ 2.38	0.002 *	0.001 *	0.574
Jitter	3.34 $\pm$ 2.89	2.50 $\pm$ 1.62	3.07 $\pm$ 2.34	0.596	0.596	0.596
RAP	1.46 $\pm$ 1.73	0.96 $\pm$ 0.97	1.26 $\pm$ 1.36	0.537	0.498	0.537
PPQ5	1.62 $\pm$ 2.48	1.06 $\pm$ 1.21	1.43 $\pm$ 1.79	0.565	0.540	0.565
DDP	4.27 $\pm$ 5.22	2.79 $\pm$ 2.90	3.73 $\pm$ 4.07	0.537	0.498	0.537
Shimmer	1.07 $\pm$ 0.25	0.94 $\pm$ 0.20	1.01 $\pm$ 0.24	0.026 *	0.003 *	0.272
APQ3	4.73 $\pm$ 1.36	3.89 $\pm$ 1.11	4.25 $\pm$ 1.28	0.001 *	<0.001 *	0.466
APQ5	5.79 $\pm$ 1.26	4.86 $\pm$ 1.05	5.23 $\pm$ 1.22	<0.001 *	<0.001 *	0.377
APQ11	5.20 $\pm$ 2.84	4.45 $\pm$ 2.33	4.63 $\pm$ 2.62	<0.001 *	<0.001 *	0.343
DDA	13.60 $\pm$ 3.86	11.20 $\pm$ 3.15	12.20 $\pm$ 3.75	0.001 *	<0.001 *	0.466

\* Significant difference.

TABLE A.3.

Mean  $\pm$  Standard Deviation of Estimated Acoustic Voice Features Estimated From the Reading Passage Task, and Their Respective *P* Value From Linear Mixed Models (LMM) Comparison Between Microphone and Smartphone, and Microphone and Tablet From the Control Group (\* Indicates Significant Difference Following Bonferroni Correction)

Feature	Mean $\pm$ Standard Deviation			<i>P</i> Value (LMM)		
	Microphone	Smartphone	Tablet	Microphone Vs Smartphone	Microphone Vs Tablet	Smartphone Vs Tablet
F0	156.28 $\pm$ 42.62	156.62 $\pm$ 42.17	155.96 $\pm$ 42.48	0.801	0.801	0.801
SD F0	9.06 $\pm$ 7.09	10.97 $\pm$ 8.49	11.78 $\pm$ 8.36	0.078	0.001*	0.099
HNR	11.05 $\pm$ 3.09	9.18 $\pm$ 2.78	8.45 $\pm$ 2.67	<0.001*	<0.001*	<0.001*
Jitter	2.11 $\pm$ 1.65	2.87 $\pm$ 2.33	3.57 $\pm$ 3.31	0.018	<0.001*	0.007
RAP	0.73 $\pm$ 1.05	1.12 $\pm$ 1.30	1.57 $\pm$ 1.97	0.035	<0.001*	0.003*
PPQ5	0.84 $\pm$ 1.20	1.19 $\pm$ 1.60	1.61 $\pm$ 2.57	0.171	0.003*	0.067
DDP	2.10 $\pm$ 3.09	3.20 $\pm$ 3.87	4.41 $\pm$ 5.78	0.044	<0.001*	0.009
Shimmer	0.90 $\pm$ 0.29	1.09 $\pm$ 0.30	1.22 $\pm$ 0.33	<0.001*	<0.001*	<0.001*
APQ3	3.49 $\pm$ 1.32	4.61 $\pm$ 1.36	5.40 $\pm$ 1.60	<0.001*	<0.001*	<0.001*
APQ5	4.49 $\pm$ 1.50	5.83 $\pm$ 1.44	6.61 $\pm$ 1.56	<0.001*	<0.001*	<0.001*
APQ11	3.96 $\pm$ 2.51	4.91 $\pm$ 3.20	5.76 $\pm$ 3.47	0.001*	<0.001*	0.002*
DDA	9.92 $\pm$ 3.65	13.26 $\pm$ 3.86	15.55 $\pm$ 4.52	<0.001*	<0.001*	<0.001*

\* Significant differences.

TABLE A.4.

Mean  $\pm$  Standard Deviation of Estimated Acoustic Voice Features Estimated From the Reading Passage Task, and Their Respective *P* Value From Linear Mixed Models (LMM) Comparison Between Microphone and Smartphone, and Microphone and Tablet From Recordings From Participants With Hd (\* Indicates Significant Difference Following Bonferroni Correction)

Feature	Mean $\pm$ Standard Deviation			<i>P</i> Values (LMM)		
	Microphone	Smartphone	Tablet	Microphone Vs Smartphone	Microphone Vs Tablet	Smartphone Vs Tablet
F0	167.24 $\pm$ 40.83	167.90 $\pm$ 40.77	167.77 $\pm$ 40.79	0.915	0.915	0.915
SD F0	4.40 $\pm$ 1.90	4.39 $\pm$ 1.19	4.13 $\pm$ 1.36	0.976	0.677	0.677
HNR	15.49 $\pm$ 4.21	11.54 $\pm$ 3.67	11.25 $\pm$ 3.93	<0.001*	<0.001*	0.506
Jitter	0.77 $\pm$ 0.22	0.95 $\pm$ 0.29	0.90 $\pm$ 0.31	0.058	0.131	0.485
RAP	0.16 $\pm$ 0.05	0.27 $\pm$ 0.11	0.27 $\pm$ 0.15	0.011	0.011	0.979
PPQ5	0.28 $\pm$ 0.10	0.39 $\pm$ 0.15	0.40 $\pm$ 0.18	0.009	0.006	0.656
DDP	0.46 $\pm$ 0.14	0.79 $\pm$ 0.33	0.80 $\pm$ 0.45	0.008	0.008	0.925
Shimmer	0.65 $\pm$ 0.21	0.93 $\pm$ 0.19	1.02 $\pm$ 0.25	<0.001*	<0.001*	0.044
APQ3	2.87 $\pm$ 1.13	4.31 $\pm$ 1.27	4.91 $\pm$ 1.72	<0.001*	<0.001*	0.036
APQ5	3.51 $\pm$ 1.44	5.45 $\pm$ 1.66	6.38 $\pm$ 2.17	<0.001*	<0.001*	0.035
APQ11	4.42 $\pm$ 1.34	6.28 $\pm$ 1.99	7.12 $\pm$ 2.88	0.066	0.015	0.349
DDA	8.33 $\pm$ 3.27	12.55 $\pm$ 3.72	14.38 $\pm$ 5.35	<0.001*	<0.001*	0.056

\* Significant difference.

TABLE A.5.

Mean  $\pm$  Standard Deviation of Estimated Acoustic Voice Features Estimated From the Syllable Repetition Task, and Their Respective *P* Value From Linear Mixed Models (LMM) Comparison Between Microphone and Smartphone, and Microphone and Tablet From Recordings From the Control Group (\* Indicates Significant Difference Following Bonferroni Correction)

Feature	Mean $\pm$ Standard Deviation			<i>P</i> Value (LMM)		
	Microphone	Smartphone	Tablet	Microphone Vs Smartphone	Microphone Vs Tablet	Smartphone Vs Tablet
F0	167.75 $\pm$ 46.02	169.50 $\pm$ 46.82	170.64 $\pm$ 46.75	0.198	0.022	0.232
SD F0	13.86 $\pm$ 18.11	15.99 $\pm$ 21.15	15.97 $\pm$ 21.57	0.199	0.199	0.969
HNR	10.08 $\pm$ 3.96	9.29 $\pm$ 3.56	8.64 $\pm$ 3.64	<0.001*	<0.001*	<0.001*
Jitter	4.71 $\pm$ 6.55	5.53 $\pm$ 6.96	6.09 $\pm$ 8.14	0.081	0.003*	0.175
RAP	2.13 $\pm$ 3.77	2.56 $\pm$ 3.78	3.00 $\pm$ 4.65	0.085	0.002*	0.085
PPQ5	2.17 $\pm$ 3.74	2.43 $\pm$ 4.05	2.54 $\pm$ 4.41	0.673	0.673	0.683
DDP	6.11 $\pm$ 10.85	7.23 $\pm$ 10.85	8.20 $\pm$ 13.28	0.179	0.014	0.179
Shimmer	0.87 $\pm$ 0.55	1.03 $\pm$ 0.56	1.20 $\pm$ 0.78	<0.001*	<0.001*	<0.001*
APQ3	3.16 $\pm$ 1.93	4.35 $\pm$ 2.23	5.03 $\pm$ 2.44	<0.001*	<0.001*	<0.001*
APQ5	2.89 $\pm$ 2.40	4.03 $\pm$ 2.92	4.65 $\pm$ 3.18	<0.001*	<0.001*	<0.001*
APQ11	1.60 $\pm$ 3.39	2.14 $\pm$ 3.79	2.52 $\pm$ 4.61	0.047	0.002*	0.188
DDA	8.46 $\pm$ 5.22	11.83 $\pm$ 6.12	13.86 $\pm$ 6.93	<0.001*	<0.001*	<0.001*

\* Significant difference.

TABLE A.6.

Mean  $\pm$  Standard Deviation of Estimated Acoustic Voice Features Estimated From the Syllable Repetition Task, and Their Respective *P* Value From Linear Mixed Models (LMM) Comparison Between Microphone and Smartphone, and Microphone and Tablet From Recordings From Participants With Hd (\* Indicates Significant Difference Following Bonferroni Correction)

Feature	Mean $\pm$ Standard Deviation			<i>P</i> Values (LMM)		
	Microphone	Smartphone	Tablet	Microphone Vs Smartphone	Microphone Vs Tablet	Smartphone Vs Tablet
F0	200.88 $\pm$ 81.26	207.59 $\pm$ 81.86	208.24 $\pm$ 82.95	0.872	0.872	0.939
SD F0	6.93 $\pm$ 4.79	7.52 $\pm$ 5.41	6.74 $\pm$ 4.80	0.747	0.747	0.747
HNR	15.20 $\pm$ 5.10	11.66 $\pm$ 4.70	11.10 $\pm$ 4.89	<0.001*	<0.001*	0.062
Jitter	1.14 $\pm$ 1.06	1.28 $\pm$ 0.88	1.30 $\pm$ 1.04	0.936	0.936	0.936
RAP	0.29 $\pm$ 0.49	0.43 $\pm$ 0.40	0.54 $\pm$ 0.63	0.385	0.245	0.385
PPQ5	0.41 $\pm$ 0.35	0.62 $\pm$ 0.51	0.61 $\pm$ 0.55	0.151	0.151	0.929
DDP	0.88 $\pm$ 1.47	1.30 $\pm$ 1.21	1.58 $\pm$ 1.90	0.450	0.295	0.450
Shimmer	0.65 $\pm$ 0.26	1.05 $\pm$ 0.41	1.18 $\pm$ 0.67	<0.001*	<0.001*	0.129
APQ3	2.77 $\pm$ 1.40	4.89 $\pm$ 2.59	6.79 $\pm$ 2.67	0.001*	<0.001*	0.138
APQ5	3.84 $\pm$ 1.86	6.40 $\pm$ 2.77	6.79 $\pm$ 2.67	<0.001*	<0.001*	0.268
APQ11	5.20 $\pm$ 3.18	8.10 $\pm$ 3.90	8.35 $\pm$ 3.86	<0.001*	<0.001*	0.714
DDA	8.28 $\pm$ 4.23	14.56 $\pm$ 7.82	17.04 $\pm$ 12.23	0.002*	<0.001*	0.131

\* Significant difference.

**TABLE A.7.**  
***P* Values for the Effect of Biological Sex on the Estimated Features (\* Indicates Significant Difference Following Bonferroni Correction)**

Feature	Sustained Vowel		Reading Passage		Syllable Repetition	
	Control	HD	Control	HD	Control	HD
F0	0.000 *	0.067	<0.001 *	<0.001 *	<0.001 *	0.089
SD F0	0.286	0.234	0.551	0.221	0.003 *	0.761
HNR	0.000 *	0.618	<0.001 *	0.061	<0.001 *	0.063
Jitter	0.074	0.262	0.011	0.311	<0.001 *	0.577
RAP	0.217	0.206	0.016	0.194	<0.001 *	0.810
PPQ5	0.156	0.254	0.025	0.183	<0.001 *	0.687
DDP	0.217	0.206	0.016	0.173	<0.001 *	0.834
Shimmer	0.018	0.855	<0.001 *	0.054	<0.001 *	0.123
APQ3	0.043	0.631	<0.001 *	0.018	0.001 *	0.082
APQ5	0.020	0.680	<0.001 *	0.076	0.783	0.090
APQ11	0.004	0.593	0.003 *	0.742	0.098	0.366
DDA	0.043	0.631	<0.001 *	0.029	0.009	0.080

**TABLE A.8.**  
***P* Values for the Effect of Age on the Estimated Features (\* Indicates Significant Difference Following Bonferroni Correction)**

Feature	Sustained Vowel		Reading Passage		Syllable Repetition	
	Control	HD	Control	HD	Control	HD
F0	0.067	0.143	0.131	0.035	0.128	0.374
SD F0	0.050	0.288	0.120	0.934	0.130	0.740
HNR	0.372	0.699	0.385	0.919	0.111	0.557
Jitter	0.066	0.237	0.055	0.822	0.252	0.679
RAP	0.182	0.167	0.046	0.689	0.223	0.803
PPQ5	0.056	0.171	0.035	0.863	0.099	0.818
DDP	0.182	0.167	0.044	0.664	0.185	0.824
Shimmer	0.439	0.780	0.697	0.632	0.744	0.342
APQ3	0.467	0.551	0.794	0.250	0.212	0.199
APQ5	0.398	0.664	0.475	0.131	0.039	0.212
APQ11	0.383	0.790	0.102	0.070	0.769	0.149
DDA	0.467	0.551	0.642	0.190	0.066	0.188



TABLE A.9.

**P Values for the Features Estimated From the Entire Signal for Sustained Vowel for Both Groups (\* Indicates Significant Difference Following Bonferroni Correction)**

Feature	Control			HD		
	Microphone Vs. Smartphone	Microphone vs. Tablet	Smartphone Vs. Tablet	Microphone Vs. Smartphone	Microphone Vs. Tablet	Smartphone Vs. Tablet
F0	0.998	0.998	0.998	0.873	0.873	0.873
SD F0	0.448	0.050	0.147	0.424	0.424	0.766
HNR	0.000 *	0.000 *	0.041	0.002 *	0.001 *	0.575
Jitter	0.397	0.037	0.141	0.514	0.514	0.734
RAP	0.363	0.030	0.137	0.434	0.376	0.615
PPQ5	0.461	0.019	0.067	0.398	0.320	0.595
DDP	0.363	0.030	0.137	0.434	0.376	0.615
Shimmer	0.000 *	0.000 *	0.124	0.004 *	<0.001 *	0.315
APQ3	0.001 *	0.000 *	0.235	<0.001 *	<0.001 *	0.617
APQ5	0.000 *	0.000 *	0.317	<0.001 *	<0.001 *	0.558
APQ11	0.000 *	0.000 *	0.047	<0.001 *	<0.001 *	0.226
DDA	0.001 *	0.000 *	0.235	<0.001 *	<0.001 *	0.617

## REFERENCES

- Hou MY, Hurwitz S. Using daily text-message reminders to improve adherence with. *Obstet Gynecol.* 2010;116:633–640. [Online]. Available at: [http://journals.lww.com/greenjournal/Abstract/2010/09000/Using\\_Daily\\_Text\\_Message\\_Reminders\\_to\\_Improve.13.aspx](http://journals.lww.com/greenjournal/Abstract/2010/09000/Using_Daily_Text_Message_Reminders_to_Improve.13.aspx).
- Reyes BA, Reljin N, Kong Y, et al. Tidal volume and instantaneous respiration rate estimation using a volumetric surrogate signal acquired via a smartphone camera. *IEEE J Biomed Heal Informatics.* 2017;21:764–777. <https://doi.org/10.1109/JBHI.2016.2532876>.
- Nam Y, Kong Y, Reyes B, et al. Monitoring of heart and breathing rates using dual cameras on a smartphone. *PLoS One.* 2016;11:1–15. <https://doi.org/10.1371/journal.pone.0151013>.
- Doheny EP, Lowery MM, Russell A, et al. Estimation of respiration rate and sleeping position using a wearable accelerometer. *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS.* 2020;2020:4668–4671. <https://doi.org/10.1109/EMBC44109.2020.9176573>.
- Ginis P, Nieuwboer A, Dorfman M, et al. Feasibility and effects of home-based smartphone-delivered automated feedback training for gait in people with Parkinson's disease: a pilot randomized controlled trial. *Parkinsonism & related disorders.* 2016;22:28–34. <https://doi.org/10.1016/j.parkreldis.2015.11.004>.
- Larson EC, Goel M, Boriello G, et al. SpiroSmart: using a microphone to measure lung function on a mobile phone. In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* 2012:280–289.
- Vatanparvar K, Nathan V, Nemat E, et al. SpeechSpiro: lung function assessment from speech pattern as an alternative to spirometry for mobile health tracking. In: *Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS.* IEEE; 2021:7237–7243. <https://doi.org/10.1109/EMBC46164.2021.9630077>.
- Horin AP, McNeely ME, Harrison EC, et al. Usability of a daily mHealth application designed to address mobility, speech and dexterity in Parkinson's disease. *Neurodegener Dis Manag.* 2019;9:97–105. <https://doi.org/10.2217/nmt-2018-0036>.
- Wilson R, Small J. Care staff perspectives on using mobile technology to support communication in long-term care: mixed methods study. *JMIR Nurs.* 2020;3:e21881. <https://doi.org/10.2196/21881>.
- Hussein WF, et al. The mobile health readiness of people receiving in-center hemodialysis and home dialysis. *Clin J Am Soc Nephrol.* 2021;16:98–106. <https://doi.org/10.2215/CJN.11690720>.
- Volkman J, Hefter H, Lange HW, et al. Impairment of temporal organization of speech in basal ganglia diseases. *Brain Lang.* 1992;43:386–399. [https://doi.org/10.1016/0093-934X\(92\)90108-Q](https://doi.org/10.1016/0093-934X(92)90108-Q).
- Scott Kelso JA, Tuller B, Harris KS. A 'dynamic pattern' perspective on the control and coordination of movement. *The Production of Speech.* 1983:137–173. [https://doi.org/10.1007/978-1-4613-8202-7\\_7](https://doi.org/10.1007/978-1-4613-8202-7_7).
- Smith A, Mcfarland DH, Weber CM. Interactions between speech and finger movements. *J Speech Lang Hear Res.* 1986;29:471–480. <https://doi.org/10.1044/jsr.2904.471>.
- Oliveira G, Fava G, Baglione M, et al. Mobile digital recording: adequacy of the iRig and IOS device for acoustic and perceptual analysis of normal voice. *J. Voice.* 2017;31:236–242. <https://doi.org/10.1016/j.jvoice.2016.05.023>.
- Vogel AP, Rosen KM, Morgan AT, et al. Comparability of modern recording devices for speech analysis: smartphone, landline, laptop, and hard disc recorder. *Folia Phoniatr Logop.* 2014;66:244–250. <https://doi.org/10.1159/000368227>.
- Jannetts S, Schaeffler F, Beck J, et al. Assessing voice health using smartphones: bias and random error of acoustic voice parameters captured by different smartphone types. *Int J Lang Commun Disord.* 2019;54:292–305. <https://doi.org/10.1111/1460-6984.12457>.
- Zhang C, Jepson K, Lohfink G, et al. Comparing acoustic analyses of speech data collected remotely. *J Acoust Soc Am.* 2021;149:3910–3916. <https://doi.org/10.1121/10.0005132>.
- Schaeffler F, Jannetts S, Beck J. Reliability of Clinical Voice Parameters Captured With Smartphones – Measurements of Added Noise and Spectral Tilt. In: *Proc. 20th Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH.* Graz, Austria: ISCA; 2019:2523–2527.
- Manfredi C, Lebacqz J, Cantarella G, et al. Smartphones offer new opportunities in clinical voice research. *J Voice.* 2017;31:111.e1–111.e7. <https://doi.org/10.1016/j.jvoice.2015.12.020>.
- Lebacqz J, Schoentgen J, Cantarella G, et al. Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research. *J Voice.* 2017;31:550–556. <https://doi.org/10.1016/j.jvoice.2017.02.017>.
- Kim G-H, Kang D-H, Lee Y-Y, et al. Recording quality of smartphone for acoustic analysis. *J Clin Otolaryngol Head Neck Surg.* 2016;27:286–294. <https://doi.org/10.35420/jcohns.2016.27.2.286>.
- Maryn Y, Ysenbaert F, Zarowski A, et al. Mobile communication devices, ambient noise, and acoustic voice measures. *J Voice.* 2017;31:248.e11–248.e23. <https://doi.org/10.1016/j.jvoice.2016.07.023>.
- van der Woerd B, Wu M, Parsa V, et al. Evaluation of acoustic analyses of voice in nonoptimized conditions. *J Speech Lang Hear Res.* 2020;63:3991–3999. [https://doi.org/10.1044/2020\\_JSLHR-20-00212](https://doi.org/10.1044/2020_JSLHR-20-00212).
- Uloza V, Ulozaitė-Staniėnė N, Petrauskas T, et al. Accuracy of acoustic voice quality index captured with a smartphone – measurements

- with added ambient noise. *J Voice*. 2021. <https://doi.org/10.1016/j.jvoice.2021.01.025>. In press.
25. Bocklet T, Steidl S, N6th E, et al. Automatic evaluation of parkinson's speech - acoustic, prosodic and voice related cues. In: *Proc Annu Conf Int Speech Commun Assoc INTERSPEECH*. 2013:1149–1153. <https://doi.org/10.21437/interspeech.2013-313>.
  26. Portnoy RA, Aronson AE. Diadochokinetic syllable rate and regularity in normal and in spastic and ataxic dysarthric subjects. *J Speech Hear Disord*. 1982;47:324–328.
  27. Orozco-Arroyave JR, V1squez-Correa JC, Klumpp P, et al. Apkinson: the smartphone application for telemonitoring Parkinson's patients through speech, gait and hands movement. *Neurodegener Dis Manag*. 2020;10:137–157. <https://doi.org/10.2217/nmt-2019-0037>.
  28. Tsanas A, Little MA, McSharry PE, et al. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng*. 2010;57:884–893. <https://doi.org/10.1109/TBME.2009.2036000>.
  29. Ruz J, Hlavnička J, Tykalova T, et al. Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. *IEEE Trans Neural Syst Rehabil Eng*. 2018;26:1495–1507. <https://doi.org/10.1109/TNSRE.2018.2851787>.
  30. Jeancolas L, Mangone G, Corvol JC, et al. Comparison of telephone recordings and professional microphone recordings for early detection of Parkinson's disease, using mel-frequency cepstral coefficients with Gaussian mixture models. *INTERSPEECH 2019: 20th Annual Conference of the International Speech Communication Association. International Speech Communication Association (ISCA)*; 2019:3033–3037.
  31. Patel RH, Rita R, Awan SN, et al. Recommended protocols for instrumental assessment of voice: American Speech- Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *Am J Speech Lang Pathol*. 2018;27:887–905. [https://doi.org/10.1044/2018\\_AJSLP-17-0009](https://doi.org/10.1044/2018_AJSLP-17-0009).
  32. Fairbanks G. The rainbow passage. *Voice Articul Drillb*. 1960;2:127.
  33. Ruz J, Tykalova T, Ramig LO, et al. Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders. *Mov Disord*. 2021;36:803–814. <https://doi.org/10.1002/mds.28465>.
  34. Kaiser JF. On a simple algorithm to calculate the 'energy' of a signal. *ICASSP IEEE Int Conf Acoust Speech Signal Process - Proc*. 1990;1:381–384. <https://doi.org/10.1109/icassp.1990.115702>.
  35. Tsanas A, Little MA, McSharry PE, et al. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *J R Soc Interface*. 2011;8:842–855. <https://doi.org/10.1098/rsif.2010.0456>.
  36. Ruz J, Saft C, Schlegel U, et al. Phonatory dysfunction as a preclinical symptom of huntington disease. *PLoS One*. 2014;9:1–7. <https://doi.org/10.1371/journal.pone.0113412>.
  37. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proc*. 1993;17:97–110. [Online]. Available at: [http://www.fon.hum.uva.nl/paul/papers/Proceedings\\_1993.pdf](http://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf).
  38. de Cheveign1 A, Kawahara H. YIN, a fundamental frequency estimator for speech and music. *J Acoust Soc Am*. 2002;111:1917–1930. <https://doi.org/10.1121/1.1458024>.
  39. Kasi K, Zahorian SA. Yet another algorithm for pitch tracking. *ICASSP IEEE Int Conf Acoust Speech Signal Process - Proc*. 2002;1:361–364. <https://doi.org/10.1109/icassp.2002.5743729>.
  40. De Cheveign1. Speech f0 extraction based on Licklider's pitch perception model. In: *Proc. ICPHS*. 1993:6. [Online]. Available at: <http://recherche.ircam.fr/equipes/pcm/cheveign/ps/ICPS91-Licklider.pdf>.
  41. McFee B, Raffel C, Liang D, et al. librosa: audio and music signal analysis in python. In: *Proc 14th Python Sci Conf*. 2015:18–24. <https://doi.org/10.25080/majora-7b98e3ed-003>.
  42. Brookes M. Voicebox: speech processing toolbox for matlab. *Software*. 1997;47:45. Available at: [Mar. 2011] from [www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html](http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html).
  43. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Int J Nurs Stud*. 2010;47:931–936. <https://doi.org/10.1016/j.ijnurstu.2009.10.001>.
  44. B Winter, "Linear models and linear mixed effects models in R with linguistic applications," arXiv Prepr. arXiv:1308.5499, 2013.
  45. D Bates, M M1achler, B Bolker, et al. "Fitting linear mixed-effects models using lme4," arXiv Prepr. arXiv:1406.5823, 2014.
  46. Uloza V, Padervinskis E, Vegiene A, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *Eur Arch Oto-Rhino-Laryngol*. 2015;272:3391–3399. <https://doi.org/10.1007/s00405-015-3708-4>.
  47. Teixeira JP, Oliveira C, Lopes C. Vocal acoustic analysis – jitter, shimmer and HNR parameters. *Proc Technol*. 2013;9:1112–1122. <https://doi.org/10.1016/j.protcy.2013.12.124>.
  48. Robin J, Harrison JE, Kaufman LD, et al. Evaluation of speech-based digital biomarkers: review and recommendations. *Digit Biomark*. 2020;4:99–108. <https://doi.org/10.1159/000510820>.
  49. Guidi A, Salvi S, Ottaviano M, et al. Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: system performance evaluation and case study. *Sensors (Switzerland)*. 2015;15:28070–28087. <https://doi.org/10.3390/s151128070>.
  50. Nittrouer S, McGowan RS, Milenkovic PH, et al. Acoustic measurements of men's and women's voices. *J Speech Lang Hear Res*. 1990;33:761–775. <https://doi.org/10.1044/jshr.3304.761>.
  51. Lin E, Hornibrook J, Ormond T. Evaluating iPhone recordings for acoustic voice assessment. *Folia Phoniatr Logop*. 2012;64:122–130. <https://doi.org/10.1159/000335874>.