JSLHR

Research Article

The Impact of Communication Modality on Voice Production

Lauren F. Tracy,[a] Roxanne K. Segina,[b] Manuel Diaz Cadiz,[b] and Cara E. Stepp[a,b,c]

Purpose: Communicating remotely using audio and audiovisual technology is ubiquitous in modern work and social environments. Remote communication is increasing in medicine and in voice therapy delivery, and this evolution may have an impact on speakers' voices. This study sought to determine whether these communication modalities impact the voice production of typical speakers.
Method: The speech acoustics of 12 participants with healthy voices were recorded as they held standardized conversations with a single investigator using three communication modalities: in-person, remote-audio, and remote-audiovisual. Participants rated their vocal effort on a 100-mm visual analog scale.
Results: Compared to in-person communication, self-ratings of vocal effort were statistically significantly increased for remote-audiovisual communication; vocal effort during remote-audio and in-person communication were not significantly different. In comparison to in-person communication, vocal intensity and smoothed cepstral peak prominence (CPPS) were statistically significantly higher during remote-audio and remote-audiovisual communication. Effect sizes for CPPS changes were larger than for sound pressure level (SPL), and changes in CPPS and SPL between in-person and remote-audiovisual communication were not significantly correlated.
Conclusions: Vocal effort and SPL were increased when using remote-audio and remote-audiovisual communication in comparison to in-person communication. Voice quality was also impacted by technology use, with changes in CPPS that were consistent with, but not fully explained by, increases in SPL. This may impact the telepractice delivery of voice therapy, and further investigation is warranted.

Society is increasingly dependent on remote communication to stay connected. Both audio and audiovisual remote communication are routinely used for personal and professional interactions. The utilization of these resources is expanding with an increase in the number of telecommuting remote workers over the last decade. There has been a recent surge in telehealth delivery through video and phone conferencing, and this may have a lasting presence.

Audiovisual communication may present a superior option for information exchange in comparison to audio-only communication. Audiovisual communication allows observation of facial expressions and visual cues to clarify meanings. Combining audio and visual information has been shown to increase retention and comprehension in numerous settings (OSHA, 1996). A systematic review comparing remote videoconferencing to telephone use in health care delivery identified fewer medication errors and improved diagnostic accuracy for videoconferencing visits (Rush et al., 2018). Telehealth delivery of voice therapy has created improved access to this limited resource, and its efficacy is well-supported (Fu et al., 2015; Mashima & Brown, 2011). Audiovisual technology allows multiple advantages for delivery of remote voice therapy, including clinician modeling, shared screen use for speech stimuli, and biofeedback. Given these reported benefits, there has been substantial increase in the utilization of audiovisual communication in the workforce.

Unsurprisingly, there is a known impact on vocal function for frequent users of remote audio communication. Both telemarketers and emergency telecommunicators report an increase in voice complaints relative to the general population (dos Santos et al., 2016; Johns-Fiedler & van Mersbergen, 2015; Jones et al., 2002). However, the impact of audiovisual communication on vocal function has not been evaluated. Furthermore, studies in these individuals

[a]Department of Otolaryngology—Head and Neck Surgery, Boston University School of Medicine, MA
[b]Department of Speech, Language and Hearing Sciences, Boston University, MA
[c]Department of Biomedical Engineering, Boston University, MA
Correspondence to Lauren F. Tracy: lauren.tracy@bmc.org

Editor-in-Chief: Bharath Chandrasekaran
Editor: Kate Bunton
Received April 9, 2020
Revision received June 4, 2020
Accepted June 18, 2020
https://doi.org/10.1044/2020_JSLHR-20-00161

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

may conflate the effects of increased vocal use with the specific effects of remote communication.

It is conceivable that interaction with technology during remote communication has an impact on vocal function and vocal effort. Vocal effort includes psychological and physiological components, as well as effort relative to the communication environment (Hunter et al., 2020). These factors may all contribute to increased vocal effort during remote communication. For example, the inability to accurately determine the listener's distance during remote-audio communication may cause the speaker to increase vocal intensity (Pelegrín-García et al., 2011). Poor posture, which can be observed during telephone use, may also contribute to changes in vocal function (Cardoso et al., 2019). Additionally, experience with remote-communication during periods of limited reception may result in learned behaviors of increased vocal intensity, regardless of the current communication environment.

With increasing use of both audio and audiovisual technology for remote communication, it is important to understand the potential effects on voice, as well as possible risks associated with the utilization of this technology. Therefore, this study aims to evaluate changes in vocal function that may occur during remote communication. The effects of communication modality on self-reported vocal effort, vocal intensity, and vocal quality (estimated with two acoustic measures, the low–high (LH) ratio and the smoothed cepstral peak prominence [CPPS]) were examined using three communication modalities: in-person, remote-audio, and remote-audiovisual. Our overall hypothesis was that communication modality would have an effect on the voice production of speakers with healthy voices, with the lowest vocal effort, lowest vocal intensity, lowest CPPS, and highest LH ratio during in-person communication; and the highest vocal effort, highest vocal intensity, highest CPPS, and lowest LH ratio during remote-audio communication.

Methods

Participants

A group of 12 cisgender adults (six women, six men; M_{age} = 24.0 years, range: 19–33 years) participated in the current study. Informed consent was obtained from all participants prior to the study, in compliance with the Boston University Institutional Review Board.

All participants reported speaking standard American English as their first language and reported no history of speech, language, or hearing disorders. All but one participant passed an audiometric hearing screening at loudness levels of 25 dB HL at the frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz. The remaining participant passed the audiometric hearing screening at 35 dB HL in his left ear and 25 dB HL in his right ear.

When asked to determine their experience level with "video conferencing" (Skype, FaceTime, Zoom, etc.) across their lifetime, seven individuals reported greater than 31 hr, or "extensive"; four individuals reported between 6 and 30 hr, or "moderate"; and one individual reported between 0 and 5 hr, or "minimal." Eleven participants reported spending 1 hr or less on the phone daily, whereas one reported spending between 0 and 3 hr on the phone daily. All participants completed the Voice Related Quality of Life questionnaire, with scores ranging between 10 and 15 (M = 11.6; Hogikyan & Sethuraman, 1999).

Procedure

Two portable tablets (Samsung Galaxy Tab E; Samsung Electronics) were used to connect to a conference call using Zoom Video Communications. Once connected, one tablet was placed in a sound-attenuated booth at Boston University; the other was placed in a small conference room nearby. Both tablets were placed on tablet stands with their screens angled at approximately 45°.

A single experimenter (R. K. S.) interacted with each participant. The experimenter wore a lapel-clip microphone (JK MIC-J044 Lavalier Microphone; JK Global Trading LLC) on her collar to capture acoustic recordings of her speech. A Hot Spot transducer (K&K Sound) was placed on the anterior surface of the experimenter's neck to capture voice-related vibrations. The experimenter's microphone and transducer signals were recorded with a digital handheld audio recorder (LS-10 Linear PCM Recorder; Olympus). Signals were collected in .wav format at 44100 Hz and 16 bits.

Each participant was seated in the sound-attenuated booth. All participants wore a headset microphone over the ear (MX153 Subminiature Earset Microphone; Shure), angled 45° below midline and 7 cm away from the corner of the mouth. A neck surface accelerometer (Knowles BU-21771; Knowles Acoustics) was placed at the notch of the neck with double-sided adhesive tape. Recordings of the neck surface vibrations were obtained to capture the vocal fold vibration of the participant, free of ambient noise. Acoustic recordings were made using SONAR Artist acoustic software on a desktop computer. Audio was collected in .wav format at 44100 Hz and 16 bits. Using this acoustic recording equipment, prior to the experiment, electrolarynx output located 7 cm from the participant's microphone was recorded while its sound pressure level in dB SPL was simultaneously measured via a sound pressure level meter (CM-150; Galaxy Audio) placed at the microphone for use in calibrating speech recordings.

The study tasks occurred using three communication modalities: in-person, remote-audio, and remote-audiovisual. A Zoom video conference call was used to conduct the two remote conditions. During these tasks, the participant sat approximately 32 in. from the tablet, and the experimenter sat in the nearby conference room. During in-person communication, the experimenter was in the sound booth with the participant, seated approximately 32 in. away. The participant and experimenter used each communication modality to accomplish a specific collaborative communication task. These communication tasks were adapted from scenarios employed by Seita et al. (2018). Scenarios prompted participants to collaborate with the experimenter to decide on 10 items that they would bring with them if stranded and to rank those items in order of importance. During in-person communication, participants were asked to imagine they and the experimenter were stuck on a boat lost at sea; during remote-audio communication, participants imagined they were astronauts stuck on the moon; and during remote-audiovisual communication, participants imagined they were stranded in the desert. The order of communication modality and communication task was counterbalanced across participants.

Participants were given a clipboard with a pen and paper for note-taking during the tasks. In the remote conditions, the experimenter informed participants of the volume buttons on the side of the tablet in case they wanted to adjust the volume of the tablets. Each participant's tablet volume was set to 80% before each trial. Each communication task lasted 10 min.

Between tasks, a 5-min voice break occurred, during which participants did not engage in conversation to allow for vocal rest. During this time, participants were asked to rate their self-perceived vocal effort on a 100-mm visual analog scale (VAS) as in McKenna and Stepp (2018). Using a VAS allows explicit anchors to guide raters on a continuous scale (Gerratt et al., 1993); 0 mm on the scale was labeled "No Effort," and 100 mm on the scale was labeled "Max Effort." Participants received a new VAS for each vocal task and did not have access to previous VAS ratings.

Data Analysis

Speech acoustics (from the participant's microphone) during times at which only the participant was speaking were further analyzed. This was achieved via a custom MATLAB script by (a) time-aligning the microphone and neck vibratory signals of the participant and the experimenter, (b) thresholding the envelope of the participant's neck vibratory signals to remove pauses, and (c) thresholding the envelope of the experimenter's neck vibratory signals to remove times when the experimenter was speaking. Threshold values for each signal were tuned manually by visual inspection of the first 10 s of each recording in order to ensure reasonable pause and voice segmentation. The resulting speech acoustics for each task were an average of 112 s in duration (range: 57–191 s).

The participants' mean vocal intensity in dB SPL was calculated for each communication task by calibrating the acoustic signals with respect to the electrolarynx test tone. The CPPS and LH ratio were also computed for each communication task. These measures were selected because they can be used for analysis of voices during connected speech. The CPPS and LH ratio were both calculated using Praat acoustic software (Boersma, 2002).

In order to provide context to any changes in study outcomes, a certified speech-language pathologist specializing in voice and voice disorders rated each speech sample using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009). The speech stimuli were presented, and the CAPE-V ratings were gathered using a custom MATLAB graphical user interface. The speech samples were rated in random order, with an additional ~22% of samples repeated in order to assess intrarater reliability.

Statistical Analysis

Four one-way repeated measures analyses of variance (ANOVAs) examined the effect of communication modality (in-person, remote-audio, and remote-audiovisual) on each of the four outcomes: vocal effort, vocal intensity, CPPS, and LH ratio. Effect sizes for each ANOVA factor were calculated using a squared partial curvilinear correlation (η_p^2), designated as either small (~ .01), medium (~ .09), or large (> .25) effect size (Witte & Witte, 2010). For outcome measures that indicated a significant effect of communication modality, post hoc analyses were conducted with Bonferroni-corrected (nine comparisons) paired *t* tests. Cohen's *d* effect sizes were calculated to further assess statistically significant differences. Cohen's *d* values were designated as either small (0.2–0.3), medium (~ 0.5), or large (> 0.8; Witte & Witte, 2010). Based on the primary study results, we performed an exploratory analysis to determine the relationship between changes in voice quality as measured by CPPS and vocal intensity. Per participant, the difference in CPPS and vocal intensity between remote-audiovisual and in-person as well as remote-audio and in-person was computed. Simple linear regressions were performed, with the changes in vocal intensity between communication modalities as the independent variables and the changes in CPPS between communication modalities as the dependent variables.

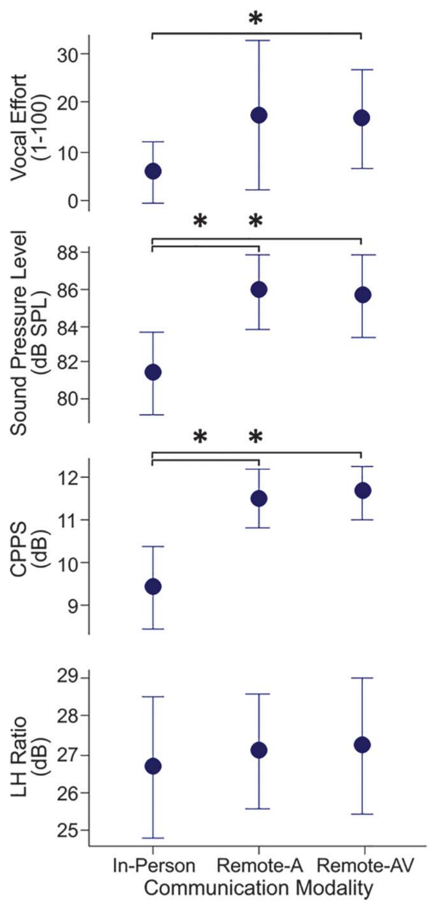
As post hoc analysis better interprets changes in study outcomes in light of auditory-perceptual measures, the CAPE-V ratings of the speech-language pathologist were documented as a function of communication modality. The intrarater reliability of the rater for each of the features was assessed via Pearson correlation coefficients as r = .84 for overall severity of dysphonia, r = .53 for roughness, r = .66 for breathiness, and r = .79 for strain. Due to the small range of values in this population with healthy voices, we also assessed intrarater reliability by examining the average absolute differences in repeated ratings (mm), which were 0.82 for overall severity of dysphonia, 1.07 for roughness, 1.26 for breathiness, and 1.28 for strain.

Results

Figure 1 shows averages and standard deviations of the studied variables. Table 1 details the results of each ANOVA.

In this cohort of 12 speakers with healthy voices, there were differences in vocal effort when comparing in-person communication versus remote communication with both audio and audiovisual technology. The mean self-rating of vocal effort was 5.33 (*SD* = 9.83) for in-person, 17.08 (*SD* = 24.76) for remote-audio, and 16.33 (*SD* = 16.39) for remote-audiovisual. The results of the ANOVA indicated a statistically significant effect of communication modality on vocal effort

Figure 1. Mean and 95% confidence intervals for average vocal effort, SPL, CPPS, and LH ratio between three communication modalities—in-person, remote-audio, and remote-audiovisual. Error bars show 95% confidence intervals. Asterisk denotes statistical significance. SPL = sound pressure level; CPPS = smoothed cepstral peak prominence; LH ratio = low–high ratio; AV = audiovisual; A = audio.



with a large effect size. The results of post hoc testing found that only remote-audiovisual was statistically significantly increased relative to in-person (p_{adj} = .025, d = 0.84).

The results of the ANOVA indicated a statistically significant effect of communication modality on vocal intensity with a large effect size. Post hoc testing indicated that vocal intensity was increased for both remote scenarios with mean vocal intensity of 85.9 dB SPL (SD = 3.6 dB; p_{adj} < .001, d = 0.73) during remote-audio and 85.6 dB SPL (SD = 3.6 dB; p_{adj} < .001, d = 1.19) during remote-audiovisual, in comparison to 81.3 dB SPL (SD = 3.6 dB) during in-person communication.

The effects of communication modality on acoustic estimates of vocal quality were conflicting. Although the results of the ANOVA on CPPS indicated a statistically significant effect of communication modality with a large effect size, the ANOVA on LH ratio did not find a statistically significant effect of communication modality. Post hoc testing indicated that participants' mean CPPS values were increased during remote communication, with mean CPPS during remote-audio at 11.5 dB (SD = 1.1 dB; p_{adj} < .001, d = 1.50) and during remote-audiovisual at 11.6 dB (SD = 1.0 dB; p_{adj} < .001, d = 1.69), in comparison to 9.4 dB (SD = 1.6 dB) for in-person communication. LH ratios were similar across the three communication modalities, with 26.7 dB (SD = 3.0 dB) for in-person communication in comparison to 27.1 dB (SD = 2.4 dB) for remote-audio and 27.2 dB (SD = 2.9 dB) for remote-audiovisual.

Table 1. Results of analysis of variance models.

Factor	df	F value	p value	Effect size (η_p^2)	Qualitative effect size
Vocal effort					
Communication modality	2	4.06	.032	.27	Large
Sound pressure level					
Communication modality	2	43.95	< .001	.80	Large
CPPS					
Communication modality	2	36.51	< .001	.77	Large
LH ratio					
Communication modality	2	0.34	.719	—	—

Note. Em dashes indicate effect size not reported if effect was not significant. CPPS = smoothed cepstral peak prominence; LH ratio = low–high ratio.

The results of the exploratory analysis to determine the relationships between changes in voice quality as measured by CPPS and vocal intensity indicated weak relationships. The regression comparing in-person and remote-audiovisual communication showed that 19.4% of the variance in the changes in CPPS was explained by the changes in vocal intensity (p = .152). Although the regression comparing in-person and remote-audio communication did indicate that they had a statistically significant relationship, only 33.3% of the variance in the changes in CPPS was explained by the changes in vocal intensity (p = .049).

Table 2 provides descriptive statistics for the average CAPE-V ratings as a function of communication modality. Ratings for all features were generally low, with mean values that did not appreciably differ by communication modality.

Evaluation of volume adjustment throughout the study identified that six of 12 participants did not adjust volume with either communication modality. Three of 12 participants increased the volume by one notch (20%) to 100% volume in both audio and audiovisual conditions. Two of 12 participants increased the volume to 100% only during remote-audio communication, and one participant increased the volume to 100% only during remote-audiovisual communication. Volume was not decreased in any studied condition.

Discussion

The aim of this study was to examine the impact of communication modality on voice production and vocal

Table 2. Average (standard deviation) statistics for Consensus Auditory-Perceptual Evaluation of Voice ratings as a function of communication modality (OS = overall severity of dysphonia).

Communication modality	OS	Roughness	Breathiness	Strain
In-person	7.79 (7.33)	2.49 (6.54)	5.82 (7.77)	2.71 (3.08)
Remote-Audio	5.6 (6.74)	3.30 (6.24)	3.61 (6.77)	2.04 (2.36)
Remote-Audiovisual	6.43 (6.79)	3.18 (5.62)	2.54 (6.64)	3.79 (3.93)

effort in speakers with healthy voices. We hypothesized that speakers would report increased vocal effort and produce differences in vocal acoustics consistent with increased vocal effort for remote in comparison to in-person communication. Our hypothesis was supported by the increases in speaker-reported vocal effort and vocal intensity during remote communication. Surprisingly, our results contrasted with those from Shewmaker et al. (2010) who did not identify a statistically significant difference in vocal intensity based on communication modality. They evaluated voice production during face-to-face communication, using landline phones, and cellular phones in both noisy and quiet conditions. They found no statistically significant differences between communication modalities when comparing the measures of self-rated vocal effort, vocal intensity, and fundamental frequency. It is likely that methodological differences between these two studies are responsible for the differences in findings. In the current study, participants were required to complete a collaborative task with a communication partner, a task that required accurate information transfer. Furthermore, the participant received explicit and implicit social-communication cues from the communication partner about the success of this information transfer. In contrast, the tasks employed by Shewmaker et al. were not conversational but rather involved a short monologue ("Please describe in detail how you make a peanut butter and jelly sandwich.") and reading a short passage. It is likely that the interactive nature of the tasks employed in the current study better represent the demands on speakers during real-life remote communication scenarios.

Basis of Changes in Vocal Intensity With Communication Modality

Analysis of speakers' vocal intensity demonstrated an increase for both remote speaking scenarios (audio and audiovisual) and no significant difference in vocal intensity between the two remote scenarios. Increases in vocal intensity during phone use have been previously attributed to the Lombard effect, wherein speakers increase intensity in background noise (Shewmaker et al., 2010). Although the intensity of received speech signal was equalized between the scenarios and each modality was used in the same quiet, sound-treated room, these results suggest that the Lombard effect may have a more general effect. Although it is often associated with noisy environments, the Lombard effect may in fact generalize to any environment in which communication is hindered or perceived to be hindered. Future research is needed to identify the types and degrees of communication deterioration that can induce the Lombard effect.

Changes in Voice Quality

In addition to vocal intensity, participants demonstrated some changes in voice quality during remote communication. CPPS is a reliable measure of dysphonia and strongly correlates with overall dysphonia severity (Heman-Ackah et al., 2014). Higher CPPS values correlate with improved periodicity of speech signals and improved vocal quality, whereas lower CPPS values represent decreased quality of voice due to disturbed periodicity (Heman-Ackah et al., 2002). Lower CPPS values also have been shown to correlate with perceptions of roughness, hoarseness, and strain (Maryn et al., 2010).

Previous work in individuals with dysphonia may suggest that lower CPPS is a straightforward indicator of poor voice use. However, the relationship between CPPS and voice quality is more complex. For instance, the association between vocal intensity and acoustic estimates of vocal quality has been well documented (Brockmann et al., 2008). Orlikoff and Kahane (1991) and Gelfer (1995) demonstrated that higher vocal intensity resulted in lower measurements of jitter and shimmer (representing less vocal perturbation) during sustained vowel production. Likewise, Awan et al. (2012) showed that even small variations in vocal intensity affect cepstral peak prominence (CPP). Phadke et al. (2020) quantified the changes in CPP and CPPS relative to vocal intensity among a group of female teachers. They found that an increase of 10 dB SPL correlated with an increase of 0.7 dB in CPP and a 1.2-dB increase in CPPS. In a similar study evaluating female speakers with normal and hyperfunctional voices, Brockmann-Bauser et al. (2019) found a higher CPPS-to-SPL ratio with a 10-dB increase in vocal intensity associated with a 2.2-dB increase in CPPS for both groups.

Given the known relationship between vocal intensity and CPPS, it is unsurprising that increases in CPP and CPPS can be found as a result of other voice quality manipulations. In a study by Rosenthal et al. (2014), when healthy speakers were asked to increase vocal effort, there was a significant increase in CPP during maximal effortful speech. Authors concluded that increases in mean SPL may have contributed to the increases in CPP. However, vocal intensity may not always be the source of changes in CPPS. MacPherson et al. (2017) examined vocal intensity, fundamental frequency, CPPS, and LH ratio in adults with healthy voices in two cognitive load conditions. Increased

CPPS and decreased LH ratio were associated with increased cognitive load and autonomic arousal, even though increased vocal intensity was not associated with the increased cognitive load. These changes may be interpreted as creation of a more “pressed” voice, with increased (periodic) energy in the higher vocal harmonics (MacPherson et al., 2017). Although the extemporaneous stimuli produced in this study design do not allow for confirmation of increased periodic energy in higher vocal harmonics, this interpretation is generally compatible with the results of the current study: Although both vocal intensity and CPPS increased in remote tasks relative to in-person communication, the CPPS increases had larger effect sizes (d = 1.5–1.69 vs. d = 0.73–1.19), and our exploratory analysis showed that only a small amount of the variance in changes in CPPS were explained by the changes in vocal intensity (19%–33%). The findings suggest that changes observed in voice quality, as measured by changes in CPPS, can in part be explained by an interaction with communication modality, and are not only a byproduct of the changes in vocal intensity.

Importantly, the relationship between vocal intensity and voice quality may differ between those speakers with voice disorders in comparison to those without voice disorders. Individuals with voice disorders characterized by auditory-perceptual strain and increased vocal effort have been shown to have lower CPP than individuals with healthy voices (Lowell et al., 2012; Watts et al., 2015). When speakers without voice disorders are asked to produce voice with increased vocal effort (McKenna & Stepp, 2018; Rosenthal et al. 2014) or are given vocal loading tasks (Sundarrajan et al., 2017), there is a consistent increase in CPP when compared to typical speaking. Therefore, results from this study of healthy, nondysphonic speakers may not be applicable to individuals with dysphonia and additional investigation in this population is warranted.

It is noteworthy that, while CPPS increased, the LH ratio did not demonstrate significant differences across vocal conditions. This finding may be due to the inclusion of only healthy, nondysphonic participants as disparities in LH ratio of acoustic energy are uniquely features of dysphonia and may not be represented well in nondysphonic speakers. Less disordered voices also demonstrate increased variability in LH ratios, which may be partly attributed to increased fundamental frequency variability in normal voices in comparison to disordered voices (Awan et al., 2010). Furthermore, LH ratios are hypothesized to be most helpful in the assessment of breathy voice quality, which is not a feature of normal healthy speakers (Hillenbrand & Houde, 1996).

Finally, although CPPS values changed as a function of communication modality, changes in voice quality were not reflected in the clinical ratings of voice quality via the CAPE-V: Average ratings of overall severity of dysphonia, roughness, breathiness, and strain were all quite low (all less than 8). This small range of ratings is unsurprising given the population of speakers with healthy voices and likely explains the relatively low intrarater reliability values: Although the Pearson correlations ranged r = .53–0.84, the average difference between repeated ratings was quite small (0.82–1.28) given the scale range of 0–100. The lack of sensitivity of the clinical auditory-perceptual ratings to speaker changes in remote communication environments is not particularly surprising. Although ratings by a single clinician are the clinical standard for the care of individuals with voice disorders, they are designed to capture the full range of the presentation of individuals with voice disorders. Thus, it is unsurprising that these features were not sensitive to the small changes in speakers with healthy voices in the current study, a finding consistent with previous work in speakers with healthy voices (Fujiki et al., 2017).

Changes in Vocal Effort

Consistent with acoustic measures, speakers' self-reported vocal effort was increased during remote communication. The overall increases in rated vocal effort for remote communication relative to in-person observed in this study are similar to those changes observed when comparing quiet communication with communication in background noise (Shewmaker et al., 2010). It is generally accepted that the added input of facial and visual cues makes remote-audiovisual communication easier to use relative to remote-audio communication. However, in this study, only remote-audiovisual communication demonstrated statistically significant increases in effort. It is conceivable that increased familiarity with remote-audio communication (such as cell phone use) relative to remote-audiovisual communication mitigated the potential benefits of audiovisual communication. However, variability across participants may also play a role in these results. There was a high degree of variability in reported effort across all three communication scenarios, which was most pronounced for remote-audio. This likely contributed to the lack of statistical significance. This large degree of variability is hypothesized to be attributed to variable previous experience with remote communication or to inherent attitudes about interacting with this technology. For example, those who primarily use remote communication in settings of noisy backgrounds may be conditioned to speak with increased vocal intensity and effort. In an increasingly remotely connected society, identifying those individuals who respond with increased vocal strain to remote communication may provide opportunities for intervention and education.

Limitations and Future Directions

One acknowledged limitation of this study is that only individuals without voice disorders were evaluated. It is conceivable that individuals with voice disorders would have different acoustic manifestations of vocal effort. Speakers with dysphonia may also have greater variations in intensity and quality under the studied conditions. Additionally, evaluating differences in both those dysphonic speakers without laryngeal pathology and for those with laryngeal pathology would improve generalizability of these results. This study evaluated typical speech and did not evaluate changes in voice production, which may occur during voice treatment with therapy. Furthermore, although these results have implications for individuals who use remote communication occupationally (e.g., telemarketers and emergency telecommunicators), it is not clear whether the changes seen here in typical speakers are representative of individuals who use remote communication regularly. For instance, although the speakers in the current study experienced changes in their vocalization in response to remote communication environments, speakers who regularly experience such environments may show a reduction (due to long-term adaptation) or an enhancement (due to vocal susceptibility or maladaptive compensations; Stepp et al., 2017) of these responses. Finally, this study incorporated a relatively small sample of young adults. Future investigations should include a larger number of participants of variable age to improve generalizability.

Finally, this project used simulated remote conferencing in "ideal" acoustic environments in order to acquire speech acoustic data from which valid acoustic estimates of voice quality could be computed. However, the use of remote conferencing modalities in more typical environments, with background noise, may lead to different results. Furthermore, as we work to transition these findings to telepractice settings in voice therapy, we need to consider what modes of assessment, other than patient-reported measures, can be reliably accessed in nonideal acoustic environments. Additionally, participants in this study were not given any particular instructions for optimizing their vocalization during the remote communication conditions. Future work should explore the mitigating impacts of remote communication through optimization of positioning, posture, and background noise.

Conclusions

In this preliminary study, speakers without voice disorders communicated in three different scenarios: in-person, remote-audio, and remote-audiovisual. Self-rated vocal effort was greater during remote-audio and remote-audiovisual communication in comparison to in-person. Remote-audio and remote-audiovisual communication resulted in increased vocal intensity and, to a greater degree, changes in voice quality as measured by increases in CPPS. Differences in vocal effort and voice production between in-person and remote communication warrant further investigation and should be examined in speakers with voice disorders and during remote delivery of voice therapy.

Acknowledgments

This work was supported by Grants DC015570 (awarded to Cara E. Stepp) and DC015446 (awarded to Robert E. Hillman) from the National Institute on Deafness and Other Communication Disorders. The authors thank Paul Park for assistance with data processing.

References

Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice, 26*(5), 670.e15–670.e6.7E20. https://doi.org/10.1016/j.jvoice.2011.12.001

Awan, S. N., Roy, N., Jetté, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24*(9), 742–758. https://doi.org/10.3109/02699206.2010.492446

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345.

Brockmann, M., Stork, C., Carding, P. N., & Drinnan, M. J. (2008). Voice loudness and gender effects on jitter and shimmer in healthy adults. *Journal of Speech, Language, and Hearing Research, 51*(5), 1152–1160. https://doi.org/10.1044/1092-4388(2008/06-0208)

Brockmann-Bauser, M., Van Stan, J. H., Carvalho Sampaio, M., Bohlender, J. E., Hillman, R. E., & Mehta, D. D. (2019). Effects of vocal intensity and fundamental frequency on cepstral peak prominence in patients with voice disorders and vocally healthy controls. *Journal of Voice*. Advance online publication. https://doi.org/10.1016/j.jvoice.2019.11.015

Cardoso, R., Lumini-Oliveira, J., & Meneses, R. F. (2019). Associations between posture, voice, and dysphonia: A systematic review. *Journal of Voice, 33*(1), 124.e1–124.e12. https://doi.org/10.1016/j.jvoice.2017.08.030

dos Santos, C. T., Santos, C., Lopes, L. W., Silva, P. O. C., & de Lima-Silva, M. F. B. (2016). Relationship between working and voice conditions self-reported by telemarketers of an emergency call center. *Communication Disorders, Audiology and Swallowing, 28*(5), 583–594. https://doi.org/10.1590/2317-1782/20162015125

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

Fujiki, R. B., Chapleau, A., Sundarrajan, A., McKenna, V., & Sivasankar, M. P. (2017). The interaction of surface hydration and vocal loading on voice measures. *Journal of Voice, 31*(2), 211–217. https://doi.org/10.1016/j.jvoice.2016.07.005

Gelfer, M. P. (1995). Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability. *Journal of Speech, Language, and Hearing Research, 38*(6), 1189–1198. https://doi.org/10.1044/jshr.3806.1189

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech, Language, and Hearing Research, 36*(1), 14–20. https://doi.org/10.1044/jshr.3601.14

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

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., & Hillenbrand, J. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice, 28*(6), 783–788. https://doi.org/10.1016/j.jvoice.2014.05.005

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech.

Journal of Speech, Language, and Hearing Research, 39(2), 311–321. https://doi.org/10.1044/jshr.3902.311

Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13*(4), 557–569. https://doi.org/10.1016/S0892-1997(99)80010-1

Hunter, E. J., Cantor-Cutiva, L. C., van Leer, E., van Mersbergen, M., Nanjundeswaran, C. D., Bottalico, P., Sandage, M. J., & Whitling, S. (2020). Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *Journal of Speech, Language, and Hearing Research, 63*(2), 509–532. https://doi.org/10.1044/2019_JSLHR-19-00057

Johns-Fiedler, H., & van Mersbergen, M. (2015). The prevalence of voice disorders in 911 emergency telecommunicators. *Journal of Voice, 29*(3), 389.e1–389.e10. https://doi.org/10.1016/j.jvoice.2014.08.008

Jones, K., Sigmon, J., Hock, L., Nelson, E., Sullivan, M., & Ogren, F. (2002). Prevalence and risk factors for voice problems among telemarketers. *Archives of Otolaryngology–Head & Neck Surgery, 128*(5), 571–577. https://doi.org/10.1001/archotol.128.5.571

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)

Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology & Laryngology, 121*(8), 539–548. https://doi.org/10.1177/000348941212100808

MacPherson, M. K., Abur, D., & Stepp, C. E. (2017). Acoustic measures of voice and physiologic measures of autonomic arousal during speech as a function of cognitive load. *Journal of Voice, 31*(4), 504.e1–504.e9. https://doi.org/10.1016/j.jvoice.2016.10.021

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice, 24*(5), 540–555. https://doi.org/10.1016/j.jvoice.2008.12.014

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

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

Orlikoff, R. F., & Kahane, J. C. (1991). Influence of mean sound pressure level on jitter and shimmer measures. *Journal of Voice, 5*(2), 113–119. https://doi.org/10.1016/S0892-1997(05)80175-4

OSHA. (1996). *United States Department of Labor. Presenting effective presentations with visual aids*. http://www.osha.gov

Pelegrín-García, D., Smits, B., Brunskog, J., & Jeong, C.-H. (2011). Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America, 129*(4), 1981–1990. https://doi.org/10.1121/1.3552881

Phadke, K. V., Laukkanen, A.-M., Ilomäki, I., Kankare, E., Geneid, A., & Švec, J. G. (2020). Cepstral and perceptual investigations in female teachers with functionally healthy voice. *Journal of Voice, 34*(3), 485.E33–485.E43. https://doi.org/10.1016/j.jvoice.2018.09.010

Rosenthal, A. L., Lowell, S. Y., & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice, 28*(2), 144–153. https://doi.org/10.1016/j.jvoice.2013.09.007

Rush, K. L., Howlett, L., Munro, A., & Burton, L. (2018). Videoconference compared to telephone in healthcare delivery: A systematic review. *International Journal of Medical Informatics, 118,* 44–53. https://doi.org/10.1016/j.ijmedinf.2018.07.007

Seita, M., Albusays, K., Kafle, S., Stinson, M., & Huenerfauth, M. (2018). *Behavioral changes in speakers who are automatically captioned in meetings with deaf or hard-of-hearing peers*. Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, ACM.

Shewmaker, M. B., Hapner, E. R., Gilman, M., Klein, A. M., & Johns, M. M., III (2010). Analysis of voice change during cellular phone use: A blinded controlled study. *Journal of Voice, 24*(3), 308–313. https://doi.org/10.1016/j.jvoice.2008.09.002

Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Pieter Noordzij, J., & Lupiani, A. A. (2017). Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research, 60*(6), 1545–1550. https://doi.org/10.1044/2017_JSLHR-S-16-0282

Sundarrajan, A., Huber, J. E., & Sivasankar, M. P. (2017). Respiratory and laryngeal changes with vocal loading in younger and older individuals. *Journal of Speech, Language, and Hearing Research, 60*(9), 2551–2556. https://doi.org/10.1044/2017_JSLHR-S-17-0106

Watts, C. R., & Awan, S. N. (2015). An examination of variations in the cepstral spectral index of dysphonia across a single breath group in connected speech. *Journal of Voice, 29*(1), 26–34. https://doi.org/10.1016/j.jvoice.2014.04.012

Witte, R. S., & Witte, J. S. (2010). *Statistics* (9th ed.). Wiley.