JASA ARTICLE

Impact of naturalistic smartphone positioning on acoustic measures of voice[a)]

Victoria S. McKenna,[1,b)] Rachel M. Roberts,[1] Aaron D. Friedman,[2] Savannah N. Shanley,[1] and Andres F. Llico[3]
[1]Department of Communication Sciences and Disorders, University of Cincinnati, Cincinnati, Ohio 45267, USA
[2]Department of Otolaryngology–Head and Neck Surgery, University of Cincinnati, Cincinnati, Ohio 45267, USA
[3]Department of Biomedical Engineering, University of Cincinnati, Cincinnati, Ohio 45221, USA

ABSTRACT:
Smartphone technology has been used for at-home health monitoring, but there are few available applications (apps) for tracking acoustic measures of voice for those with chronic voice problems. Current apps limit the user by restricting the range of smartphone positions to those that are unnatural and non-interactive. Therefore, we aimed to understand how more natural smartphone positions impacted the accuracy of acoustic measures in comparison to clinically acquired and derived measures. Fifty-six adults (11 vocally healthy, 45 voice disordered, aged 18–80 years) completed voice recordings while holding their smartphones in four different positions (e.g., as if reading from the phone, up to the ear, etc.) while a head-mounted high-quality microphone attached to a handheld acoustic recorder simultaneously captured voice recordings. Comparisons revealed that mean fundamental frequency (Hz), maximum phonation time (s), and cepstral peak prominence (CPP; dB) were not impacted by phone position; however, CPP was significantly lower on smartphone recordings than handheld recordings. Spectral measures (low-to-high spectral ratio, harmonics-to-noise ratio) were impacted by the phone position and the recording device. These results indicate that more natural phone positions can be used to capture specific voice measures, but not all are directly comparable to clinically derived values.
https://doi.org/10.1121/10.0020176

(Received 21 February 2023; revised 21 May 2023; accepted 29 June 2023; published online 14 July 2023)
[Editor: Benjamin V. Tucker] Pages: 323–333

I. INTRODUCTION

Nearly 1/3 of the United States population will encounter voice problems (dysphonia) at some point in their lives (Roy *et al.*, 2004; Roy *et al.*, 2005). Dysphonia is characterized by a variety of symptoms, including hoarseness, throat clearing, fatigue, discomfort, and difficulty communicating (Misono *et al.*, 2014). All of these obstacles can contribute to psychosocial distress and, in many scenarios, can even impact the individual's ability to meet the demands of their occupation (de Medeiros *et al.*, 2012; Misono *et al.*, 2014).

The evaluation and treatment of voice disorders often require multiple healthcare visits over several weeks (Wood *et al.*, 2014), months (Johns, 2003; Krause *et al.*, 2022), or years (Simonyan *et al.*, 2021; Yang *et al.*, 2022), depending on the patient's diagnosis. The general course of intervention involves an initial office-based instrumental assessment (e.g., laryngeal stroboscopy) for diagnosis and then repeated office visits to assess progression and/or receive treatment, such as voice therapy. Therefore, clinical monitoring of voice progress (or deterioration) is necessary to determine an appropriate clinical intervention schedule and guide clinical decisions. However, there are few ways for patients and clinicians to determine optimal treatment follow-up schedules; thus, they rely on pre-selected follow-up timeframes (e.g., 3 months) or patient-initiated follow-up for a self-perceived change in their vocal baseline. Patient reports of vocal severity are only, at best, moderately related to more objective measures of voice (e.g., voice acoustics, aerodynamics; Awan *et al.*, 2014; Gillespie *et al.*, 2014) and only provide a partial picture of the patient profile. This can lead to either unnecessary follow-up visits or delayed intervention by waiting too long for clinical treatment. Not only do unnecessary visits increase healthcare costs (Berwick and Hackbarth, 2012), but also delayed clinical intervention can lead to unwanted disease progression and potentially more dire health consequences (e.g., airway obstruction; Costa *et al.*, 2020).

At-home voice monitoring that includes acoustic measures of voice may provide more information to patients and clinicians when determining office follow-up and/or intervention timing. In recent years, healthcare has transformed to involve remote health monitoring to provide more patient-centric care from home (Layfield *et al.*, 2020; Majumder and Deen, 2019). Digital health resources can support patients who reside in remote locations (Jennett *et al.*, 2003; Layfield *et al.*, 2020; Philips *et al.*, 2019; Wolfe *et al.*, 2020), are susceptible to infection (Hakim *et al.*, 2020; Layfield *et al.*, 2020), or lack

[a)]Portions of this work were presented at the Fall Voice Conference, San Francisco, CA, USA, October 2022 and the 183rd Meeting of the Acoustical Society of America, Nashville, TN, USA, December 2022.
[b)]Also at: Department of Otolaryngology–Head and Neck Surgery, University of Cincinnati, Cincinnati, OH 45267, USA. Electronic mail: mckennvs@ucmail.uc.edu

the ability to pay for healthcare visits (Jennett *et al.*, 2003; Philips *et al.*, 2019), resolving obstacles for a wide array of circumstances. While there has been a surge in smartphone-based health monitoring applications (apps; Wallace and Kanegaonkar, 2020), the ambulatory voice monitoring space is relatively deficient in remote tools to provide biofeedback for a patient and inform clinicians of their patient's vocal health status. Providing patients and clinicians with objective data could contribute to more accurately monitoring the progression of dysphonia and could be instrumental in addressing concerns at the appropriate time (van Leer *et al.*, 2017). Thus, there is high demand for a smartphone app that can provide immediate, user-friendly acoustic biofeedback comparable to acoustics gathered in the clinical setting. However, the variability of vocal data resulting from smartphone positioning must be assessed in parallel to gold-standard clinical techniques to ensure app acoustic accuracy.

A. Acoustic recording considerations

Clinicians follow specific microphone distance and positioning standards to accurately record measures of acoustic amplitude, frequency, quality, and duration (Patel *et al.*, 2018). Current guidelines recommend that the microphone be placed a consistent distance from the mouth (i.e., 4–10 cm for head-mounted microphones) and at an angle of 45°–90° from the lips (Patel *et al.*, 2018). Microphone-to-mouth distances are critically important because it has been well-established that distance impacts signal amplitude, in which the amplitude is attenuated when the microphone is moved further away from the sound source. Low frequencies are differentially impacted because they are attenuated at different rates depending on the distance of the microphone (i.e., proximity effect; Švec and Granqvist, 2010). On the other hand, microphone position (angle) influences how aerodynamics, such as high frequency aspiration and affrication noise, impact the acoustic signal (Price and Sataloff, 1988). Subsequently, microphone distance and placement can affect amplitude- and spectral-based voice measures like vocal intensity, harmonics-to-noise ratio (HNR), and low-to-high (L/H) spectral ratio.

Decisions about the appropriate location of the microphone may depend on the acoustic recording environment. Clinical voice recordings are conventionally acquired in a quiet room or sound booth to ensure an adequate signal-to-noise ratio (SNR; Bottalico *et al.*, 2020; Švec and Granqvist, 2010), calculated as the signal intensity minus the noise intensity (Maryn and Zarowski, 2015). Current guidelines recommend that the acoustic signal be a minimum of 15 dB higher than the background noise (Patel *et al.*, 2018) but preferentially 30 dB higher to ensure an accurate extraction of acoustic perturbation measures and 42 dB higher to reach 99% accuracy in measure detection (Deliyski *et al.*, 2005). Subsequently, microphone distance can be leveraged as a strategy to improve signal amplitude and ensure an appropriate SNR. An adequate SNR can be accomplished at microphone distances of 30 cm as long as there is a relatively low level of background noise (i.e., <40 dBA; Švec and Granqvist, 2010). Reducing the microphone-to-mouth distance to 10 cm would continue to allow for an adequate SNR in louder environments in which background noise reaches up to 50 dBA (Dejonckere *et al.*, 2001).

Replicating these noise and distance standards during remote acoustic acquisition on personal smartphones is important for reliable acoustic estimation. A study by Maryn *et al.* (2017) investigated the impact of background noise levels on common clinical acoustic measures of voice across different smartphone platforms (i.e., iOS, Android, Windows). Pre-recorded voice samples were played over a loudspeaker at a fixed distance of 10 cm and 45° angle from the smartphone's microphone in a sound booth, with various levels of additive noise. Results showed that fundamental frequency (f_o) was the most resistant to background noise across all platforms. Conversely, measures of cepstral peak prominence (CPP) and HNR were significantly lower when background noise levels were greater than 47 and 56 dBA, respectively, compared to values obtained with a background noise level of 20.5 dBA.

In addition to accounting for background noise, it may be difficult to control the location and distance of the microphone from the speaker in the home setting. To date, there are no published guidelines for smartphone placement during voice recordings, resulting in variable recommendations for phone holding positions. An app by van Leer *et al.* (2017) calculated CPP, f_o, and jitter during home voice practice. The researchers instructed the participants to hold the phone approximately 20 cm from the mouth during tasks. However, information regarding how participants were able to measure and maintain this distance at home was not described. Grillo *et al.* (2016) examined differences between smartphone recording devices and a stationary head-mounted microphone, each placed 4 cm from the mouth (with smartphones parallel to the floor) in 20 vocally healthy adults. They reported that this smartphone position resulted in values comparable to those acquired on the stationary microphone for frequency, perturbation (e.g., jitter), and quality measures. Current directions for their app, known as VoiceEvalU8, recommend using a 4 cm measuring stick to facilitate a consistent phone distance and placement directly in front of the mouth.

Close microphone distances, such as 4 cm, reduce the naturalistic component of smartphone voice recordings and may be unrealistic for patients to regularly achieve outside of the clinic. Holding the phone at a comfortable distance and angle enables users to view the screen while using the recording app. In this way, patients could interact with the screen and view speech stimuli (such as when asked to speak a standardized reading passage), instead of needing to memorize or read the sentences from a separate source during recordings. Moreover, fixed-distance acquisition assumes a steady hand and ability to maintain a specific posture over the course of a recording. These physical demands may not be possible for individuals with neurological or motor impairments, such as those with essential tremor and

Parkinson's disease. Therefore, it is critical to understand how patients interact with and hold their phones to ensure adequate reliability when there is likely user variation and error.

B. Purpose

To attain reliable acoustic estimations of voice from smartphone recordings, more information is needed on user variation during typical smartphone interactions. Therefore, this project aimed to evaluate how variation in phone holding position impacted specific voice measures (e.g., vocal pitch, quality, and duration) on a smartphone compared to traditional, commercially available clinical tools. We sought to describe typical phone holding patterns across smartphone users and discuss their influence on acoustic measures. Based on previous research, we hypothesized that pitch and duration parameters [i.e., mean f_o, maximum phonation time (MPT)] would be impervious to smartphone positioning, as these are time-based parameters that should not be impacted by microphone distance and position. Conversely, we expected that acoustic measures dependent upon amplitude, spectral shape, and noise (i.e., HNR, L/H ratio, CPP) would be impacted by phone position due to known effects of distance and position on the acoustic signal.

II. METHOD

A. Participants

Participants were recruited prospectively from January to June 2022 from the Robin Cotton and Rocco dal Vera Professional Voice, Swallowing, and Airway Center at the University of Cincinnati Medical Center and the Voice and Swallow Mechanics Lab at the University of Cincinnati. Institutional Review Board (IRB) approval was acquired prior to enrollment (IRB#2021-0724), and all participants provided informed consent prior to participation.

A total of 56 participants completed the study protocol [28 cisgender women, 2 transgender women, 26 cisgender men; aged 18–80 years; mean (M) = 53.8 years, standard deviation (SD) = 19.2 years]. Of these, 11 participants were vocally healthy, and 45 participants were diagnosed with a voice disorder by a laryngologist through a formal clinical visit. A range of voice diagnoses were captured, including 20 with neurogenic disorders (e.g., vocal tremor, laryngeal dystonia), 12 with benign lesions (e.g., vocal fold nodules), 7 with neoplastic lesions (e.g., laryngeal cancer, recurrent respiratory papillomatosis), and 6 with functional disorders (e.g., muscle tension dysphonia).

B. Voice questionnaire and dysphonia ratings

Participants self-reported the degree of their voice impairment via the Voice-Related Quality of Life (V-RQOL; Hogikyan and Sethuraman, 1999). This standardized, validated ten-item questionnaire assesses the impact of voice problems on daily quality of life using a Likert scale. A statement rating of 1 indicated that the statement is "not a problem," and a rating of 5 indicated the "problem is as bad as it can be." The sum of the questions (out of a total score of 50) is then transformed to a 0–100 scale, in which a score of 100 indicates no impact of voice problems on daily life, and an overall score of 0 indicates the greatest possible problem.

A voice-specialized speech-language pathologist (SLP) with 5 years of clinical experience blindly completed auditory-perceptual ratings using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster *et al.*, 2009). Voice recordings were played via over-the-ear headphones (Sennheiser, 280 Pro) set to a comfortable listening level that the SLP could adjust, as needed. The SLP rated overall dysphonia, defined as the "global, integrated impression of voice deviance" (Kempster *et al.*, 2009), by placing a single mark on a 100-mm visual-analog scale. A rating toward the left side of the scale (0) indicated no vocal deviance, whereas a score toward the right side of the scale (100) indicated severe vocal deviance. The mark was measured with a ruler to the nearest mm, and the final rating was placed in an Excel sheet.

C. Procedure

Participants were seated in a comfortable position in a quiet room. The background noise of the room was measured via a sound pressure level (SPL) meter, showing an average sound level of 37.4 dBA. Participants were then fitted with a headset microphone (C555L, Shure, Niles, IL) placed 8.5 cm from the center of the lips, 45° from midline. The headset microphone was attached to a handheld recorder (H4N, Zoom, Tokyo, Japan) to acquire a high-quality acoustic signal at a sampling rate of 44.1 kHz and 16 bits. The headset microphone was calibrated to dB SPL by playing a pure tone of 1000 Hz from a smartphone app (Frequency Generator for Android; Sonic for iOS) placed at the lips while dB SPL was measured at the headset microphone via a sound level meter.

Participants then downloaded a voice recording app on their personal smartphone[1] (Voice Recorder or RecForge II for Android; Voice Recorder Pro for iOS). Settings of the smartphone app were adjusted to sample at a minimum of 22 050 Hz[2] and output acoustics in a waveform audio file format (.wav). This format was chosen to mitigate the compression that can occur during smartphone audio acquisition and offloading of files (Cavalcanti *et al.*, 2023; Vogel and Morgan, 2009). Then simultaneous recordings were made on the personal smartphone and the handheld recorder to directly compare the two recorded acoustic signals.

To understand how phone position impacted voice acoustics, four positions were assessed (Fig. 1): Ear, holding the phone to the ear, as if speaking on the phone [Fig. 1(A)]; Reading, holding the phone in front of the face, as if reading from it [Fig. 1(B)]; Speaker phone, holding the phone as if on speaker phone [Fig. 1(C)]; Controlled distance, holding the phone 8.5 cm directly in front of the mouth, parallel to the floor [Fig. 1(D)]. The distance of 8.5 cm was chosen because this is the size of a typical identification or credit

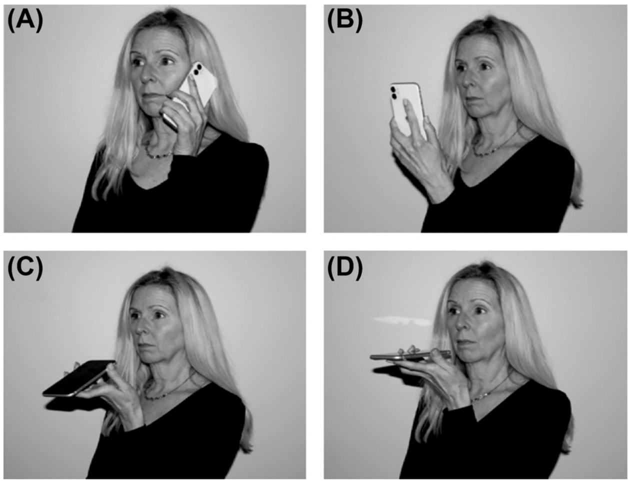

FIG. 1. Phone positions. (A) Ear position. (B) Reading position. (C) Speaker phone position. (D) Controlled distance position of 8.5 cm from the lips, parallel to the ground.

card and would be an item a patient could use to measure the mouth-to-phone distance outside of the clinic. The position of the phone directly in front of the mouth was used because this is the current recommendation from Grillo *et al*. (2016). The distance from the participant's mouth to the phone was also measured for the Reading and Speaker phone positions to describe the distances each speaker naturally used for these positions. All participants completed recordings in all four positions, but the positions were randomized for each speaker to mitigate any order effects.

Speech tasks consistent with those completed during standard voice evaluations were completed in each phone position. These tasks included (i) sustained vowel /ɑ/ for 5 s (referred to as "vowel," here forward); (ii) sustained vowel /ɑ/ for as long as possible (referred to as "maximum phonation"); and (iii) the second and third sentences of the rainbow passage (referred to as "read speech").

D. Acoustic extraction

Voice acoustic measures were chosen based on current recommendations for acoustic measures for voice evaluation (Patel *et al.*, 2018) and the standard practices of the University of Cincinnati Professional Voice, Swallowing, and Airway Center. The measures included mean f_o (Hz), MPT (s), HNR (dB), CPP (dB), and L/H spectral ratio (dB). See Table I for a complete list of measures extracted from each speech task.

Vocal intensity was not a target smartphone measure in this study because it has already been established that microphone distance and directionality impact signal amplitude (Švec and Granqvist, 2010). However, we did extract vocal intensity (dB SPL) from the handheld recordings, to ascertain whether participants altered their vocality intensity when the phone position changed. That is, we measured vocal intensity to verify that participants were not speaking louder because the phone was held further away.

TABLE I. Acoustic measures extracted for each speech task.

Speech task	Acoustic processing software	Acoustic measure
Sustained vowel /ɑ/ × 5 s	PRAAT	Mean f_o HNR
	ADSV	CPP L/H ratio
Sustained vowel /ɑ/ for as long as possible	PRAAT	Mean f_o HNR MPT
Second and third sentences of the rainbow passage	PRAAT	Mean f_o HNR
	ADSV	CPP L/H ratio

1. Mean f_o, MPT, HNR, and vocal intensity

Measures of mean f_o, MPT, HNR, and vocal intensity were determined via PRAAT (Boersma, 2001), a free, open source software used in research and clinical practice. Two trained technicians met training criteria of inter-rater reliability standards [intraclass correlation coefficient (ICC) > 0.90] on targeted acoustic measures, prior to extracting experimental data. Each technician extracted acoustic measures from half of the participants. Acoustics were extracted from both the smartphone recordings and from the handheld recordings, except for intensity that was only measured for the handheld recordings.

First, pitch ranges were set based on the gender of the participant. For women, a range of 90–500 Hz was used, whereas for men, a lower range of 60–300 Hz was applied. Rarely, pitch settings needed to be adjusted to ensure accurate pitch estimation and prevent aliasing. Next, technicians followed a protocol for extraction for each speech task. For sustained vowels, the mid-portion of the vowel was determined through visual inspection of the waveform and spectrogram. The mid-portion (approximately 3 s in the 5-s sample) was chosen to reflect vowel steady-state and to eliminate phonatory onset/offset behaviors. For maximum phonation, the entire vowel segment was selected to measure MPT and extract additional pitch, quality, and intensity measures of mean f_o, HNR, and vocal intensity, respectively. Finally, the entirety of the two rainbow sentences was selected to extract acoustic measures. The "voice report" function was used to determine mean f_o and HNR, whereas the highlighting time-measurement capability of PRAAT was employed to measure MPT. The "get intensity" function was used to measure vocal intensity. All vocal intensity measures were further corrected using the known intensity measured with the sound level meter during the calibration procedure described above.

Following initial data extraction, inter- and intra-rater reliability was determined on targeted acoustic measures

https://doi.org/10.1121/10.0020176

from ∼10% of samples (six participants) more than 2 months later. Technicians were blinded to previous extractions. Inter-rater reliability was deemed "excellent" across measures (Koo and Li, 2016), with ICC = 0.98, 0.99, and 0.98 for mean f_o, HNR, and MPT, respectively. Intra-rater reliability was determined via Pearson correlation coefficients for each measure and rater, with "excellent" reliability as well (average $r = 0.98$, range = 0.91–0.99).

2. CPP and L/H ratio

Analysis of Dysphonia in Speech and Voice (ADSV; version 4.0), a proprietary software used in research and clinical practice, was used to extract measures of CPP and L/H ratio from the sustained vowel and read speech tasks. First, settings were adjusted to reflect the gender of the speaker, in which a frequency range of 90–500 Hz was used for women and a range of 60–300 Hz was used for men. Default ADSV settings were applied, including a spectral window size of 1024, frame overlap of 75%, and seven-frame averaging. Data were downsampled to 22 050 Hz, as opposed to the default value of 25 000 Hz. The L/H cutoff ratio was set to 4000 Hz (Hillenbrand *et al.*, 1994), and vocalic detection was turned "on."

Next, the ADSV "Sustained Vowel" protocol option was applied when extracting CPP and L/H ratio from the vowel productions, and the "Rainbow Passage" protocol option was applied when extracting the same measures from the read speech task. Like extraction guidelines for PRAAT, the mid-portion of the vowel (i.e., middle 3-s) was identified by a trained technician. The entire read speech task was selected and used for the analysis.

Following data extraction, 10% of samples (six participants) were blindly re-extracted for reliability purposes. Inter-rater reliability was deemed "excellent" (Koo and Li, 2016) with CPP ICC = 0.98 and L/H ratio ICC = 0.99. Intra-rater reliability was also "excellent," with an average $r = 0.97$ (range = 0.95–0.99) across raters for the two measures.

E. Statistical plan

1. Participant and sample characteristics

Summary data were provided to describe the sample, including average V-RQOL scores and CAPE-V dysphonia ratings. Two-sample *t*-test comparisons were made between Android-based and iPhone users for the continuous variables of age, V-RQOL scores, and CAPE-V dysphonia ratings. A paired *t*-test was calculated to compare mouth-to-phone distances for the Reading and Speaker phone positions. Finally, a repeated measures analysis of variance (ANOVA) was employed to assess whether vocal intensity varied by phone position (Ear, Reading, Speaker phone, Controlled distance), task (vowel, maximum phonation, read speech), and their interaction (phone position × task). Effect sizes were calculated as partial eta square (η_p^2) and interpreted based on the recommendations of Cohen (Witte and Witte, 2017). *Post hoc* analyses were completed using Tukey's test for paired comparisons. Significance was set to $p < 0.05$ for each model and comparison. All statistics were completed in Minitab (version 20.2).

2. Acoustical analyses

Mixed-effects models were analyzed for each acoustic measure. Fixed effects were recording device (smartphone, handheld Zoom recorder), phone position (Ear, Reading, Speaker phone, Controlled distance), and speech task (vowel, maximum phonation, read speech), with a random effect of "participant." Two-way interactions were analyzed for fixed effects. The continuous covariates of background noise and CAPE-V dysphonia rating were accounted for in each model. Significance was set to $p < 0.05$ for all variables, and each model coefficient of determination was reported (R^2_{adj}). If indicated, Tukey's *post hoc* analyses were completed with significance again at $p < 0.05$, as Tukey's test adjusts for family-wise error at the time of the analysis.

All models were checked for model assumption (e.g., normality of residuals), and no data transformations or adjustments were required, except for the variable MPT. MPT required a square root transformation to meet normality criteria (Anderson–Darling test, $p > 0.05$). Models were further adjusted based on the specific acoustic measure analyzed. For example, MPT was only measured for the maximum phonation task, so the fixed variable "speech task" was not included in that model as there were no MPT measurements for sustained vowel or read speech. Likewise, CPP and L/H ratio were only extracted from the vowel and read speech tasks, so in these models, there were only two levels for the fixed variable "speech task."

3. Missing data

Mean f_o and HNR could not be determined from four participants because PRAAT's pulse function was unable to consistently track vocal cycles. Therefore, these analyses only include 52 participants. These four participants each had overall dysphonia ratings of 100 mm, indicating that they represented some of the most severe voices in our sample. CPP and L/H ratio were not able to be calculated in three instances (two from the phone and one from the handheld). One participant with instances of missing CPP and L/H ratio was the same who was also missing mean f_o and HNR (dysphonia rating of 100 mm), whereas the other had to do with an error in that particular recording (dysphonia rating of 4 mm). There were no missing MPT values.

III. RESULTS

A. Participant characteristics

Participants reported a wide range of V-RQOL scores (0–100), with an average score of 67.3 (SD = 27.9). The overall dysphonia ratings on the CAPE-V were an average of 57.5 mm, (SD = 36.9 mm, range = 0–100 mm). A total of 12 participants had typical voices, 6 had mild dysphonia, 11 had moderate dysphonia, and 27 had severe dysphonia, as classified via the severity rating labels on the CAPE-V form.

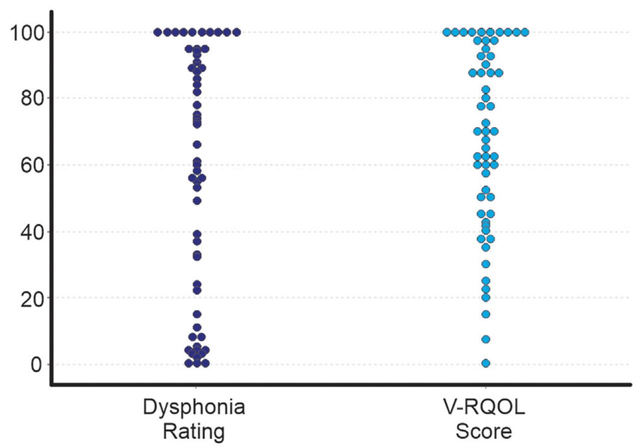

FIG. 2. (Color online) Distribution plot of dysphonia ratings and V-RQOL scores across all participants.

See Fig. 2 for a distribution plot of V-RQOL and auditory-perceptual ratings.

A total of 33 participants had iPhones, and 23 participants had Android-based phones. Participants with Android-based phones represented a significantly older demographic with an average age of 61.0 years, compared to those with iPhones, who were an average age of 48.7 years ($p = 0.010$). Participants with Android-based phones also had significantly more severe dysphonia ratings ($p = 0.022$) and trended toward having lower V-RQOL scores compared to those with iPhones, although this was not significant (Table II).

In general, the Reading position had lower mouth-to-phone distance variability and range (M = 27.7 cm, SD = 5.9 cm, range = 15–43 cm) compared to the Speaker phone position (M = 28.8 cm, SD = 11.6 cm, range = 13–61 cm). However, the average distances were not statistically different from one another ($p = 0.524$).

A repeated measures ANOVA was calculated to understand how vocal intensity may have changed by phone position and task. All measures were taken from the stationary headset microphone. Results showed a significant impact of phone position ($p < 0.001$; $\eta_p^2 = 0.03$, small effect size) and speech task ($p < 0.001$; $\eta_p^2 = 0.21$, medium-to-large effect size) on vocal intensity, but no interaction effect ($p = 0.837$). Tukey's *post hoc* analysis revealed that the Ear (86.76 dB SPL) and Controlled distance (87.36 dB SPL) positions were not significantly different from one another ($p = 0.213$). The Speaker phone (85.93 dB SPL), and Reading (86.56 dB SPL) positions were both significantly lower than the Controlled distance position; however, they were not significantly different from the Ear position. The read speech task resulted in significantly lower vocal intensity (84.50 dB SPL) compared to both the vowel (87.28 dB SPL) and maximum phonation (87.14 dB SPL) tasks.

TABLE II. Participant demographics by phone type. *, significant differences between Android and iPhone users.

Factor	Android-based mean (range)	iPhone mean (range)	*p*-value
Age (years)	61.0 (34–79)	48.7 (18–80)	0.010*
Dysphonia rating (mm)	70.7 (0–100)	48.3 (0–100)	0.022*
V-RQOL score	60.1 (0–100)	72.3 (25–100)	0.132

B. Mean f_o

The mixed-effects model showed no significant impact of recording device ($p = 0.669$) or phone position ($p = 0.518$) on mean f_o; however, the main effect of speech task was significant ($p < 0.001$) with *post hoc* testing showing that the mean f_o of the read speech task was significantly lower (154.81 Hz), compared to both vowel (161.38 Hz) and maximum phonation (162.69 Hz). No significant interaction effects were found, and the covariates of background noise and CAPE-V rating were not significant as well. The variance accounted for by all variables was 88% ($R^2_{adj} = 0.88$). See Table III for acoustic averages by recording device and speech task.

C. MPT

There was no impact of recording device ($p = 0.931$) or phone position ($p = 0.332$) on MPT, as well as no interaction effect ($p = 0.962$). Background noise did not significantly impact MPT ($p = 0.624$), but CAPE-V rating was a significant covariate ($p = 0.012$), in which a higher CAPE-V rating resulted in lower MPT ($\beta = -0.008$). The R^2_{adj} of the model was 0.93.

D. Spectral measures: HNR and L/H ratio

HNR was impacted by the variables of recording device ($p < 0.001$), phone position ($p = 0.001$), and speech task ($p < 0.001$). Interaction effects of recording device × phone position ($p = 0.010$) and recording device × speech task ($p = 0.033$) were also significant (Fig. 3); however, phone position × speech task was not ($p = 0.892$). Background noise was not a significant covariate ($p = 0.284$), but CAPE-V rating was significant ($p < 0.001$), with higher CAPE-V ratings leading to lower HNR values ($\beta = -0.108$). Variables accounted for 79% of the variance in the model ($R^2_{adj} = 0.79$).

Post hoc analysis of interaction effects showed that all handheld recordings resulted in significantly higher HNR compared to all smartphone recording positions; however, within the smartphone recording positions, the Ear position (14.02 dB) was significantly higher than the Reading (12.50 dB) and Speaker phone positions (12.29 dB) but the same as the Controlled distance (13.14 dB). HNRs from both the vowel and maximum phonation tasks were significantly greater for the recordings captured on the handheld compared to the smartphone (all $p < 0.001$). HNR for the read speech task was lower compared to all other tasks, but differentially, with handheld HNR from read speech (12.52 dB) still being significantly greater than smartphone HNR read speech (9.88 dB).

L/H ratio was also impacted by recording device ($p < 0.001$), phone position ($p < 0.001$), and speech task ($p = 0.001$). Further interaction effects of recording device

https://doi.org/10.1121/10.0020176

TABLE III. Means (SD) for each acoustic measure by phone position, recording device, and speech task.

		Position							
		Ear		Reading		Speaker phone		Controlled distance	
Acoustic measure	Speech task	Phone	Handheld	Phone	Handheld	Phone	Handheld	Phone	Handheld
Mean f_o (Hz)	Vowel	162.32 (42.14)	163.37 (43.56)	161.86 (45.57)	162.49 (42.85)	158.81 (43.85)	160.50 (43.54)	161.86 (41.95)	161.61 (41.70)
	Maximum phonation	163.58 (93.94)	162.02 (39.73)	162.18 (42.14)	161.95 (42.37)	162.46 (48.31)	163.83 (45.02)	164.27 (40.26)	164.23 (40.75)
	Read speech	154.38 (37.25)	153.50 (88.53)	155.40 (38.59)	156.36 (36.89)	155.52 (37.50)	155.71 (36.66)	155.11 (37.53)	154.29 (37.19)
MPT (s)	Maximum phonation	9.49 (5.96)	9.44 (5.97)	9.55 (5.59)	9.51 (5.61)	9.52 (5.46)	9.61 (5.37)	9.31 (5.53)	9.32 (5.51)
HNR (dB)	Vowel	15.74 (7.84)	18.52 (7.78)	13.95 (7.07)	18.35 (8.35)	13.69 (6.88)	17.74 (7.74)	14.66 (7.60)	18.11 (7.88)
	Maximum phonation	15.72 (7.40)	18.16 (7.69)	14.17 (6.79)	18.82 (7.75)	13.90 (7.23)	18.28 (8.23)	14.73 (7.43)	17.87 (7.82)
	Read speech	10.70 (3.38)	12.67 (3.61)	9.382 (3.07)	12.483 (3.72)	9.27 (3.00)	12.30 (3.73)	10.16 (3.26)	12.61 (3.54)
L/H ratio (dB)	Vowel	35.07 (10.63)	37.41 (7.02)	31.46 (8.75)	37.20 (6.50)	29.79 (8.75)	36.10 (6.83)	28.68 (8.97)	36.45 (6.81)
	Read speech	31.92 (8.25)	36.01 (5.76)	30.04 (7.69)	35.94 (5.62)	30.08 (5.76)	35.97 (5.61)	28.38 (7.40)	35.72 (5.59)
CPP (dB)	Vowel	8.52 (4.27)	9.58 (4.04)	8.56 (4.36)	9.65 (4.51)	8.16 (4.36)	9.30 (4.58)	8.81 (4.50)	9.62 (4.51)
	Read speech	4.89 (2.18)	6.34 (2.71)	4.79 (1.99)	6.27 (2.66)	4.71 (2.09)	6.31 (2.86)	5.01 (2.16)	6.45 (2.72)

× phone position ($p < 0.001$) and phone position × speech task ($p = 0.034$) were found, but there was no significant interaction effect of recording device × speech task ($p = 0.654$). Background noise was not a significant covariate ($p = 0.567$), whereas CAPE-V rating was significant ($p < 0.001$). Once again, results showed that as CAPE-V ratings increased, L/H ratio decreased ($\beta = -0.082$). The variance accounted for in this model was 76% ($R^2_{adj} = 0.76$).

Similar to HNR results, the L/H ratios from handheld recordings were significantly higher than those from smartphone recordings. The Ear position showed the highest L/H ratio values (33.54 dB), and the Controlled distance showed the lowest values (28.53 dB) for the smartphone [Fig. 4(A)]. *Post hoc* analysis of phone position × speech task showed L/H ratio from the Ear position for the vowel was significantly higher than all other conditions. Controlled distance read speech was significantly lower than Ear position read speech and Reading position vowel [Fig. 4(B)].

E. CPP

CPP was significantly different between recording devices ($p < 0.001$), with smartphone CPP being lower (6.68 dB) than CPP obtained from the handheld (7.93 dB). CPP was also significantly lower by speech task ($p < 0.001$), with CPP from read speech (5.61 dB) being lower than vowel CPP (9.00 dB). These findings drove an interaction effect of recording device × speech task ($p = 0.046$). *Post hoc* testing of the interaction effect showed that CPP was significantly different across each condition. That is, the vowel task on the handheld (9.51 dB), the vowel task on the smartphone (8.49 dB), the read speech task on the handheld (6.34 dB), and the read speech task on the smartphone (4.87 dB) showed significantly different CPPs from one another. There was no impact of phone position ($p = 0.169$) or background noise ($p = 0.891$) on CPP measures, as well as no other two-way interactions. CAPE-V rating was a significant covariate in the model ($p < 0.001$), with higher CAPE-V ratings resulting in lower CPP values ($\beta = -0.064$). The variables accounted for 83% of the variance in CPP ($R^2_{adj} = 0.83$).

1. Secondary analysis

To further investigate the impact of recording device on CPP, a mixed linear regression model between all values obtained from the phone and handheld were analyzed (with "participant" as a random factor), resulting in a model

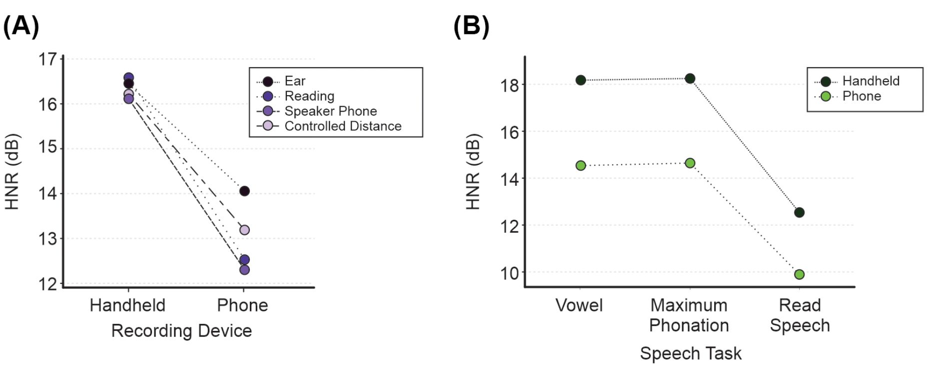

FIG. 3. (Color online) Interaction plots for HNR. (A) Interaction of recording device × phone position. (B) Interaction of recording device × speech task.

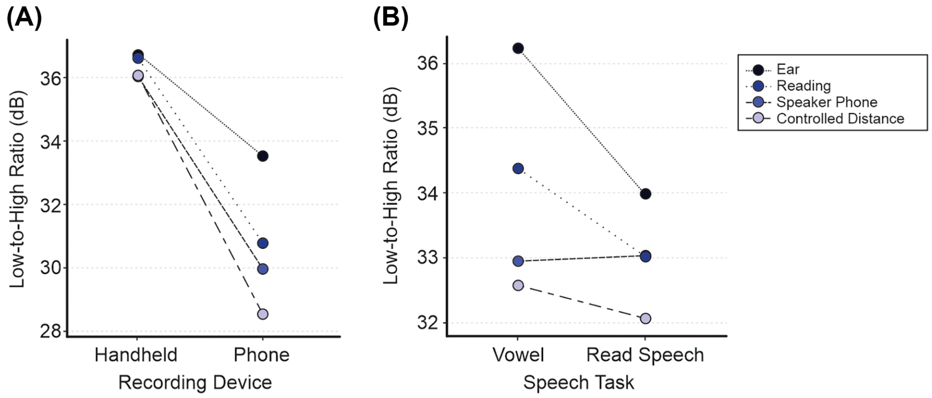

FIG. 4. (Color online) Interaction plots for L/H spectral ratio. (A) Interaction of recording device × phone position. (B) Interaction of phone position × speech task.

$R^2_{\mathrm{adj}} = 0.97$, as demonstrated in Fig. 5. A line of best fit equation was calculated for all data, revealing a 1.29 dB offset.

IV. DISCUSSION

We evaluated the impact of smartphone position on acoustic measures of voice, with a focus on more naturalistic positions that are often used in daily phone communication. We examined five acoustic measures that provide important information about vocal pitch, duration, and quality that are commonly monitored during clinical voice evaluation and treatment. Our results found that three of the five measures were not impacted by phone position and may be appropriate for interactive apps and at-home health monitoring.

First, consistent with our hypothesis and previous research (Maryn *et al.*, 2017), mean f_o and MPT were robust measures that were not impacted by phone position. This was expected due to the time-based nature of the measures and the relatively quiet rooms that the recordings were captured in. However, contrary to our hypothesis, we also determined that CPP was not impacted by phone position either. This was unexpected due to the known effects of vocal intensity on CPP, in which speaking more loudly (Awan *et al.*, 2012) or with additional effort (Rosenthal *et al.*, 2014) has been shown to increase CPP values. We investigated the potential for vocal intensity to vary based on phone position, finding only a small amount of change (∼1 dB SPL) across positions. This small change is reflective of natural variation found in speech (Brown *et al.*, 1976), instead of purposeful changes in loudness brought on by phone position. Thus, vocal intensity was likely not a factor influencing measures across phone positions and did not impact acoustic measures in our study.

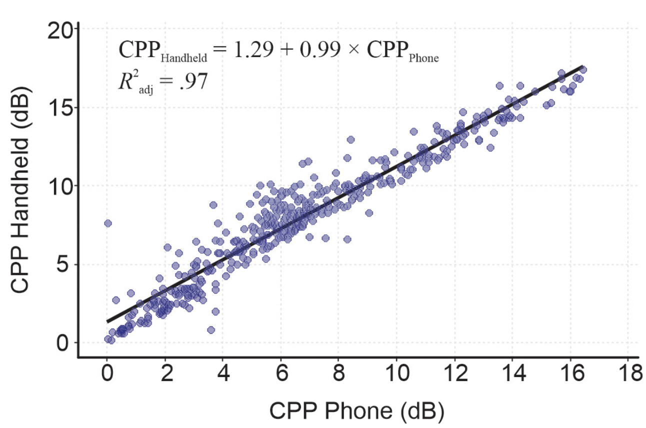

FIG. 5. (Color online) Scatter plot and line of best fit of CPP from recordings taken from the phone and handheld.

Still, the CPP results must be interpreted in the context of the smartphone vs handheld comparison, in which CPP values were still significantly lower on the smartphone. Further investigation into this relationship showed a strong correspondence between the two recording devices ($R^2_{\mathrm{adj}} = 0.97$) but an offset of ∼1.3 dB. Therefore, it would not be appropriate to compare absolute CPP between the smartphone and clinically derived values, but the *relative* change in CPP would likely be consistent. Comparing relative CPP measured over time on the same smartphone by the same user could be used to objectively assess temporal changes in vocal function. This has potential clinical implications in terms of assessing a patient's response to treatment, since CPP has been shown to be sensitive to auditory-perception of overall dysphonia (Heman-Ackah *et al.*, 2014) and breathiness (Hillenbrand *et al.*, 1994). Future studies should evaluate the test-retest reliability of CPP from the same user cued for the same holding position over several days.

Our other hypotheses were supported when both HNR and L/H ratio were impacted by phone position and recording device, increasing the likelihood of inaccuracies when using a natural smartphone position. These findings support the need for controlled mouth-to-phone distances for specific voice measures, as has been noted in the on-going work by Grillo and colleagues. The VoiceEvalU8 app requires a consistent mouth-to-microphone distance of 4 cm for all acoustic recordings and has shown that measures of pitch (e.g., mean f_o, f_o SD), perturbation (e.g., jitter, shimmer), and quality (e.g., noise-to-harmonics ratio, CPP) could be accurately recorded on smartphones as long as the same smartphone and off-app processing software were used over time (Grillo *et al.*, 2016).

To our knowledge, we are some of the first to examine naturalistic phone holding behaviors and their impact on acoustic outcomes. The average distance participants held the phone from the mouth was about 28 cm for both the Reading and the Speaker phone positions. This is consistent with previous studies that have quantified the average distance phones are held from the eyes for viewing, showing a range of 29–36 cm (Bababekova *et al.*, 2011; Boccardo, 2021; Long *et al.*, 2017). Still, previous studies have reported a large range of distances (Bababekova *et al.*, 2011; Boccardo, 2021), which we also observed here. The range for the Speaker phone position showed distances up to 61 cm, whereas the furthest distance for the Reading position was only 43 cm. Moreover, many participants also placed the phone down on the table and/or toward their laps when advised to use the Speaker phone position. Given this information, we recommend the Reading position over the Speaker phone position due to the closer, and potential for more consistent, placement in front of the face/mouth, as well as allowing for an interactive app design during voice acquisition.

With the possible range of mouth-to-phone distances in mind, the recommendation for voice acquisition in a quiet room becomes even more important for home-based monitoring. With a 30 cm mouth-to-phone distance, background noise should not exceed 40 dBA to preserve an adequate SNR for the majority of acoustic measures of voice (Švec and Granqvist, 2010). In the present study, the average background noise of the quiet room was 37.4 dBA, but the upper end of the range was 46.3 dBA, which is consistent with other reports of background noise levels in clinical rooms (Bottalico *et al.*, 2020). Although recordings were made in a quiet room, these noise levels may be of particular concern for the measurement of CPP. CPP measured via a smartphone placed 10 cm from the sound source was significantly lower at higher background noise levels (47 dBA) compared to those made in a quieter sound booth (20.5 dBA; Maryn *et al.*, 2017). When adjusted for a 30 cm mouth-to-microphone distance, as was the average distance during the naturalistic Reading position, CPP would then be significantly lower when background noise level reaches 37 dBA. This could be one explanation for the lower values noted during phone-based recordings in our study. Our preliminary investigation into background noise showed that it was not a significant factor in the CPP model; however, background noise was not controlled or evaluated at several levels within the same speaker. A more structured paradigm is needed to determine whether there is a threshold for background noise that impacts CPP. Having a smartphone app that could additionally screen for background noise and tell the user to move to quieter conditions to improve accuracy would be beneficial for users trying to record in the home setting.

A. Limitations and future directions

The major objective of this work was to determine whether more naturalistic phone positions could be appropriate for smartphone acquisition of acoustic parameters. To do so, we employed participants' own smartphones, holding positions, and natural room environments. However, utilizing realistic phone-use scenarios, which are ecologically valid, comes with limitations for experimental control. First, we were unable to assess phone type (Android-based, iPhone) due to unbalanced population distribution and symptoms between the two groups, in that those with Android-based phones were older and had more severe ratings of clinician-perceived dysphonia. Second, we could not evaluate the model of the phones due to several different phone types being owned across subjects, with different durations of ownership that may have influenced the smartphone software acquisition capabilities and/or the microphone hardware. (e.g., some phones were almost 6 years old, whereas others were newer versions). Smartphones represent rapidly developing technology that may result in different microphone specifications, amplification gain factors, and sampling rates. The ability to evaluate all combinations of smartphone specifications is limited not only due to the variety of phones on market, but also the proprietary nature of technology, where many of these specifications are not publicly available. Nevertheless, this mix of technology represents the variety of smartphones that would be expected with any app designed for clinical practice and patient use. Temporal tracking of voice quality should likely be limited to the same user and same physical phone, reducing the impact of changing technology on clinical decision making within a single patient. Third, data were collected in quiet rooms that simulated typical home recording environments, but further information is needed on louder, more controlled noise conditions during naturalistic acoustic acquisition. There may be a threshold for background noise that reduces accuracy for measures such as CPP. Further, we did not account for room reverberation time (Rakerd *et al.*, 2018) that, when combined with background noise, has negative impacts on vocal perturbation measures (Bottalico *et al.*, 2020). In summary, the above-described variables make it challenging to complete studies that can provide standardized recommendations and generalizations across all smartphone users. Therefore, researchers and clinicians alike should continue to evaluate best practice and guidelines as technology develops.

In addition to the above environmental and technological challenges, future investigations should also evaluate within-subject variability of phone holding position. A study that evaluates the same speaker cued to complete the same phone position over several days is needed. For example, it is likely that the cue to "hold the phone as if reading from it" results in slightly different positions each time. This variation, in combination with speaker variation (coming from natural speech variability as well as dysphonia severity that can vary even within a day) should be tested to fully understand the limitations of at-home app monitoring.

V. CONCLUSION

We determined that measures of mean f_o, MPT, and CPP are not impacted by phone position, making them viable acoustic options for at-home monitoring. However, absolute CPP values are likely not comparable to clinically derived values, but relative change in CPP seems to be appropriate within-subject. Therefore, alternative phone positions may be used for the acquisition of specific voice acoustic measures. This has significant implications for future user interface designs of smartphone apps intended to record patient voices outside of the clinic. Future work should evaluate the impact of speaker variability, including variation in how someone holds the phone over multiple trials, in combination with typical fluctuations in speech parameters over several days. This is necessary to fully understand the potential for apps to provide healthcare professionals with reliable information and to tease out meaningful clinical changes in acoustics from natural variation inherent in the way the recordings were obtained.

ACKNOWLEDGMENTS

This work was supported by the University Research Council Faculty Scholars Grant Award (V.S.M.). We would like to thank Haley Doerr, Jacqueline Meriwether, Samantha Neltner, and Mary Wilkens for their assistance with acoustical data collection and processing and Lauren Mikhail, M.A., CCC-SLP for her assistance with perceptual dysphonia ratings. We thank Elizabeth Viviano for modeling the different phone positions.

[1]The specific smartphone brand and type were reported for 21 participants enrolled in this study. These included Apple iPhone 7, 7+, 8+, XS, 11, 12, 13, 13 Pro Max; Samsung Galaxy S9, Galaxy S10e; Google Pixel 5, LG K30, K51; and Motorola Moto G Power 2021.

[2]Variation in smartphone sampling rate was dependent upon the recording app used. Only four participants had acoustic recordings acquired at a sampling rate of 22 050 Hz, whereas the remaining 52 participants were recorded at a rate of 44 100 Hz.

Awan, S. N., Giovinco, A., and Owens, J. (**2012**). "Effects of vocal intensity and vowel type on cepstral analysis of voice," J. Voice **26**(5), 670.E15–670.E20.

Awan, S. N., Roy, N., and Cohen, S. M. (**2014**). "Exploring the relationship between spectral and cepstral measures of voice and the voice handicap index (VHI)," J. Voice **28**(4), 430–439.

Bababekova, Y., Rosenfield, M., Hue, J. E., and Huang, R. R. (**2011**). "Font size and viewing distance of handheld smart phones," Optom. Vis. Sci. **88**(7), 795–797.

Berwick, D. M., and Hackbarth, A. D. (**2012**). "Eliminating waste in US health care," JAMA **307**(14), 1513–1516.

Boccardo, L. (**2021**). "Viewing distance of smartphones in presbyopic and non-presbyopic age," J. Optom. **14**(2), 120–126.

Boersma, P. (**2001**). "PRAAT, a system for doing phonetics by computer," Glot. Int. **5**, 341–345.

Bottalico, P., Codino, J., Cantor-Cutiva, L. C., Marks, K., Nudelman, C. J., Skeffington, J., Shrivastav, R., Jackson-Menaldi, M. C., Hunter, E. J., and Rubin, A. D. (**2020**). "Reproducibility of voice parameters: The effect of room acoustics and microphones," J. Voice **34**(3), 320–334.

Brown, W. S., Murry, T., and Hughes, D. (**1976**). "Comfortable effort level: An experimental variable," J. Acoust. Soc. Am. **60**(3), 696–699.

Cavalcanti, J. C., Englert, M., Oliveira, M., and Constantini, A. C. (**2023**). "Microphone and audio compression effects on acoustic voice analysis: A pilot study," J. Voice **37**, 162–172.

Costa, C. C., Ramos, H. V. L., Alves, W., Lamounier, P., Velasco, L., de Castro Velasco, L., and El Cheikh, M. R. (**2020**). "Partial laryngectomy and reconstruction with rotation of the epiglottis in the treatment of a rare laryngeal schwannoma: A case report," J. Med. Case Rep. **14**(1), 229.

Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier-Buchman, L., Friedrich, G., Van De Heyning, P., Remacle, M., and Woisard, V. (**2001**). "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," Eur. Arch. Otorhinolaryngol. **258**(2), 77–82.

Deliyski, D. D., Shaw, H. S., and Evans, M. K. (**2005**). "Adverse effects of environmental noise on acoustic voice quality measurements," J. Voice **19**(1), 15–28.

de Medeiros, A. M., Assunção, A. Á., and Barreto, S. M. (**2012**). "Absenteeism due to voice disorders in female teachers: A public health problem," Int. Arch. Occup. Environ. Health **85**(8), 853–864.

Gillespie, A. I., Gooding, W., Rosen, C., and Gartner-Schmidt, J. (**2014**). "Correlation of VHI-10 to voice laboratory measurements across five common voice disorders," J. Voice **28**(4), 440–448.

Grillo, E. U., Brosious, J. N., Sorrell, S. L., and Anand, S. (**2016**). "Influence of smartphones and software on acoustic voice measures," Int. J. Telerehab. **8**(2), 9–14.

Hakim, A. A., Kellish, A. S., Atabek, U., Spitz, F. R., and Hong, Y. K. (**2020**). "Implications for the use of telehealth in surgical patients during the COVID-19 pandemic," Am. J. Surg. **220**(1), 48–49.

Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., Rubin, A., Eller, R., Chandran, S., Abaza, M., Lyons, K., Divi, V., Lott, J., Johnson, J., and Hillenbrand, J. (**2014**). "Quantifying the cepstral peak prominence, a measure of dysphonia," J. Voice **28**(6), 783–788.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech. Lang. Hear. Res. **37**(4), 769–778.

Hogikyan, N. D., and Sethuraman, G. (**1999**). "Validation of an instrument to measure voice-related quality of life (V-RQOL)," J. Voice **13**(4), 557–569.

Jennett, P. A., Hall, L. A., Hailey, D., Ohinmaa, A., Anderson, C., Thomas, R., Young, B., Lorenzetti, D., and Scott, R. E. (**2003**). "The socio-economic impact of telehealth: A systematic review," J. Telemed. Telecare **9**(6), 311–320.

Johns, M. M. (**2003**). "Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts," Curr. Opin. Otolaryngol. Head Neck Surg. **11**(6), 456–461.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., and Hillman, R. E. (**2009**). "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," Am. J. Speech Lang. Pathol. **18**(2), 124–132.

Koo, T. K., and Li, M. Y. (**2016**). "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," J. Chiropr. Med. **15**(2), 155–163.

Krause, A. J., Walsh, E. H., Weissbrod, P. A., Taft, T. H., and Yadlapati, R. (**2022**). "An update on current treatment strategies for laryngopharyngeal reflux symptoms," Ann. N.Y. Acad. Sci. **1510**(1), 5–17.

Layfield, E., Triantafillou, V., Prasad, A., Deng, J., Shanti, R. M., Newman, J. G., and Rajasekaran, K. (**2020**). "Telemedicine for head and neck ambulatory visits during COVID -19: Evaluating usability and patient satisfaction," Head Neck **42**(7), 1681–1689.

Long, J., Cheung, R., Duong, S., Paynter, R., and Asper, L. (**2017**). "Viewing distance and eyestrain symptoms with prolonged viewing of smartphones," Clin. Exp. Optom. **100**(2), 133–137.

Majumder, S., and Deen, M. J. (**2019**). "Smartphone sensors for health monitoring and diagnosis," Sensors **19**(9), 2164.

Maryn, Y., Ysenbaert, F., Zarowski, A., and Vanspauwen, R. (**2017**). "Mobile communication devices, ambient noise, and acoustic voice measures," J. Voice **31**(2), 248.e11–248.e23.

Maryn, Y., and Zarowski, A. (**2015**). "Calibration of clinical audio recording and analysis systems for sound intensity measurement," Am. J. Speech Lang. Pathol. **24**(4), 608–618.

Misono, S., Peterson, C. B., Meredith, L., Banks, K., Bandyopadhyay, D., Yueh, B., and Frazier, P. A. (**2014**). "Psychosocial distress in patients presenting with voice concerns," J. Voice **28**(6), 753–761.

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., and Hillman, R. (**2018**). "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. **27**(3), 887–905.

Philips, R., Seim, N., Matrka, L., Locklear, B., Moberly, A. C., Inman, M., and Essig, G. (**2019**). "Cost savings associated with an outpatient otolaryngology telemedicine clinic," Laryngoscope Investig. Otolaryngol. **4**(2), 234–240.

Price, D. B., and Sataloff, R. T. (**1988**). "A simple technique for consistent microphone placement in voice recording," J. Voice **2**(3), 206–207.

Rakerd, B., Hunter, E. J., Berardi, M., and Bottalico, P. (**2018**). "Assessing the acoustic characteristics of rooms: A tutorial with examples," Perspect. ASHA Spec. Interest Groups **3**(19), 8–24.

Rosenthal, A. L., Lowell, S. Y., and Colton, R. H. (**2014**). "Aerodynamic and acoustic features of vocal effort," J. Voice **28**(2), 144–153.

Roy, N., Merrill, R. M., Gray, S. D., and Smith, E. M. (**2005**). "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," Laryngoscope **115**(11), 1988–1995.

Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., and Smith, E. M. (**2004**). "Prevalence of voice disorders in teachers and the general population," J. Speech Lang. Hear. Res. **47**(2), 281–293.

Simonyan, K., Barkmeier-Kraemer, J., Blitzer, A., Hallett, M., Houde, J. F., Jacobson Kimberley, T., Ozelius, L. J., Pitman, M. J., Richardson, R. M., Sharma, N., Tanner, K., and The NIH/NIDCD Workshop on Research Priorities in Spasmodic Dysphonia/Laryngeal Dystonia (**2021**). "Laryngeal dystonia: Multidisciplinary update on terminology, pathophysiology, and research priorities," Neurology **96**(21), 989–1001.

Švec, J. G., and Granqvist, S. (**2010**). "Guidelines for selecting microphones for human voice production research," Am. J. Speech Lang. Pathol. **19**(4), 356–368.

van Leer, E., Pfister, R. C., and Zhou, X. (**2017**). "An iOS-based cepstral peak prominence application: Feasibility for patient practice of resonant voice," J. Voice **31**(1), 131.e9–131.e16.

Vogel, A. P., and Morgan, A. T. (**2009**). "Factors affecting the quality of sound recording for speech and voice analysis," Int. J. Speech Lang. Pathol. **11**(6), 431–437.

Wallace, J., and Kanegaonkar, R. (**2020**). "The role of smartphone applications in clinical practice: A review," J. Laryngol. Otol. **134**(2), 96–103.

Witte, R. S., and Witte, J. S. (**2017**). *Statistics*, 11th ed. (Wiley, New York).

Wolfe, M. K., McDonald, N. C., and Holmes, G. M. (**2020**). "Transportation barriers to health care in the United States: Findings from the National Health Interview Survey, 1997–2017," Am. J. Public Health **110**(6), 815–822.

Wood, J. M., Athanasiadis, T., and Allen, J. (**2014**). "Laryngitis," BMJ **349**, g5827.

Yang, J., Xie, Z., and Seyler, B. C. (**2022**). "Comparing KTP and CO_2 laser excision for recurrent respiratory papillomatosis: A systematic review," Laryngoscope Investig. Otolaryngol. **7**(4), 970–981.