Research Article

Accuracy of Acoustic Measures of Voice via Telepractice Videoconferencing Platforms

Hasini R. Weerathunge,[a,b] (iD) Roxanne K. Segina,[b] Lauren Tracy,[c] (iD) and Cara E. Stepp[a,b,c] (iD)

Purpose: Telepractice improves patient access to clinical care for voice disorders. Acoustic assessment has the potential to provide critical, objective information during telepractice, yet its validity via telepractice is currently unknown. The current study investigated the accuracy of acoustic measures of voice in a variety of telepractice platforms.
Method: Twenty-nine voice samples from individuals with dysphonia were transmitted over six video conferencing platforms (Zoom with and without enhancements, Cisco WebEx, Microsoft Teams, Doxy.me, and VSee Messenger). Standard time-, spectral-, and cepstral-based acoustic measures were calculated. The effect of transmission condition on each acoustic measure was assessed using repeated-measures analyses of variance. For those acoustic measures for which transmission condition was a significant factor, linear regression analysis was performed on the difference between the original recording and each telepractice platform, with the overall severity of dysphonia, Internet speed, and ambient noise from the transmitter as predictors.
Results: Transmission condition was a statistically significant factor for all acoustic measures except for mean fundamental frequency (f_o). Ambient noise from the transmitter was a significant predictor of differences between platforms and the original recordings for all acoustic measures except f_o measures. All telepractice platforms affected acoustic measures in a statistically significantly manner, although the effects of platforms varied by measure.
Conclusions: Overall, measures of f_o were the least impacted by telepractice transmission. Microsoft Teams had the least and Zoom (with enhancements) had the most pronounced effects on acoustic measures. These results provide valuable insight into the relative validity of acoustic measures of voice when collected via telepractice.
Supplemental Material: https://doi.org/10.23641/asha.14794812

Voice disorders affect approximately 3%–9% of the U.S. population (Ramig & Verdolini, 1998; Roy et al., 2005) and can negatively impact daily communication as well as quality-of-life (Cohen et al., 2006; Franic et al., 2005; Krischke et al., 2005; Murry & Rosen, 2000; Rasch et al., 2005; Smith et al., 1994). All available voice assessments have different strengths and weaknesses, and therefore, current recommendations for complete voice assessment stress a multifactorial approach (Patel et al., 2018; Roy et al., 2013). Contemporary voice assessment commonly includes a case history, patient-reported outcomes,

laryngeal videostroboscopy, aerodynamic evaluation, acoustic assessment, and auditory-perceptual evaluation (Roy et al., 2013). However, there is weak evidence to support this standard voice evaluation during the utilization of telepractice, which has been growing in the past decade due to cost efficiencies, improved access, client demand, and quality of service (Grillo, 2019).

Due to the COVID-19 pandemic, speech-language pathologists (SLPs) had limited access to those voice evaluation methods that require direct patient interaction. As a result, telepractice has become a necessity for the clinical assessment of voice (Castillo-Allendes et al., 2020). Telepractice, in this context, refers to the application of telecommunication services to deliver speech-language pathology services remotely, linking clinicians to clients for assessment, intervention, or consultation, either synchronously or asynchronously (American Speech-Language-Hearing Association [ASHA], 2018). Telepractice has improved access to diagnosis

[a]Department of Biomedical Engineering, Boston University, MA
[b]Department of Speech, Language and Hearing Sciences, Boston University, MA
[c]Department of Otolaryngology—Head & Neck Surgery, Boston University School of Medicine, MA

Correspondence to Hasini R. Weerathunge: hasiniw@bu.edu

and treatment of voice disorders for those in rural communities and has allowed delivery of services to those who would otherwise have limited access to care (Barkmeier-Kraemer & Patel, 2016; Keck & Doarn, 2014; Kelchner, 2013; Mashima et al., 2003; Mashima & Brown, 2011; Mashima & Doarn, 2008; Molini-Avejonas et al., 2015). Effectiveness of telepractice treatment delivery has been validated for a range of voice disorders, including vocal fold nodules, muscle tension dysphonia, and Parkinson's disease as demonstrated by variations in outcome measures over time (Constantinescu et al., 2011; Fu et al., 2015; Grillo, 2019; Howell et al., 2009; Mashima et al., 2003; Rangarathnam et al., 2015; Tindall et al., 2008; Towey, 2012). Although earlier studies used specialized hardware and software within laboratory confines to conduct telepractice, videoconferencing platforms have become more readily available and accessible, enabling clinicians to reach clients via commercially available platforms and mobile telecommunication options. A variety of videoconferencing platforms, such as Skype (Skype Technologies), Cisco WebEx (Cisco Systems), Zoom (Zoom Video communications), VSee Messenger (VSee Lab, Inc.), Microsoft Teams (Microsoft Corporation), GoToMeetings (LogMeln, Inc.), and Doxy.me (Doxy.me, LLC), can be used to conduct telepractice in accordance with the Health Insurance Portability and Accountability Act of 1996 standards of privacy and data security (Nosowsky & Giordano, 2006).

Laryngeal videostroboscopy and instrumented aerodynamic evaluation are not possible when using videoconferencing platforms, as they require specialized equipment and direct patient contact. Case history and patient-reported outcomes can be evaluated via telepractice and provide valuable information, but these assessments are limited by their subjectivity and lack of specificity (Branski et al., 2010). Acoustic measures may be of use during telepractice voice assessment, as it only requires an acoustic recording of the client's voice, which is readily available through both synchronous and asynchronous telepractice methods. However, no study to date has assessed the validity of acoustic measures captured via video conferencing platforms commonly utilized in telepractice.

Acoustic measures have the potential to provide critical, objective information about voice disorders (Maryn & Weenink, 2015). Standard acoustic measures of voice used in clinical settings include measures of sound pressure level (SPL; mean and variance measures), fundamental frequency (f_o; mean and variance measures), and measures related to voice quality (Mehta & Hillman, 2008; Roy et al., 2013). However, as many of these measures are susceptible to environmental noise, they are typically obtained in noise-treated environments (Deliyski et al., 2005; Maryn & Weenink, 2015; Maryn et al., 2017; Yiu, 1999). Telepractice platforms, which can be utilized via desktop computers or a variety of portable devices (i.e., laptops, tablets, smartphones), are likely to introduce noise due to connection bandwidths and recording environments. However, the impact of telepractice platforms on acoustic measures of voice has not been directly assessed for commercially available telepractice platforms. Thus, clinicians currently do

not have a solid evidence base to support incorporating acoustic measures into telepractice voice assessment protocols. It is yet unclear whether incorporating these measures adds value, or instead adds additional uncertainty, to voice assessment. Information about the validity of acoustic measures of voice via telepractice is necessary to allow patients and clinicians to evaluate the future risk-to-benefit ratios of in-person versus telepractice voice care.

The recommended core set of acoustic parameters related to SPL and f_o are the mean SPL, the minimum and maximum SPL, the mean f_o, the f_o standard deviation, and the minimum and maximum f_o (Patel et al., 2018). Unfortunately, mean, minimum, and maximum SPL values require calibration, which is not easily feasible via telepractice platforms. However, SPL variance measures (such as the range and standard deviation) do not require calibration and are based on the relative changes of signal amplitude, and may be minimally affected by noise; thus, measures of SPL variance may provide useful information that is robust to transmission via telepractice. Likewise, standard clinical measures of f_o also have a high potential for assessment via telepractice; given a signal of sufficient signal periodicity, f_o measures may be minimally susceptible to signal noise. For instance, Maryn et al. (2017) reported that f_o measures were resistant to the recording system, environmental noise, and their combination and another study comparing the variability of acoustic measures captured via a variety of smartphone types reported that f_o showed acceptable levels of error (Jannetts et al., 2019).

Acoustic measures related to voice quality are also essential to standard clinical assessment, as changes in voice quality are a primary concern in most individuals with voice disorders. The majority of the traditional acoustic measures of voice quality (e.g., jitter, shimmer, harmonic-to-noise ratio [HNR]; Teixeira et al., 2013) are obtained via time-based algorithms to extract information about the dominant frequency and signal perturbation (Diercks et al., 2013). These measures are highly sensitive to ambient noise and thus may be expected to be impacted by transmission via telepractice platforms. Marsano-Cornejo et al. (2020) explored the influence of background noise levels on HNR and observed that HNR statistically significantly decreased as the background noise increased over 47.7 dB(A). Lebacq et al. (2017) further observed that speech recorded via smartphones was distorted by the signal processing applied by the devices and that it significantly influenced values of jitter and noise-to-harmonic ratio (the inverse of HNR). Finally, there is strong evidence that both jitter and shimmer are highly impacted by noise: in fact, these measures fail to retain accuracy and reliability when signal-to-noise ratios (SNR) drop below certain levels (30 dB; Deliyski et al., 2005) and thus are not compatible with telepractice transmission.

More recently, spectral and cepstral measures such as the low–high ratio (L/H ratio; Awan et al., 2014) and smoothed cepstral peak prominence (CPPS; Heman-Ackah et al., 2003) have been introduced to assess the amount of noise in acoustic signals and have been shown to correlate with voice quality percepts of breathiness and the overall

severity of dysphonia (Heman-Ackah et al., 2002, 2003; Hillenbrand et al., 1994; Hillenbrand & Houde, 1996; Klatt & Klatt, 1990). These acoustic measures of voice quality are based explicitly on noise or perturbation and therefore are theoretically highly susceptible to ambient noise levels in acoustic signals. However, Jannetts et al. (2019) examined CPPS values from speech recorded with a variety of smartphones, and although smartphone-based CPPS measures were statistically significantly lower than those derived from a reference microphone, the level of error was small. Thus, CPPS is impacted by the measurement noise induced by smartphone acquisition but may be more robust than other correlates of voice quality.

Relative fundamental frequency (RFF; Stepp et al., 2010) is a newer measure that reflects short-term changes in f_o surrounding voiceless obstruents and has been shown to correlate with listener perceptions of the voice quality percept of strain (Lien et al., 2015; McKenna & Stepp, 2018; Stepp et al., 2012). Although f_o is expected to be relatively robust to environmental noise, RFF requires accurate identification of phonatory cycles near the offset and onset of voicing. Therefore, accurate RFF measurement may be more difficult when signals have high levels of noise and may therefore be problematic if obtained via telepractice platforms.

An ideal clinical setting for voice evaluation is in a sound-treated environment, which minimizes ambient noise levels, while using an omnidirectional standard microphone to capture the full spectrum of the speech signal. However, when voice is transmitted through telepractice platforms, several factors may contribute towards modified sound quality. Ambient noise, as well as microphone characteristics, can affect the recording quality of the signal. To combat these issues, telepractice platforms have built-in sound enhancement algorithms to improve sound quality. Although they improve usability for most applications, these audio enhancements distort the original voice signal in terms of both amplitude and frequency. Noise suppression is a common enhancement that detects sustained sounds and reduces their intensity (Gunawan et al., 2014; Jagadeesan & Surazski, 2006; Zoom Video Communications, 2020). However, noise suppression could significantly affect sustained vowel productions utilized in acoustic measurements in voice telepractice. Automatic microphone volume control is another enhancement that normalizes telepractice platform outputs (Irukuvajhula et al., 2019). However, it may artificially remove dynamic changes inherent to the voice signal and could potentially affect acoustic measures of f_o and SPL.

Although audio enhancements are common in many telepractice platforms, several platforms offer the option to disable all enhancements. This occurs at the expense of increased noise in the signals; however, signal distortions can be avoided. For example, Zoom has the option "original sound," which disables noise suppression techniques, high-pass filtering, and automatic gain control. Similarly, Cisco WebEx and VSee Messenger provide the option to switch off automatic gain control of the microphone. However, not all telepractice platforms offer this option, and have unalterable enhancements. In addition to these audio enhancements

occurring at the telepractice platform level, the computers and Internet connections of both the transmitter and receiver could lead to potential alterations to the acoustic signal. Therefore, it is necessary to examine acoustic measures of voice to determine if they are valid when transmitted and recorded via telepractice platforms.

The main objective of the current study was to determine the accuracy of acoustic measures of voice in a variety of videoconferencing platforms used for telepractice. Leveraging an existing database of voice signals, common acoustic measures with the potential for telepractice were calculated for voice signals transmitted over five popular videoconferencing platforms as well as for the original signals that were recorded in a sound-treated environment. Furthermore, the Zoom platform was examined both with and without sound enhancements to capture the platform-specific ramifications of these audio enhancement algorithms. We hypothesized that all acoustic measures explicitly based on noise, such as CPPS and L/H ratio, or based on signal perturbation, such as HNR, would be significantly impacted by transmission condition with large effect sizes, whereas RFF and SPL variance would be impacted but with a small effect size, due to the lack of explicit reliance on noise. Finally, we hypothesized that the f_o mean and standard deviation would not be significantly impacted by transmission condition.

Method

Participants

Voice samples from a group of 29 cisgender participants (female = 14, male = 15) with a variety of voice disorder diagnoses and over a large age range (19–82 years; $M = 51.8$, $SD = 18.0$) were selected for the current study from an existing database of over 1,400 participant speech samples.[1] An a priori power analysis suggested that the use of 28 speakers would allow detection of small to medium effect sizes (e.g., $\eta_p^2 = .06$) with $\alpha = 0.005$ and power of 80%. Diagnoses for individuals with voice disorders included: Parkinson's disease ($N = 8$), muscle tension dysphonia ($N = 9$), adductor laryngeal dystonia (i.e., spasmodic dysphonia; $N = 5$), vocal fold nodules ($N = 4$), unilateral vocal fold polyp ($N = 2$), vocal fold scar ($N = 1$), and unilateral vocal fold paralysis ($N = 1$). All participants received their clinical diagnosis from a neurologist (for Parkinson's disease) or a laryngologist (for all other diagnoses), and they were all speakers of American English with no other history of speech, language, or hearing disorders. All participants passed a hearing screening at 30 dB HL for octave frequencies from 500 to 4000 Hz (ASHA, 2005; Burk & Wiley,

---

[1]During participant selection, we focused on identifying a relatively uniform distribution of CAPE-V overall severity of dysphonia, in order to (a) replicate a sample set of participants found in a clinical setting and (b) have a set of voices across a spectrum of overall severity scores such that overall severity could be utilized as a covariate in the statistical analysis for variabilities.

2004). An SLP specializing in voice completed blinded ratings of overall severity of dysphonia for each participant using the 100-mm visual analog scale of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009). Overall severity ranged from 0 to 64.4 mm on the CAPE-V. For intrarater reliability evaluation, the same rater repeated 20% of the ratings on a later date, and for interrater reliability evaluation, another SLP specializing in voice performed 20% of the ratings. Interrater reliability was .65 (moderately reliable) and intrarater reliability was .84 (good reliability). Informed consent was obtained from all participants, in compliance with the Boston University Institutional Review Board.

Experimental Procedure

Original Voice Recordings

All acoustic recordings were obtained in a sound-attenuated booth at Boston University. Audio signals were pre-amplified using the RME Quadmic II, sampled at 44100 Hz and 16 bits using a MOTU UltraLite-mk3 Hybrid sound card. A Shure WH20QTR headset microphone, angled 45° from the midline and 7 cm away from the corner of the mouth (Patel et al., 2018), was used to collect voice recordings. The mean SNR of original recordings was 30.69 dB (SD = 4.32 dB; recommended ASHA SNR guidelines ≥ 30 dB; Patel et al., 2018). Prerecorded voice clips were used intentionally to ensure that identical recordings were transmitted via each teleconferencing platform.

Stimuli

Recorded tasks included the production of three sustained vowels /ɑ/, /i/, and /u/, which were repeated 3 times each; the production of three vowel–consonant–vowel (VCV) utterances /ɑfɑ/, ifi/, and /ufu/, which were repeated 3 times each; and connected speech using the first paragraph of the Rainbow Passage (Fairbanks, 1960). All vowel tokens, VCV utterances, and connected speech segments were consolidated into one .wav file for each participant. A total data set of 29 participants' concatenated audio files was organized into two sets.

Experimental Setup

Six telepractice platforms, Zoom used with enhancements, Zoom used without enhancements (i.e., "original sound" option), Cisco WebEx, Microsoft Teams, Doxy.me, and VSee Messenger, and the original recorded signals were selected as the transmission conditions for the current study. These platforms were chosen based on their adherence to the Health Insurance Portability and Accountability Act of 1996 standards (Nosowsky & Giordano, 2006) and their prevalence of use for voice telepractice (Grillo, 2019). Each platform permits host/provider screen sharing, bandwidth optimization based on a client network by automatically adjusting for video and audio quality when Internet speeds reduce, encryption of transmitted information, and the immediate deletion of any transmitted information (Grillo, 2017).

Three experimenters were assigned the role of the *transmitter,* and four experimenters were assigned the role of *receiver*. Transmitters would transmit voice recordings of participants via the telepractice platform for the receiver to collect. Each videoconferencing session required one transmitter and one receiver. Each of the three transmitters was seated in their home environments, resulting in three distinct ambient noise conditions. The first transmitter utilized a Toshiba Satellite C50-A (Toshiba) computer with a Windows 10 operating system; the second transmitter utilized a Microsoft Surface Pro 3 (Microsoft Corporation) computer with a Windows 10 operating system; and the third transmitter utilized a MacBook Pro Retina 2015 (Apple Inc.) computer with an iOS X operating system. Two receivers utilized computers with Windows 10 operating systems and two receivers utilized computers with iOS X operating systems. The pairing of Windows and iOS operating systems was randomly assigned to have various combinations of operating system combinations regarding both the transmitter and receiver.

The sound enhancements of each computer's microphone and speaker were disabled for both the transmitter and receiver prior to connecting to the video conferencing call (see Supplemental Material S2 for an illustrated user guide on disabling sound enhancements). The microphone volume of the transmitter and the speaker volume of the receiver were set to 100% via their personal computer's sound settings as well as via each telepractice platform's sound settings. Within the VSee Messenger telepractice platform, however, the receiver's output volume was set to 50% to avoid signal clipping observed in pilot testing.

All concatenated audio files from the participants were saved to a handheld recording device (LS-10 Linear PCM Recorder; Olympus Corporation). A 1-kHz pure tone was available on the handheld recorder for calibration. The handheld recorder signal was transmitted to an external speaker (Soundcore Motion+ A3116011; Anker), in order to amplify the signal to an SPL similar to spoken voice. The volume of the handheld recorder was set to 100%, and the external speaker volume was set to 50% in order to mimic a spoken voice intensity in a room (approximately 80 dB SPL). The external speaker sound equalizer setting was set such that there was 0 dB amplification for the frequency range 80 Hz–12 kHz to provide a nondistorted output of the data.

The distance from the external speaker to each transmitter's computer microphone was 58 cm, which was determined based on pilot testing by investigators: the average distance calculated by 15 speakers, who measured the distance from the corner of their mouth to their personal computer microphones while seated comfortably and partaking in a videoconference. At the beginning of each recording block, the transmitter recorded the ambient noise in dB SPL using an SPL meter (CM-150; Galaxy Audio).

Videoconferencing calls took place during morning (9:00 a.m. through 12:00 p.m.) and afternoon (3:00 p.m. through 6:00 p.m.) sessions of 2-hr blocks in order to provide variability in Internet connection speeds. Once connected

to the videoconferencing call on each platform, both the transmitter and the receiver calculated their respective Internet speeds [latency (ms)/ downlink bandwidth (Mbps)/uplink bandwidth (Mbps)] using a desktop Internet speed test application (http://www.speedtest.com). The required bandwidth requirement for undisrupted audio and video conferencing is 1.2 Mbps (uplink/downlink) and 80 kbps respectively (Zoom Video Communications, 2021). The minimum bandwidths recorded for the current study (uplink/downlink: 4.56 Mbps/41.10 Mbps) was above this minimum requirement, and no connectivity loses were encountered that lasted for more than 1–2 s (i.e., the minimum duration of an utterance).

During each videoconferencing call, the receiver was asked to mute their microphone to mitigate the addition of any receiver ambient noise in the received transmission. The receiver was then asked to record each concatenated audio file on their personal computer using Praat acoustic software (Version 6.1.24; Boersma, 2020) at a sampling frequency of 44100 Hz. Using one telepractice platform at a time, the transmitter would present one of the two sets of concatenated audio files at a time from the handheld recorder. Once all recordings of the session were complete, the receiver saved the recordings as ".wav" files (see Supplemental Material S1 for an illustrated user guide on recording from telepractice platforms in real-time via Praat Software).

Data Analysis

All acoustic files were normalized prior to acoustic analysis. The objective acoustic measures mean f_o, f_o variation (standard deviation and range), SPL variation (standard deviation and range), HNR, L/H ratio, and CPPS were calculated offline, using Praat acoustic software. Three trained technicians manually annotated the beginning and end of each utterance. The middle 1-s region of each vowel and the entire utterance for connected speech stimuli and VCV utterances were used to calculate acoustic measures. All measures related to f_o, SPL, and HNR were calculated via the output of the "Voice Report" in Praat (pitch range [Hz] = 50, 600; maximum period factor = 1.3; maximum period amplitude = 1.6; silence threshold = 0.03; voicing threshold = 0.75). L/H ratio was calculated using a custom script in Praat and with a cutoff frequency of 4 kHz. CPPS was calculated as defined by Hillenbrand et al. (1994) using the "get CPPS" function in Praat. RFF (Vojtech et al., 2019) was calculated using an automated MATLAB algorithm (MATLAB: Version 2018a; MathWorks). A selected set of acoustic measures were calculated for each of the three types of voice stimuli. Using the sustained vowel stimuli, measures of HNR, L/H ratio, and CPPS were calculated. RFF Offset 10 and Onset 1 were calculated using the VCV utterances. Using the connected speech, L/H ratio, CPPS, mean f_o, f_o standard deviation, f_o range, and SPL range were calculated (see Table 1).

Statistical Analysis

Statistical analyses were completed using Minitab Statistical Software (Version 19; Minitab, Inc.). To account for conducting multiple analyses of variance (ANOVAs) and regressions, the significance for all statistical testing was set a priori at a conservative $p < .005$. Automated algorithms failed to provide conclusive RFF outputs for 7%–64% for RFF Onset 1 and 17%–75% for RFF Offset 10 of the data transmitted via the examined telepractice platforms. Given these differences in the number of valid samples per platform, no statistical tests were performed on RFF measures.

A repeated-measures ANOVA was performed for each remaining acoustic measure: HNR, L/H ratio (sustained vowel), CPPS (sustained vowel), L/H ratio (connected speech), CPPS (connected speech), mean f_o, f_o standard deviation, f_o range, and SPL range. For each model, transmission condition (i.e., original signal, Zoom used with enhancements, Zoom used without enhancements, Cisco WebEx, Microsoft Teams, Doxy.me, and VSee Messenger) was included as a fixed factor, participant was included as a random factor, and the speaker's overall severity of dysphonia was included as a covariate. For each acoustic measure, if transmission condition was a statistically significant factor, the partial eta squared (η_p^2) was calculated to determine its effect size, and post hoc Dunnett's tests were used to evaluate differences between the original signal and each of the telepractice platforms, with Cohen's d to calculate effect sizes for any statistically significant differences. For acoustic measures for which the ANOVA showed a statistically significant effect of transmission condition, a linear regression was carried out on the difference between each platform and the original recording, with the transmitter's Internet uplink speed, recipient's Internet downlink speed, transmitter's ambient noise level, and the speaker's overall severity of dysphonia as continuous predictor variables.

Results
Effects of Transmission Condition

For repeated measures ANOVAs, there were statistically significant main effects of the transmission condition on all acoustic measures except the mean f_o (see Table 2). There was a statistically significant main effect of transmission condition on the SPL range with a large effect size. The f_o standard deviation, CPPS (connected speech), CPPS (sustained vowel), L/H ratio (sustained vowel), and HNR all had a statistically significant main effect of transmission condition with medium effect sizes. There was a statistically significant main effect of transmission condition on f_o range and L/H ratio (connected speech) with small effect sizes. Post hoc Dunnett's tests were carried out for all acoustic measures that had a statistically significant main effect of transmission condition, with original recording as the control condition. The direction of each significant difference and the associated Cohen's d effect sizes are provided in Table 3. Results from Dunnett's tests indicated that measures of f_o, SPL range, and L/H ratios (connected speech and sustained vowels) were significantly different from the control condition only for specific platforms. However, the acoustic measures HNR (sustained vowel) and CPPS

Table 1. Stimuli type versus acoustic measures calculated.

Stimulus	Vocal f_o mean (Hz)	Vocal f_o SD (Hz)	Vocal f_o range (Hz)	SPL range (dB)	HNR (dB)	L/H ratio (dB)	CPPS (dB)	RFF Offset 10	RFF Onset 1
Sustained vowels	.	.	.	.	√	√	√	.	.
VCV utterences	.	.	.	.	.	.	.	√	√
Connected speech	√	√	√	√	.	√	√	.	.

Note. f_o = fundamental frequency; SPL = sound pressure level; HNR = harmonic-to-noise ratio; L/H = low–high ratio; CPPS = cepstral peak prominence (smoothed); RFF = relative fundamental frequency; VCV = vowel–consonant–vowel.

(connected speech and sustained vowel) were significantly different from the control condition for all videoconferencing platforms considered. Supplemental Material S3 indicates the average differences between measures after transmission relative to the original recordings.

Effects of Overall Severity of Dysphonia, Ambient Noise, and Internet Speed

Linear regressions were carried out for the difference in means between each platform and the original recording, with overall severity of dysphonia, ambient noise for the transmitter, uplink Internet speed for the transmitter, and downlink Internet speed for the receiver as continuous predictors (see Table 4). Ambient noise level for the transmitter was a statistically significant predictor for all acoustic measures, except for f_o measures. For f_o range, CPPS (connected speech and sustained vowel), and HNR, the participant's overall severity of dysphonia was a statistically significant predictor. For L/H ratio (connected speech and sustained vowel) and CPPS (sustained vowel), the transmitter's Internet uplink speed was a statistically significant predictor. For f_o mean and standard deviation, SPL range, L/H ratio (connected speech), and CPPS (connected speech and sustained vowel), the receiver's Internet downlink speech was a statistically significant predictor.

Discussion

The current study investigated the accuracy of common acoustic measures of voice in a variety of teleconferencing platforms. The results provide evidence that all investigated telepractice platforms significantly degrade the signal quality of speech signals. All acoustic measures calculated, except for f_o means, were statistically significantly affected by the telepractice platforms utilized. However, the effects of platform varied by measure, and the results provide valuable insights about which acoustic measures can be most accurately calculated across telepractice platforms, as well as which telepractice platforms have the least impact on acoustic measures.

Effects of Transmission Condition on SPL Range

SPL range was the only acoustic measure that had a significant main effect of transmission condition with a large effect size. Post hoc testing indicated that Cisco WebEx, Zoom used with enhancements, and Zoom used without enhancements caused statistically significant changes in the SPL range of speech samples relative to the original recordings. SPL range was increased in Cisco WebEx and Zoom used with enhancements with large effect sizes, whereas the SPL range was decreased in Zoom used without

Table 2. Results table for repeated-measures analysis of variance for acoustic measures with transmission condition as a fixed factor, participant as a random factor, and overall severity of dysphonia as a covariate.

Acoustic Measure	Overall severity of dysphonia					Transmission condition				
	df	F	η_p^2	Effect size	p	df	F	η_p^2	Effect size	p
Vocal f_o mean (Hz)	1	20.65	.04	Small	< .001	6	3.10	—	—	.005
Vocal f_o SD (Hz)	1	8.39	.02	Small	.004	6	14.93	.14	Medium	< .001
Vocal f_o range (Hz)	1	0.20	—	—	.656	6	3.46	.04	Small	.002
SPL range (dB)	1	0.64	—	—	.424	6	88.75	.50	Large	< .001
HNR (vowel; dB)	1	108.65	.17	Medium	< .001	6	12.71	.13	Medium	< .001
L/H ratio (dB)	1	10.30	.09	Small	.001	6	8.57	.09	Small	< .001
L/H ratio (vowel; dB)	1	2.11	—	—	.147	6	22.22	.20	Medium	< .001
CPPS (dB)	1	101.84	.16	Medium	< .001	6	20.48	.18	Medium	< .001
CPPS (vowel; dB)	1	108.32	.17	Medium	< .001	6	14.55	.14	Medium	< .001

Note. Level of significance: $p < .005$ (nonsignificant values dashed out); η_p^2 effect sizes: small (.01–.09), medium (.09–.25), large (>.25); f_o = fundamental frequency; SPL = sound pressure level; HNR = harmonic-to-noise ratio; L/H ratio = low–high ratio; CPPS = cepstral peak prominence (smoothed).

Table 3. Results table for post hoc Dunnett's tests with Cohen's *d* values and difference direction.

Telepractice platform	Vocal f_o SD (Hz)	Vocal f_o range (Hz)	SPL range (dB)	HNR (vowel; dB)	L/H ratio (dB)	L/H ratio (vowel; dB)	CPPS (dB)	CPPS (vowel; dB)
Cisco WebEx	0.77↑	.	3.54↑	−1.38↓	.	−1.28↓	−1.08↓	−1.53↓
Doxy.me	0.96↑	0.76↑	.	−1.28↓	−0.99↓	.	−1.40↓	−1.57↓
Microsoft Teams	.	.	.	−1.13↓	.	.	−1.79↓	−1.52↓
VSee Messenger	0.96↑	0.71↑	.	−1.27↓	−0.90↓	.	−1.51↓	−1.56↓
Zoom (with enhancements)	0.91↑	0.73↑	0.94↑	−0.96↓	−1.01↓	.	−1.10↓	−1.41↓
Zoom (without enhancements)	1.28↑	.	−0.62↓	−1.48↓	.	.	−1.66↓	−1.43↓

Note. Cohen's *d* effect sizes: 0.2 = small, 0.5 = medium, 0.8 = large; f_o = fundamental frequency; SPL = sound pressure level; HNR = harmonic-to-noise ratio; L/H ratio = low–high ratio; CPPS = cepstral peak prominence (smoothed).

enhancements with a medium effect size. We hypothesized that the SPL range would be minimally affected (i.e., with small effect sizes) by telepractice transmission, as SPL is minimally affected by additive noise in a voice signal. However, a factor we did not consider in our original hypotheses was the enhancements added by different telepractice platforms. To further illustrate the effects of transmission on SPL range variations, Figure 1 shows the normalized raw acoustic waveforms of a single participant as recorded over all platforms, for all three transmitters. When used without enhancements, Zoom introduces substantial levels of noise to the signal, a possible reason for the reduced SPL range measurement. However, the enhancements included in other platforms, most notably Cisco WebEx, manipulate the acoustic signals such that the amplitudes of the vowels are artificially sustained at a peak level. Thus, the significant increase in SPL range in Cisco WebEx observed is likely due to platform amplitude enhancements. However, for recordings via Transmitter 2, which was the transmitter with the highest level of ambient noise, Cisco WebEx signal amplitudes seem suppressed, which may be due to the enhancement algorithm in the platform failing to identify the sustained vowels as prominent signals to transmit, thus suppressing them as part of the ambient noise suppressions.

### Effects of Transmission Condition on Spectral and Cepstral Measures

As hypothesized, measures aimed at aperiodicity and/or noise were heavily impacted by telepractice transmission. CPPS (for both connected speech and sustained vowels) and HNR were statistically significantly affected by transmission condition with medium effect sizes, while L/H ratio (sustained vowels) was affected with a small effect size. Post hoc testing indicated that CPPS (connected speech and sustained vowels) and HNR (sustained vowels) had statistically significant decreases relative to the original speech signal when transmitted across all platforms, all with large associated effect sizes. Likewise, the L/H ratio was often significantly decreased after telepractice transmission. L/H ratio for connected speech showed statistically significant decreases, all with large effect sizes, when transmitted by Doxy.me, VSee Messenger, and Zoom used with enhancements; L/H ratio for sustained vowels was statistically significantly different from original recordings

Table 4. Results table for linear regression of differences in acoustic measures between the original recordings and signals transmitted via telepractice, with overall severity of dysphonia, ambient noise, uplink Internet speed for the transmitter, and downlink Internet speed for the receiver as continuous predictors.

Acoustic Measure	Overall severity of dysphonia		Ambient noise (dB SPL)		Uplink speed for the transmitter (Mbps)		Downlink speed for the receiver (Mbps)	
	β	*p*	β	*p*	β	*p*	β	*p*
Vocal f_o mean (Hz)	0.60	.329	−1.81	.047	−2.10	.027	**4.49**	**< .001**
Vocal f_o SD (Hz)	−1.16	.131	−0.43	.709	−2.40	.044	**5.13**	**< .001**
Vocal f_o range (Hz)	**11.77**	**< .001**	−5.68	.137	1.94	.625	6.45	.048
SPL range (dB)	0.65	.422	**7.24**	**< .001**	3.24	.010	**−7.07**	**< .001**
HNR (vowel; dB)	**1.13**	**< .001**	**0.80**	**< .001**	0.71	.003	−0.19	.324
L/H ratio (dB)	0.07	.766	**−1.34**	**< .001**	**4.42**	**< .001**	**−1.54**	**< .001**
L/H ratio (vowel; dB)	−0.19	.577	**−5.83**	**< .001**	**3.63**	**< .001**	0.54	.208
CPPS (dB)	**0.59**	**< .001**	**0.57**	**< .001**	0.06	.318	**−0.20**	**< .001**
CPPS (vowel; dB)	**1.24**	**< .001**	**1.00**	**< .001**	**0.44**	**< .001**	**−0.28**	**.004**

Note. Level of significance: $p < .005$ (significant differences bolded in the table). β = standardized beta coefficient; Mbps = Megabits per second; f_o = fundamental frequency; SPL = sound pressure level; HNR = harmonic-to-noise ratio; L/H ratio = low–high ratio; CPPS = cepstral peak prominence (smoothed).

Figure 1. Acoustic waveforms for normalized sustained vowel /a/ (three repetitions) for Participant ID Tele01, listed for all telepractice platforms and all transmitters.
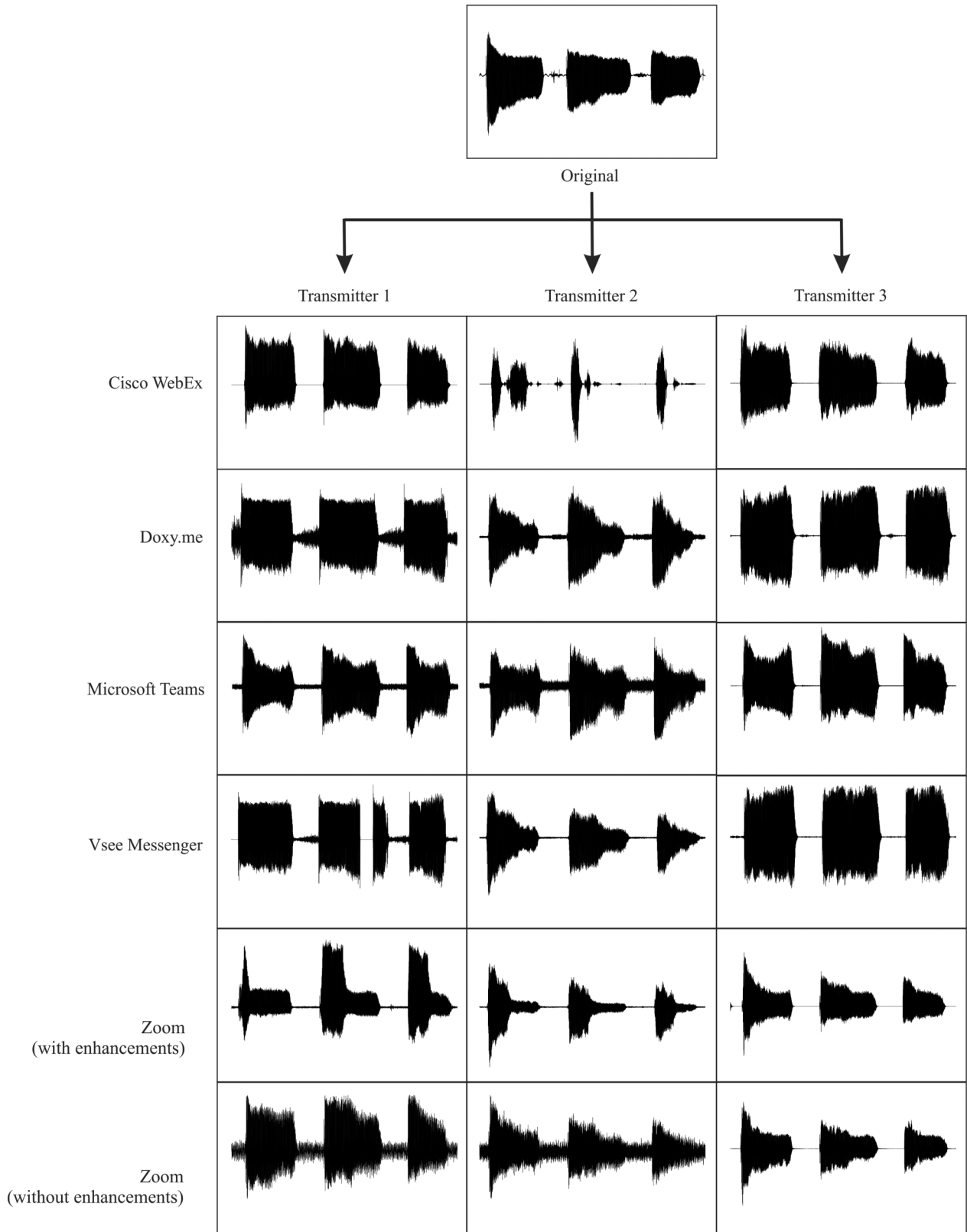
when transmitted by Cisco WebEx (with a large effect size). These decreases in CPPS, L/H ratio, and HNR are likely due to transmission and ambient noise. This is supported by the linear regression results, which indicate that ambient noise for the transmitter was a significant predictor for these measures. However, the L/H ratio showed significant effects only for specific platforms; thus, the effects on the L/H ratio may also be driven by the differences in audio enhancements provided via different platforms.

Effects of Transmission Condition on f_o Measures

Perhaps surprisingly, measures of f_o variability were impacted somewhat by telepractice transmission. f_o standard deviation was statistically significantly affected with a medium effect size and f_o range with a small effect size. Post hoc testing showed that differences in f_o standard deviation relative to the original signals were significantly increased for all platforms except Microsoft Teams; effect sizes for differences were all large except for Cisco WebEx, which was medium. Differences in f_o range due to transmission were even less compelling, with medium increases for only Doxy.me, VSee Messenger, and Zoom used with enhancements. These results provide evidence that standard time-based acoustic measures (i.e., f_o measures) might be less affected by telepractice platform compared to noise- or perturbation-based measures.

Possible Factors for Platform Differences

While the current results cannot fully predict which factors contribute to platform-based variations, there are possible candidates contributing to signal modifications. We observe in the acoustic waveforms in Figure 1 that there are a variety of signal enhancement algorithms being used across platforms in various degrees. Ambient noise suppression algorithms seem to affect sustained vowel phonations adversely, such that, in some cases they completely suppressed the amplitude of the signal, and in other cases they artificially amplified the signal such that the original pattern of the envelope was removed. The VSee Messenger platform signal for Transmitter 1 illustrates an example of package drop in signal transmission. This shows that Internet bandwidths and uplink/downlink Internet speeds are critical factors to maintain signal quality. These signal drops affect both time- and spectral-based acoustic measures, as indicated by linear regression results that downlink Internet speed for the receiver was a significant predictor for a majority of the acoustic measures. A factor that may not be visible through the current study results is the effect of signal compression applied in Voice over Internet Protocol for telecommunication platforms.

Effects of Overall Severity of Dysphonia, Ambient Noise, and Internet Speed

Linear regression results in Table 4 confirm findings of prior research that indicate spectral and cepstral measures

of voice are affected by aperiodicity in acoustic signals, caused by either ambient noise or dysphonia of the speaker (Deliyski et al., 2005; Heman-Ackah et al., 2002; Hillenbrand & Houde, 1996; Maryn et al., 2017; Watts & Awan, 2011). Internet uplink and downlink speeds were significant predictors of differences in measures between each platform and the original recording for several acoustic measures. For spectral and cepstral measures, L/H ratio, and CPPS, transmitter uplink speed was a significant predictor of the difference in measures between those transmitted over each platform and the original recording, which can be explained by the bandwidth reduction of the spectrum of the acoustic signal transmitted via the Internet (Xue & Lower, 2010). Downlink Internet speed is mainly responsible for lags in video and audio streams at the receiver end as well as spectral bandwidth reductions, which explains why many time-based and spectral measures of voice were affected significantly by downlink speed (Fuchs & Maxwell, 2016; Zhu et al., 2010).

Clinical Implications

In general, the differences observed in these acoustic measures across different telepractice platforms, compared to the original recordings, are both statistically significant and likely clinically meaningful. The differences observed in standard measures of f_o, as well as in standard measures targeting voice quality, were comparable in size for most cases. In some instances, the differences were many times larger than differences reported in the literature between individuals with and without voice disorders. For instance, in the current study, the difference in mean values for f_o measures across platforms were in the ranges of 7–25 Hz, 6–28 Hz, and 14–28 Hz for vocal f_o mean, standard deviation, and range, respectively (see Supplemental Material S3). The corresponding differences in these measures between individuals with and without vocal nodules have been reported to be 26, −2, and 67 Hz (Peppard et al., 1988). Differences in HNR and L/H ratio demonstrated an even greater impact due to telepractice transmission relative to clinically meaningful differences. In the current study, differences in HNR were 5.44–8.84 dB and differences in L/H ratio during sustained vowels were 1.00–8.90 dB due to telepractice transmission, whereas differences between individuals with and without dysphonia have been reported to be only 1.04 dB (HNR; Lathadevi & Guggarigoudar, 2018) and 1.13 dB (L/H ratio during sustained vowels; Lee et al., 2019). Surprisingly, while CPPS (connected speech) is clearly impacted by telepractice transmission, it may be more robust for clinical decision making than other measures targeted at voice quality. We found differences in CPPS from 1.44–2.34 dB across different telepractice platforms compared to original speech recordings, which is a range less than the 2.62-dB difference between speakers with and without dysphonia reported by Sauder et al. (2017). Therefore, if incorporating acoustic measures of voice for assessment during telepractice, measures of f_o and CPPS are likely to provide the most clinically relevant information.

Several recommendations can be provided with respect to which telepractice platforms may offer the best acoustic outcomes. Within the examined platforms, Microsoft Teams demonstrated the fewest effects on acoustic measures: For acoustic signals transmitted via the Teams platform, measures of f_o and SPL were not significantly impacted. Platforms such as Cisco WebEx, VSee Messenger, and Doxy.me appeared to boost signal amplitudes for sustained vowels in an artificial manner. Thus, they appear to be better optimized for connected speech acoustic measures in comparison to sustained vowel acoustic measures. Cisco WebEx also seemed to suppress noise in signals, such that the original signal noise levels were also removed. Finally, even though the use of Zoom without enhancements theoretically had minimal enhancements on the signal, the ambient noise levels in recorded signals were quite high, adversely affecting signal quality.

The main takeaway for clinicians of the current study is that a subset of objective acoustic measures of voice (i.e., measures of f_o and CPPS) can be measured without clinically significant differences, over specific videoconferencing platforms (i.e., Microsoft Teams). Moreover, based on the results, connected speech stimuli are the least affected by platform-based enhancements. In order to limit distortions applied to acoustic signals via videoconferencing platforms, the authors recommend that sound enhancement features of telecommunication devices and software should be disabled prior to telepractice sessions. Based on current results, clinicians can use several strategies during treatment planning to mitigate teleconferencing platform effects. They can request clients to send separate prerecorded voice samples prior to the telepractice session and utilize them for voice evaluations. If clinicians are using voice samples taken during telepractice sessions, they may have to revise which acoustic measures, and which teleconferencing platforms, should be utilized for voice evaluations.

Limitations

There are several differences in a typical telepractice setting compared to the recording sessions conducted in the current study. In the current study, previous speech recordings from participants recorded in a sound-treated room were played back via an external speaker at a specific distance from the transmitting computer. This method was followed to ensure that all telepractice platforms transmitted identical participant recordings that could later be compared with the original recording for different acoustic measures. However, in a typical telepractice session, patients vocalize in the presence of ambient noise. The effect of ambient noise in a typical telepractice recording environment on a participant's sound production is not considered in the current study. Thus, to fully replicate an ecologically valid telepractice environment, future studies should focus on replicating this study with real-time participant vocalizations.

Controlled aspects of our methods may limit the generalization of this study. The participant recordings were reproduced via an external speaker that was placed 58 cm

from the transmitter's computer microphone array. In order to reduce the effects of ambient noise, telepractice patients can utilize high-quality microphones and use head-mounted microphones that are only 5–10 cm away from the mouth. The current study only replicated a situation in which the patient is using a computer microphone from a distance typical of someone working on a laptop computer. The data collection was performed in quiet environments, and the transmitters ensured that unexpected ambient noises were not included in the recordings (e.g., doors closing, mobile phone notifications, construction/traffic sounds, air conditioner noises, bird sounds, etc.). Although it is preferred that patients would also adhere to connecting from a quiet environment for telepractice sessions, this cannot be ensured.

Other methodological aspects may also limit generalization to all voice telepractice clients. For instance, transmitters and receivers were counterbalanced to minimize the effects of variabilities in hardware and software of the computers utilized in each session. However, a limited number of transmitters and receivers were incorporated, and their Internet connectivity and ambient noise may not be representative of all clinicians and clients. In an attempt to minimize confounding effects of varied software enhancements available in the computers utilized for the study, all audio enhancements added via computers at both the transmitter and receiver sides were disabled prior to recording sessions, so that only the enhancements applied via telepractice platforms were applicable for the current study. Moreover, the automatic microphone volume control at the transmitter end and automatic speaker volume control at the receiver end was disabled. This intervention may not occur during a typical voice telepractice session, but based on the findings of this investigation, the authors recommend that voice telepractice providers instruct patients to disable audio enhancements prior to recording signals for acoustic measures.

A final limitation of this study was our inability to determine the impact of telepractice transmission on the accuracy of mean RFF values. Current automated RFF algorithms (Vojtech et al., 2019) include a variety of processing safeguards that remove RFF stimuli if they do not conform to expected signal conditions. The goal of this process is to ensure that the output of these algorithms is valid, even if signals are collected in less-than-ideal conditions. Unfortunately, the resulting signal quality caused by telepractice transmission required such a large portion of RFF stimuli to be discarded that it was not possible to compare across conditions. Future studies using manual calculation of RFF and, perhaps, algorithms that are specific to the signal quality expected after telepractice transmission may be needed to fully determine whether RFF can be validly collected via telepractice. However, based on the results of the current study, it is clear that the currently available automated algorithms are not well-suited for clinical assessment via telepractice.

Future Work

Further investigations should be carried out to identify if other voice assessment techniques are also affected

via telepractice platforms. For instance, although subjective, auditory-perceptual ratings (e.g., the CAPE-V) are an essential component of clinical voice assessment. However, it is not clear whether changes to signals due to telepractice platform transmission that result in degradation in acoustic measures of voice will likewise reduce the accuracy and reliability of auditory-perceptual measures. Previous work examining auditory-perceptual ratings of speech in individuals with Parkinson's disease when transmitted via a proprietary teleconferencing platform found comparable reliability scores when compared to face-to-face assessment (Constantinescu et al., 2011; Theodoros et al., 2006, 2016). However, this finding may not generalize to commercial telepractice platforms and a more diverse group of voices. Another area of future study is the extent to which telepractice platforms may be used to validly measure changes in voice acoustic measures over time. The results of the current study indicate that acoustic measures are significantly affected by telepractice platforms. However, the effects of a specific telepractice platform transmission for a single participant with a similar level of ambient noise setting could be similar over time. An alternate argument would be that the noise effects of the platform on acoustic measures are not constant, and thus introduce within-platform variabilities in measures. Thus, it is important to determine whether acoustic measure variations across time remain similar when transmitted via a specific telepractice platform. If so, acoustic measures taken over time from a single participant may present a useful outcome measure to assess the success of treatment. Other than objective acoustic measures, certain noninstrumental aerodynamic measures such as maximum phonation time (the maximum duration a person can sustain phonation of /ɑ/; Speyer et al., 2010) and s/z ratio (the ratio between the maximum phonation times of /s/ and /z/; Eckel & Boone, 1981) have the potential to be utilized as objective measures via telepractice platforms. However, their validity needs to be further investigated to ensure that ambient noise suppression algorithms in teleconferencing platforms do not suppress sustained phonation by misclassifying it as sustained ambient noise. Furthermore, the acoustic enhancement algorithms and data compression techniques used in each telepractice platform are proprietary, and it is difficult to clearly understand the effect of the enhancements by looking at a typical complex speech signal transmitted via each platform. In order to understand the signal-specific characteristic altered via each platform, a comprehensive study utilizing synthesized speech waveforms is needed. Such a study may provide insight into how to "undo" enhancements in postprocessing after telepractice platform transmission, which could allow for valid acoustic measurements. Finally, we acknowledge that telepractice is carried out via a variety of devices, ranging from desktop computers, laptops, tablets, to smartphones. In this study, the device type was kept constant (i.e., laptop computers) to avoid the confounding effects of device type used for telepractice transmission. Thus, the current work should be expanded to study the additional effects of the device type used for telepractice on acoustic measures of voice.

Conclusions

In the current study, we comprehensively investigated a set of commercially available teleconferencing platforms that are commonly used in telepractice in order to identify the accuracy of standard acoustic measures when transmitted over these platforms. The results of the study indicate that all acoustic measures, except for f_o mean, are statistically significantly affected by telepractice transmission. Overall, measures of f_o (mean, standard deviation, range) were the least impacted by telepractice transmission. SPL variability and acoustic measures aimed at voice quality were impacted by most telepractice platforms. Changes in acoustic measures of voice quality due to transmission were as large or larger than differences reported between individuals with and without voice disorders in previous work, suggesting that telepractice platform transmission imposes clinically relevant degradations to these measures. Microsoft Teams had the least, and Zoom used with enhancements had the most pronounced effects on acoustic measures overall. These results should provide insight into the relative validity of acoustic measures of voice when collected via telepractice platforms.

References

American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry*. https://www.asha.org/policy/gl2005-00014/

American Speech-Language-Hearing Association. (2018). *Telepractice: Overview*. https://www.asha.org/practice-portal/professional-issues/telepractice/

Awan, S. N., Roy, N., & Cohen, S. M. (2014). Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). *Journal of Voice, 28*(4), 430–439. https://doi.org/10.1016/j.jvoice.2013.12.008

Barkmeier-Kraemer, J. M., & Patel, R. R. (2016). *The next 10 years in voice evaluation and treatment*. Thieme Medical Publishers.

Boersma, P. (2020). *Praat: Doing phonetics by computer* [Computer program] (Version 6.1.24). https://www.praat.org

Branski, R. C., Cukier-Blaj, S., Pusic, A., Cano, S. J., Klassen, A., Mener, D., Patel, S., & Kraus, D. H. (2010). Measuring quality of life in dysphonic patients: A systematic review of content development in patient-reported outcomes measures. *Journal of Voice, 24*(2), 193–198. https://doi.org/10.1016/j.jvoice.2008.05.006

Burk, M. H., & Wiley, T. L. (2004). Continuous versus pulsed tones in audiometry. *American Journal of Audiology, 13*(1), 54–61. https://doi.org/10.1044/1059-0889(2004/008)

Castillo-Allendes, A., Contreras-Ruston, F., Cantor, L., Codino, J., Guzman, M., Malebran, C., Manzano, C., Pavez, A., Vaiano, T., Wilder, F., & Behlau, M. (2020). Voice therapy in the context of the COVID-19 pandemic: Guidelines for clinical practice.

Journal of Voice. Advance online publication. https://doi.org/10.1016/j.jvoice.2020.08.001

Cohen, S. M., Dupont, W. D., & Courey, M. S. (2006). Quality-of-life impact of non-neoplastic voice disorders: A meta-analysis. *Annals of Otology, Rhinology & Laryngology, 115*(2), 128–134. https://doi.org/10.1177/000348940611500209

Constantinescu, G., Theodoros, D., Russell, T., Ward, E., Wilson, S., & Wootton, R. (2011). Treating disordered speech and voice in Parkinson's disease online: A randomized controlled non-inferiority trial. *International Journal of Language & Communication Disorders, 46*(1), 1–16. https://doi.org/10.3109/13682822.2010.484848

Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice, 19*(1), 15–28. https://doi.org/10.1016/j.jvoice.2004.07.003

Diercks, G. R., Ojha, S., Infusino, S., Maurer, R., & Hartnick, C. J. (2013). Consistency of voice frequency and perturbation measures in children using cepstral analyses: A movement toward increased recording stability. *JAMA Otolaryngology–Head & Neck Surgery, 139*(8), 811–816. https://doi.org/10.1001/jamaoto.2013.3926

Eckel, F. C., & Boone, D. R. (1981). The S/Z ratio as an indicator of laryngeal pathology. *Journal of Speech and Hearing Disorders, 46*(2), 147–149. https://doi.org/10.1044/jshd.4602.147

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). Harper & Row.

Franic, D. M., Bramlett, R. E., & Bothe, A. C. (2005). Psychometric evaluation of disease specific quality of life instruments in voice disorders. *Journal of Voice, 19*(2), 300–315. https://doi.org/10.1016/j.jvoice.2004.03.003

Fu, S., Theodoros, D. G., & Ward, E. C. (2015). Delivery of intensive voice therapy for vocal fold nodules via telepractice: A pilot feasibility and efficacy study. *Journal of Voice, 29*(6), 696–706. https://doi.org/10.1016/j.jvoice.2014.12.003

Fuchs, R., & Maxwell, O. (2016). *The effects of mp3 compression on acoustic measurements of fundamental frequency and pitch range*. ISCA. https://doi.org/10.21437/SpeechProsody.2016-107

Grillo, E. U. (2017). Results of a survey offering clinical insights into speech-language pathology telepractice methods. *International Journal of Telerehabilitation, 9*(2), 25–30. https://doi.org/10.5195/ijt.2017.6230

Grillo, E. U. (2019). Building a successful voice telepractice program. *Perspectives of the ASHA Special Interest Groups, 4*(1), 100–110. https://doi.org/10.1044/2018_PERS-SIG3-2018-0014

Gunawan, D., Dickins, G. N., Holmberg, P., & Cartwright, R. J. (2014). *Spectral and spatial modification of noise captured during teleconferencing*. Dolby Laboratories Licensing Corp.

Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., Horman, M., & Sataloff, R. T. (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology, Rhinology & Laryngology, 112*(4), 324–333. https://doi.org/10.1177/000348940311200406

Heman-Ackah, Y. D., Michael, D. D., & Goding, G. S., Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice, 16*(1), 20–27. https://doi.org/10.1016/S0892-1997(02)00067-X

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research, 37*(4), 769–778. https://doi.org/10.1044/jshr.3704.769

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research, 39*(2), 311–321. https://doi.org/10.1044/jshr.3902.311

Howell, S., Tripoliti, E., & Pring, T. (2009). Delivering the Lee Silverman Voice Treatment (LSVT) by web camera: A feasibility study. *International Journal of Language & Communication Disorders, 44*(3), 287–300. https://doi.org/10.1080/13682820802033968

Irukuvajhula, S., Nalla, R. K., & Mantri, S. (2019). *Adaptive volume control using speech loudness gesture*. Polycom Inc.

Jagadeesan, R., & Surazski, L. K. (2006). *Devices, software and methods for generating aggregate comfort noise in teleconferencing over VoIP networks*. Cisco Technology Inc.

Jannetts, S., Schaeffler, F., Beck, J., & Cowen, S. (2019). Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types. *International Journal of Language & Communication Disorders, 54*(2), 292–305. https://doi.org/10.1111/1460-6984.12457

Keck, C. S., & Doarn, C. R. (2014). Telehealth technology applications in speech-language pathology. *Telemedicine and e-Health, 20*(7), 653–659. https://doi.org/10.1089/tmj.2013.0295

Kelchner, L. (2013). Telehealth and the treatment of voice disorders: A discussion regarding evidence. *SIG 3 Perspectives on Voice and Voice Disorders, 23*(3), 88–94. https://doi.org/10.1044/vvd23.3.88

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America, 87*(2), 820–857. https://doi.org/10.1121/1.398894

Krischke, S., Weigelt, S., Hoppe, U., Köllner, V., Klotz, M., Eysholdt, U., & Rosanowski, F. (2005). Quality of life in dysphonic patients. *Journal of Voice, 19*(1), 132–137. https://doi.org/10.1016/j.jvoice.2004.01.007

Lathadevi, H., & Guggarigoudar, S. P. (2018). Objective acoustic analysis and comparison of normal and abnormal voices. *Journal of Clinical & Diagnostic Research, 12*(12), MC01–MC04. https://doi.org/10.7860/JCDR/2018/36782.12310

Lebacq, J., Schoentgen, J., Cantarella, G., Bruss, F. T., Manfredi, C., & DeJonckere, P. (2017). Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research. *Journal of Voice, 31*(5), 550–556. https://doi.org/10.1016/j.jvoice.2017.02.017

Lee, S. J., Choi, H.-S., & Kim, H. (2019). Acoustic Psychometric Severity Index of Dysphonia (APSID): Development and clinical application. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2019.11.006

Lien, Y.-A. S., Michener, C. M., Eadie, T. L., & Stepp, C. E. (2015). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research, 58*(3), 566–575. https://doi.org/10.1044/2015_JSLHR-S-14-0194

Marsano-Cornejo, M.-J., Roco-Videla, Á., Capona-Corbalán, D., & Silva-Harthey, C. (2020). Variation of the acoustic parameter harmonic-to-noise ratio in relation to different background noise levels. *Acta Otorrinolaringologica Espanola, S0001-6519*(20), 30141–30142.

Maryn, Y., & Weenink, D. (2015). Objective dysphonia measures in the program Praat: Smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice, 29*(1), 35–43. https://doi.org/10.1016/j.jvoice.2014.06.015

Maryn, Y., Ysenbaert, F., Zarowski, A., & Vanspauwen, R. (2017). Mobile communication devices, ambient noise, and acoustic

voice measures. *Journal of Voice, 31*(2), 248.e211–248.e223. https://doi.org/10.1016/j.jvoice.2016.07.023

Mashima, P. A., Birkmire-Peters, D. P., Syms, M. J., Holtel, M. R., Burgess, L., & Peters, L. J. (2003). Telehealth. *American Journal of Speech-Language Pathology, 12*(4), 432–439. https://doi.org/10.1044/1058-0360(2003/089)

Mashima, P. A., & Brown, J. E. (2011). Remote management of voice and swallowing disorders. *Otolaryngologic Clinics of North America, 44*(6), 1305–1316. https://doi.org/10.1016/j.otc.2011.08.007

Mashima, P. A., & Doarn, C. R. (2008). Overview of telehealth activities in speech-language pathology. *Telemedicine and e-Health, 14*(10), 1101–1117. https://doi.org/10.1089/tmj.2008.0080

McKenna, V. S., & Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America, 144*(3), 1643–1658. https://doi.org/10.1121/1.5055234

Mehta, D. D., & Hillman, R. E. (2008). Voice assessment: Updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(3), 211–215. https://doi.org/10.1097/MOO.0b013e3282fe96ce

Molini-Avejonas, D. R., Rondon-Melo, S., de La Higuera Amato, C. A., & Samelli, A. G. (2015). A systematic review of the use of telehealth in speech, language and hearing sciences. *Journal of Telemedicine and Telecare, 21*(7), 367–376. https://doi.org/10.1177/1357633X15583215

Murry, T., & Rosen, C. A. (2000). Outcome measurements and quality of life in voice disorders. *Otolaryngologic Clinics of North America, 33*(4), 905–916. https://doi.org/10.1016/S0030-6665(05)70251-6

Nosowsky, R., & Giordano, T. J. (2006). The Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule: Implications for clinical research. *Annual Review of Medicine, 57,* 575–590. https://doi.org/10.1146/annurev.med.57.121304.131257

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Svec, J. G., & 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. *American Journal of Speech-Language Pathology, 27*(3), 887–905. https://doi.org/10.1044/2018_AJSLP-17-0009

Peppard, R. C., Bless, D. M., & Milenkovic, P. (1988). Comparison of young adult singers and nonsingers with vocal nodules. *Journal of Voice, 2*(3), 250–260. https://doi.org/10.1016/S0892-1997(88)80083-3

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41*(1), S101–S116. https://doi.org/10.1044/jslhr.4101.s101

Rangarathnam, B., McCullough, G. H., Pickett, H., Zraick, R. I., Tulunay-Ugur, O., & McCullough, K. C. (2015). Telepractice versus in-person delivery of voice therapy for primary muscle tension dysphonia. *American Journal of Speech-Language Pathology, 24*(3), 386–399. https://doi.org/10.1044/2015_AJSLP-14-0017

Rasch, T., Günther, S., Hoppe, U., Eysholdt, U., & Rosanowski, F. (2005). Voice-related quality of life in organic and functional voice disorders. *Logopedics Phoniatrics Vocology, 30*(1), 9–13. https://doi.org/10.1080/14015430510006640

Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M. P., Mehta, D., Paul, D., & Hillman, R. (2013). Evidence-based clinical voice assessment: A systematic review. *American Journal of Speech-Language Pathology, 22*(2), 212–226. https://doi.org/10.1044/1058-0360(2012/12-0014)

Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope, 115*(11), 1988–1995. https://doi.org/10.1097/01.mlg.0000179174.32345.41

Sauder, C., Bretl, M., & Eadie, T. (2017). Predicting voice disorder status from smoothed measures of cepstral peak prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice, 31*(5), 557–566. https://doi.org/10.1016/j.jvoice.2017.01.006

Smith, E., Verdolini, K., Gray, S., Nichols, S., Lemke, J., Barkmeier, J., Dove, H., & Hoffman, H. (1994). Effect of voice disorders on quality of life. *NCVS Status Program Report, 7,* 1–17.

Speyer, R., Bogaardt, H. C., Passos, V. L., Roodenburg, N. P., Zumach, A., Heijnen, M. A., Baijens, L. W. J., Fleskens, S. J. H. M., & Brunings, J. W. (2010). Maximum phonation time: Variability and reliability. *Journal of Voice, 24*(3), 281–284. https://doi.org/10.1016/j.jvoice.2008.10.004

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53*(5), 1220–1226. https://doi.org/10.1044/1092-4388(2010/09-0234)

Stepp, C. E., Sawin, D. E., & Eadie, T. L. (2012). The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 55*(6), 1887–1896. https://doi.org/10.1044/1092-4388(2012/11-0294)

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis—Jitter, shimmer and HNR parameters. *Procedia Technology, 9,* 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

Theodoros, D. G., Constantinescu, G., Russell, T. G., Ward, E. C., Wilson, S. J., & Wootton, R. (2006). Treating the speech disorder in Parkinson's disease online. *Journal of Telemedicine and Telecare, 12*(Suppl. 3), 88−91. https://doi.org/10.1258/135763306779380101

Theodoros, D. G., Hill, A. J., & Russell, T. G. (2016). Clinical and quality of life outcomes of speech treatment for Parkinson's disease delivered to the home via telerehabilitation: A noninferiority randomized controlled trial. *American Journal of Speech-Language Pathology, 25*(2), 214–232. https://doi.org/10.1044/2015_AJSLP-15-0005

Tindall, L. R., Huebner, R. A., Stemple, J. C., & Kleinert, H. L. (2008). Videophone-delivered voice therapy: A comparative analysis of outcomes to traditional delivery for adults with Parkinson's disease. *Telemedicine and e-Health, 14*(10), 1070–1077. https://doi.org/10.1089/tmj.2008.0040

Towey, M. P. (2012). Speech therapy telepractice for vocal cord dysfunction (VCD): MaineCare (Medicaid) cost savings. *International Journal of Telerehabilitation, 4*(1), 33–36. https://doi.org/10.5195/ijt.2012.6095

Vojtech, J. M., Segina, R. K., Buckley, D. P., Kolin, K. R., Tardif, M. C., Noordzij, J. P., & Stepp, C. E. (2019). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *The Journal of the Acoustical Society of America, 146*(5), 3184–3202. https://doi.org/10.1121/1.5131025

Watts, C. R., & Awan, S. N. (2011). Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts. *Journal of Speech, Language, and Hearing Research, 54*(6), 1525–1537. https://doi.org/10.1044/1092-4388(2011/10-0209)

Xue, S. A., & Lower, A. (2010). Acoustic fidelity of internet bandwidths for measures used in speech and voice disorders. *The Journal of the Acoustical Society of America, 128*(3), 1366–1376. https://doi.org/10.1121/1.3467764

Yiu, E. M. (1999). Limitations of perturbation measures in clinical acoustic voice analysis. *Asia Pacific Journal of Speech, Language and Hearing, 4*(3), 155–166. https://doi.org/10.1179/136132899807557475

Zhu, Y., Witt, R. E., MacCallum, J. K., & Jiang, J. J. (2010). Effects of the voice over internet protocol on perturbation analysis of normal and pathological phonation. *Folia Phoniatrica et Logopaedica, 62*(6), 288–296. https://doi.org/10.1159/000285807

Zoom Video Communications. (2020). *Background noise suppression*. Retrieved October 2, 2020, from https://support.zoom.us/hc/en-us/articles/360046244692-Background-noise-suppression

Zoom Video Communications. (2021). *System requirements for Windows, macOS, and Linux*. Retrieved February 17, 2020, from https://support.zoom.us/hc/en-us/articles/201362023-System-requirements-for-Windows-macOS-and-Linux