



OPEN

Acoustic voice characteristics with and without wearing a facemask

Duy Duong Nguyen, Patricia McCabe, Donna Thomas, Alison Purcell, Maree Doble, Daniel Novakovic, Antonia Chacon & Catherine Madill

Facemasks are essential for healthcare workers but characteristics of the voice whilst wearing this personal protective equipment are not well understood. In the present study, we compared acoustic voice measures in recordings of sixteen adults producing standardised vocal tasks with and without wearing either a surgical mask or a KN95 mask. Data were analysed for mean spectral levels at 0–1 kHz and 1–8 kHz regions, an energy ratio between 0–1 and 1–8 kHz (LH1000), harmonics-to-noise ratio (HNR), smoothed cepstral peak prominence (CPPS), and vocal intensity. In connected speech there was significant attenuation of mean spectral level at 1–8 kHz region and there was no significant change in this measure at 0–1 kHz. Mean spectral levels of vowel did not change significantly in mask-wearing conditions. LH1000 for connected speech significantly increased whilst wearing either a surgical mask or KN95 mask but no significant change in this measure was found for vowel. HNR was higher in the mask-wearing conditions than the no-mask condition. CPPS and vocal intensity did not change in mask-wearing conditions. These findings implied an attenuation effects of wearing these types of masks on the voice spectra with surgical mask showing less impact than the KN95.

Facemasks are an essential piece of personal protective equipment (PPE) and can be broadly categorized into respirators, medical masks (including surgical masks and procedure masks), and woven fabric (cloth) masks¹. Respirators and surgical masks provide different levels of barrier to prevent infectious transmission via aerosols and droplets². Masks with higher barrier levels (e.g. N95) are used in aerosol generating procedures (AGPs) and other high risk activities¹. During non-aerosol generating protocols, surgical masks offer a similar degree of protection to N95 masks against viral respiratory infections including coronaviruses in health care workers (HCWs)³. Although surgical masks do not provide the same level of protection as N95 masks, they prevent some aerosols and droplets from being released from phonation and respiratory activities, contributing to reducing the risk of transmission⁴. In the SARS-CoV2 pandemic (COVID-19), such masks have been recommended for use by not only HCWs but also the general public in areas with known or suspected widespread transmission, high population density, or settings where physical distancing cannot be effectively achieved⁵. Although masks are effective PPE⁴, wearing a mask negatively affects the physiological and psychological performance of HCWs⁶.

Masks also interfere with effective verbal communication. Certain masks particularly the N95 respirators can impact speech understanding by listeners⁷. Word intelligibility dropped between 1 and 17% while wearing respirators commonly used by HCWs in which N95 mask resulted in a mean (standard deviation, SD) of modified rhyme test (MRT) score of 83 (16.2)% compared to 92 (5.8)% in non-mask controls⁸. The use of N95 mask in background noise resulted in a significant decrease in speech perception accuracy⁹. Speaking while wearing a mask at longer distances decreases speech perception accuracy by an even greater magnitude than not wearing a mask¹⁰. A mask also physically creates a visual barrier precluding lip reading¹¹, precluding communication cues in people with hearing loss and communication disabilities such as aphasia¹². From a user's perspective, wearing masks increased perception of vocal effort, reduced auditory feedback, and difficult coordination of speech and breathing¹³. Understanding the aspects of the voice changes whilst wearing a mask is important so clinical decision, making and choice of mask is appropriate to meet infection control and optimal verbal communication.

Although it is believed that facemasks attenuate sound transmission like a low-pass filter^{10,14}, little information is available on voice characteristics whilst wearing a facemask. The scarce literature on the topic suggests possible changes in the speech spectrum. Mendel et al.¹⁵ compared speech spectral levels calculated as total root mean square (RMS) power from the Connected Speech Test (CST) stimuli produced by one speaker with and without wearing a surgical mask. They found a significant difference in the total RMS power between the two

Voice Research Laboratory, Faculty of Medicine and Health, D18, Susan Wakil Health Building, Camperdown Campus, The University of Sydney, Western Avenue, Sydney, NSW 2006, Australia. email: cate.madill@sydney.edu.au

conditions. However, the affected frequency band was not reported. Atcherson et al.¹¹ found that the total RMS values of speech signals from the CST stimuli were significantly higher when not wearing a mask compared to the conditions with a mask. They also did not mention which frequency range was affected by the mask. Goldin et al.¹⁴ utilised a head and torso simulator to play white noise via the model's mouth without a mask and with a surgical mask and a N95 respirator. They found that facemasks attenuated the sound levels at frequency regions between 2 and 7 kHz by 3–4 dB for the surgical mask and nearly 12 dB for the N95 mask compared with the non-mask condition. Their model lacked natural speech features while its face contour and surface were not similar to human face contour and skin, affecting fitting levels of the masks. However, based on their findings it seems reasonable to hypothesize that mask wearing would attenuate speech spectra at similar frequency bands.

Clear and natural speech production is necessary in accurate speech understanding and requires less listening effort than degraded speech¹⁶. Given the widespread use of facemasks in COVID-19 pandemic, it seems reasonable to further clarify the characteristics of the voice signal in speech of vocally healthy speakers who are wearing a mask. Given the above-mentioned findings of the modification of the speech spectra by the mask, the present study quantified the low- and high-frequency energy regions. Spectral analyses not only give information about the overall spectral shape that may be meaningful in speech perception¹⁷ but also provide important acoustic correlates of voice quality¹⁸. These spectral measures were selected as both the low and high frequency regions also contribute to speech recognition^{19,20}. Low-frequency spectral bands are important in recognizing vowels^{21,22} and voiced fricative consonants²³. High frequency spectral energy makes a significant contribution to speech recognition^{24,25} including the recognition of vowels²⁶, voiceless and voiced fricative consonants²⁷, spoken and sung text²⁸, and speech recognition in noise²⁹. It has also been shown that the high-frequency region provided perceptual cues for speaker identity³⁰ and gender identification³¹. In addition, the high frequency region plays an important role in the perception of clear speech: a shift of energy concentration toward higher frequency regions contributes to the clear speech effect for normal-hearing listeners³².

Presumably, the quality and audibility of the voice might also change whilst wearing a facemask as previous studies have observed voice changes in phonation with the mouth covered³³. This change may interfere with auditory-perceptual voice judgment by speech language pathologists (SLPs) and ear nose and throat specialists (ENTs). Dysphonic voice quality has also been proven to result in reduced comprehension of speech content by listeners^{34,35}. Wearing a mask may add to this effect by increasing the difficulty of understanding speech of an individual with dysphonic voice. Both voice quality and audibility can be effectively examined using acoustic analysis, which is a non-invasive objective assessment. Traditional acoustic measures of voice quality are based on frequency-based measurements³⁶ and include fundamental frequency (F0)³⁷ and noise (harmonic-to-noise ratio, HNR)^{38–41}. Amongst these, HNR has been used as a measure of vocal clarity⁴². The vocal signal can also be analysed based on spectral-based measurement of vowel and connected speech, which does not depend upon reliable tracking of vocal F0⁴³. The cepstral peak prominence (CPP) has been shown to have stronger weighted correlations with overall voice quality than other acoustic measures⁴⁴. Given that it is a measure of periodicity and harmonics strength, a signal with a strong harmonic structure would have a higher CPP than aperiodic signals⁴⁵. It has been considered a significant predictor of dysphonic severity⁴⁶. However, there are inherent limitations of cepstral analysis, that is, it is affected by vowel types and vocal intensity⁴⁷, vocal tasks⁴⁸, vocal tract⁴⁹, and the algorithm of software packages^{48,50}. Vocal audibility can be examined both by spectral energy at different frequency bands and sound intensity, which can also be estimated from the acoustic signals.

During the COVID-19 pandemic two types of masks were commercially available in Australia, including standard surgical mask and KN95 mask (China GB2626-2006)⁵¹. KN95 mask provided similar protection characteristics to N95 mask⁵². The major filtering and fitting characteristics of the KN95 mask as provided by 3 M⁵² were as follows: Filter performance $\geq 95\%$; Flow rate = 85 L/min; Inhalation resistance ≤ 350 Pa; Exhalation resistance ≤ 250 Pa; and Total inward leakage $< 8\%$. The total inward leakage indicates the amount of an aerosol that enters the mask via both filter penetration and face-seal leakage⁵². Presumably, the higher barrier level a mask can provide, the greater impact it would have on the voice signals. The aims of the present study were to (1) examine the acoustic characteristics of voice and speech whilst wearing either a surgical mask or a KN95 mask; and (2) compare the acoustic measures between the standard surgical mask and KN95 mask. We hypothesized that: (1) Low- and high-frequency spectral levels, HNR, CPPS, and vocal intensity would change during wearing these facemasks; and (2) Changes in these acoustic measures would be more pronounced with KN95 mask than with standard surgical mask.

Methods

Ethical approval. The voice and speech data analysed in this study was part of a larger project which was approved by the Human Research Ethics Committee of The University of Sydney (protocol number: 2020/399). Informed consent was obtained from all participants to participate in this study. Informed consent to publish was also obtained from a participant for publication of identifying information/image (Fig. 1) in an online open-access publication. The present study was implemented in accordance with relevant ethical guidelines and regulations. The measurement procedures used in this study conformed to the standards set by the latest revision of the Declaration of Helsinki.

Participants. Sixteen participants took part in this study (12 females, 4 males) with mean age = 43 years (range = 24–61). All were English speakers, non-smokers, and did not report any voice nor hearing problems at the time of the study. Participants were otolaryngologists (n = 2), practicing speech language pathologists (n = 13), and a registered nurse working in an Ear Nose and Throat clinic (n = 1).



Figure 1. Voice recording with a KN95 (left) and surgical mask (right).

Voice recordings. Due to social distancing measures during the COVID-19 pandemic, it was impossible for participants to have their voices recorded in the same recording environment. Therefore, voice recordings took place in a room at the practicing clinic of the participants with ambient noise ranging from 33.3 decibels (dBA) to 58.0 dBA. Participants were required to use their habitual voice to read the following standardised tasks: three repetitions of the sustained vowel /a/ for at least 10 s, the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) phrases⁵³, and the Rainbow Passage⁵⁴. These tasks were produced in three conditions with the speaker (1) not wearing a mask, (2) wearing a surgical mask, and (3) wearing a KN95 mask (Fig. 1). The order of conditions was randomised across speakers to minimize biases related to intra-speaker variability in phonation and potential compensation whilst wearing a mask. When wearing these masks, participants were required to use the highest level of fitting to ensure maximal barrier level. They were required to press the nose metal bar so that it fit tightly to the nose contour. The straps of the mask were securely placed behind the auricles and the lower side of the mask was pulled fully downward so that it covered the chin completely (Fig. 1). It has been known that in unfavourable/challenging speaking conditions, speakers may adapt a phonation style that helps improve clear phonation^{55,56}. Therefore, we required participants to maintain similar habitual voice in terms of pitch, loudness, and phonation type throughout recording sessions both with and without a mask to minimise intra-speaker variability in voice production.

All voice signals were captured using an AKG C520 ear-mounted microphone⁵⁷ placed at a constant distance of 6 cm, 45° off the mouth axis and were analog-to-digital converted using a professional external sound card (Roland Quadcapture⁵⁸) at 44.1 kHz and 16-bit resolution. The signals were processed and saved to a laptop computer using the Audacity sound editing software in *.wav format⁵⁹. Calibration of sound level in the voice signals was deemed unnecessary given that the data were used to test within-subject effects of mask and non-mask conditions.

Acoustic analysis. Voice samples were edited in Audacity to extract the middle 3 s of the sustained /a/ vowel, the 3rd CAPE-V phrase (CAPEV-3), and the 2nd and 3rd sentences of the Rainbow Passage (RP23). All acoustic data were measured using Praat version 6.0.39⁶⁰.

Mean spectral level in low (0–1 kHz) and high (1–8 kHz) frequency ranges. Spectral levels in the 0–1000 Hz and 1000–8000 Hz were measured in Praat for the /a/ vowel (averaged from three repeats) and RP23. 1000 Hz was the cut-off between the low- and high-frequency regions in this study as the spectral region above 1000 Hz has been frequently used in investigating the role of different spectral regions in speech perception²⁵. Consonant noise is mainly concentrated at frequency regions above this frequency⁶¹. Further, the 1000 Hz cut-off has been used in studies involving spectral characteristics of voice quality^{62–64}. The upper limit of 8 kHz was used as extended high frequency ranges above this frequency have minor value in speech perception⁶⁵. The protocols in Praat were as follows: From Analyse spectrum => To LTAS, set bandwidth=100 Hz and click OK. From Query => Get mean, then frequency bands were set with averaging method being “dB”. The output was then copied to an Excel spreadsheet for analysis.

Measures	ICC (<i>p</i>)	
	Single measures	Average measures
HNR	0.985 (0.000)	0.992 (0.000)
LH1000 vowel	1.000 (0.000)	1.000 (0.000)
LH1000 of RP23	1.000 (0.000)	1.000 (0.000)

Table 1. Inter-rater reliability of acoustic analyses. RP23 = The 2nd and 3rd sentences of the Rainbow Passage.

Low/high spectral energy ratio between 0–1 and 1–8 kHz (LH1000). We also evaluated the low/high energy ratio (reflecting spectral slope) which is a ratio of spectral energy levels between the low and high frequency ranges to investigate how this would be affected given the impact of mask-wearing on the speech spectrum. The low/high ratio using a 1000 Hz cut-off value (LH1000) has been used frequently in voice and speech research and has been shown to reflect voice quality^{62,63}, vocal load⁶⁴, sentence prominence in speech⁶⁶, and the effects of language⁶⁷.

The low/high energy ratio between spectral areas below 1 Hz and between 1–8 kHz was measured for the /a/ vowel (averaged from three repeats) and RP23 using the long-term average spectra (LTAS) function in Praat. The command to obtain this measure in Praat was as follows: From Analyse spectrum = > To LTAS, set bandwidth = 100 Hz and click OK. From Query = > Get slope, set averaging method = dB, low band = 0–1000 Hz, high band = 1000–8000 Hz and click OK. The value that Praat software provided was measured in dB.

Harmonics-to-noise ratio. Praat (version 6.0.39) was also used to measure harmonics-to-noise ratio (HNR) from the sustained /a/ vowel. The 3-s vowel sample was open and highlighted in Praat editing window from which HNR was obtained using the command Voice report within the Pulses menu. Data was averaged from three repeats. Prior to measurement of HNR, all edited vowel samples were signal-typed by the first author (D.D.N.) and a research assistant using criteria recommended by Titze⁶⁸ and Sprecher et al.⁶⁹ This was conducted using narrow-band spectrograms generated in Praat using settings described in Sprecher et al.⁶⁹. Signal typing was performed visually by comparing each spectrogram picture with the exemplar signal types. Signal typing was deemed necessary because the measurement of HNR relies on reliable estimation of F0, which is only feasible in type 1 and type 2 signals⁶⁹.

Cepstral peak prominence smoothed. The voice cepstrum is obtained by a Fourier transform of the logarithm power spectrum⁷⁰. A cepstral peak is identified within the dominant ‘harmonic’ corresponding to the fundamental period from which CPP is calculated as the amplitude between the peak and the regression line directly below it⁴⁵. Smoothing the individual cepstra before extracting the cepstral peak and calculating CPP can improve prediction accuracy¹⁸. CPP-smoothed (CPPS) was measured in Praat using settings as follows^{71,72}: Pitch floor (Hz) = 60, Time steps (s) = 0.002, Maximum frequency (Hz) = 5000, Pre-emphasis from (Hz) = 50, Time averaging window (s) = 0.01, Quefrency averaging window (s) = 0.001, Peak search pitch range (Hz) = 60–330, Tolerance (0–1) = 0.05, Interpolation = Parabolic, Subtract tilt before smoothing = No, Tilt line quefrency range (s) = 0.001–0.0 (= end), Line type = Straight, Fit method = Robust.

Vocal intensity. Vocal intensity (dB) was also measured from the vowel, the 3rd CAPEV phrase, and the 2nd and 3rd sentences of the Rainbow Passage using Praat with default settings. Intensity values were not calibrated to real sound pressure level as the purpose of the study was to examine within-speaker effects.

Quality check of voice recordings and reliability analysis. Because voice recordings took place in different clinic rooms with different levels of background noise, audio files were examined for signal-to-noise ratio (SNR) using a Praat script called Speech-to-noise ratio /Voice-to-noise ratio v.01.01⁷³. Only samples with a SNR greater than 30 dB were used for acoustic analyses⁷⁴.

The sound files of four participants in all conditions [$n = 4 \times 3$ conditions (no-mask, surgical mask, KN95) = 12], were randomly selected and analysed a second time by a research assistant for HNR and LH1000 to calculate inter-rater reliability using Intraclass Correlation Coefficient (ICC, two-way mixed, consistency type). The results are shown in Table 1, which indicate excellent reliability. ICC was 1.00 for LH1000 as the measurement of this was fully automated using edited voice samples. The slightly lower ICC values for HNR resulted from possible differences between the raters in selecting (highlighting) the vowel segment for HNR measurement in Praat editing window.

Statistical analyses. Data were managed in Microsoft Excel 365⁷⁵ and analysed using IBM SPSS Statistics v.25.0⁷⁶ and Prism v8.1.2⁷⁷ for Windows. One-way repeated-measures analysis of variance (ANOVA) was used to examine the effects across three conditions (no-mask, surgical mask, and KN95 mask) on acoustic measures. Significant main effects were evaluated with Bonferroni-adjusted tests. Prior to analyses, normal distribution of the data was examined using Kolmogorov–Smirnov tests⁷⁸. Mauchly’s test of sphericity was performed before ANOVA and, if sphericity assumptions were not met, a Greenhouse–Geisser adjustment was used. Effect size was calculated using partial Eta squared (η^2). Effect sizes of 0.01, 0.1, and 0.25 indicated small, medium, and large effects, respectively⁷⁹. Where normality assumption was not met, the Friedman test was used to compare data across non-mask, surgical mask, and KN95 conditions. A significance level of 0.05 was used.

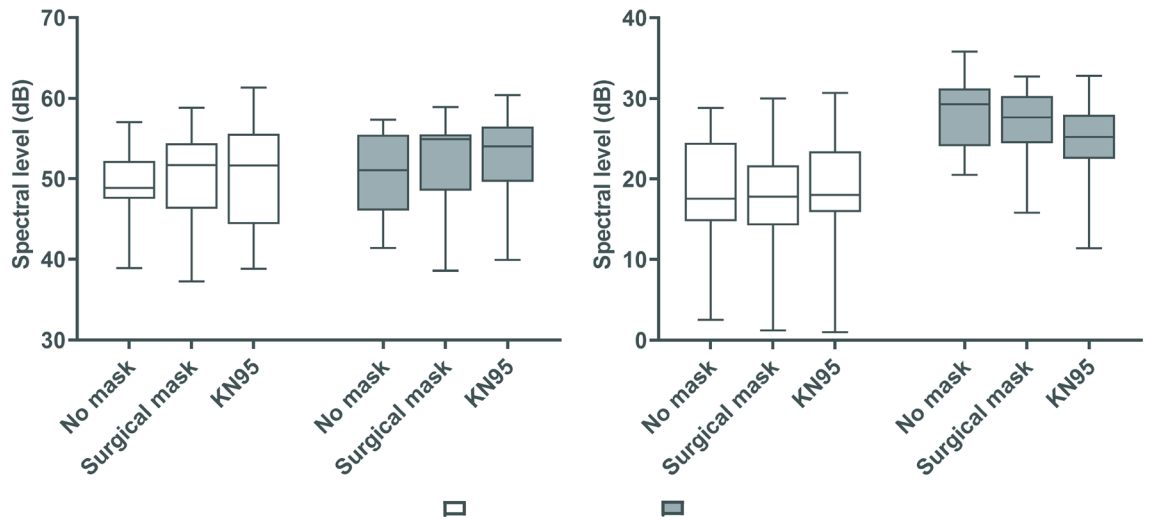


Figure 2. Mean spectral levels at low and high frequency regions across conditions. As sound pressure level was not calibrated, spectral levels were normalized so that the lowest value equalled zero. RP23 = The 2nd and 3rd sentences of the Rainbow Passage.

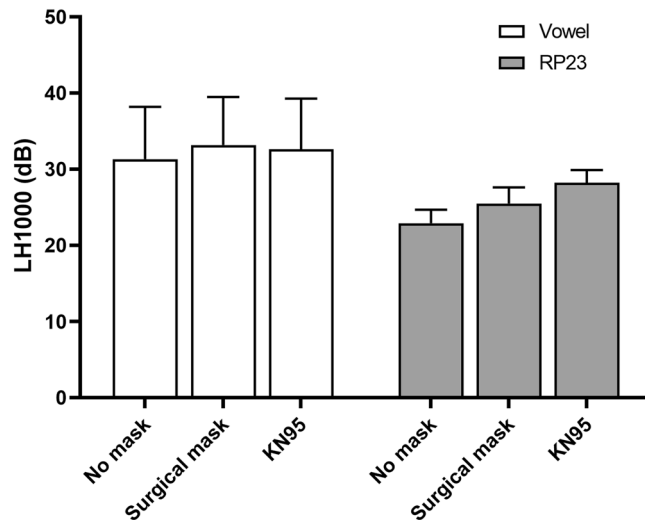


Figure 3. Mean low/high ratio at 1000 Hz for all conditions. Error bars indicate standard deviation. RP23 = The 2nd and 3rd sentences of the Rainbow Passage.

Results

Mean spectral levels at low and high frequency regions. *Mean spectral levels in 0–1 kHz region.* Figure 2 showed mean spectral levels at both frequency bands 0–1 kHz and 1–8 kHz. This figure shows that this spectral measure did not change across conditions for both vowel and connected speech. For sustained /a/ vowel phonation, no significant main effects of mask-wearing were found: $F(2, 22) = 0.396$, $p = 0.678$, partial $\eta^2 = 0.035$. For RP23, there was also no significant main effects of masks in the 0–1 kHz range $F(1.235, 13.588) = 0.808$, $p = 0.410$, partial $\eta^2 = 0.068$.

Mean spectral levels in 1–8 kHz region. Figure 2 shows mean spectral levels for both vowel and connected speech in the 1–8 kHz region. For vowel production, no significant main effects of wearing a facemask were observed $F(2, 22) = 0.024$, $p = 0.963$, partial $\eta^2 = 0.002$. However, for RP23, wearing a facemask affected mean spectral levels in the 1–8 kHz region: There was a significant main effect of mask-wearing on this measure $F(1.173, 11.735) = 16.951$, $p = 0.001$, partial $\eta^2 = 0.629$. Post-hoc tests showed that, compared with the non-mask condition, the KN95 mask attenuated the spectral levels in the 1–8 kHz region by 5.2 dB ($p = 0.005$) while the surgical mask attenuated the spectral levels in this region by 2.0 dB ($p = 0.014$).

Low/high spectral ratio (LH1000). LH1000 was calculated for the /a/ vowel and RP23. Figure 3 shows mean and SD of LH1000 for both tasks. One-way repeated-measures ANOVA was calculated to compare data



Figure 4. Harmonics-to-noise ratio in all conditions. Error bars = standard deviation.

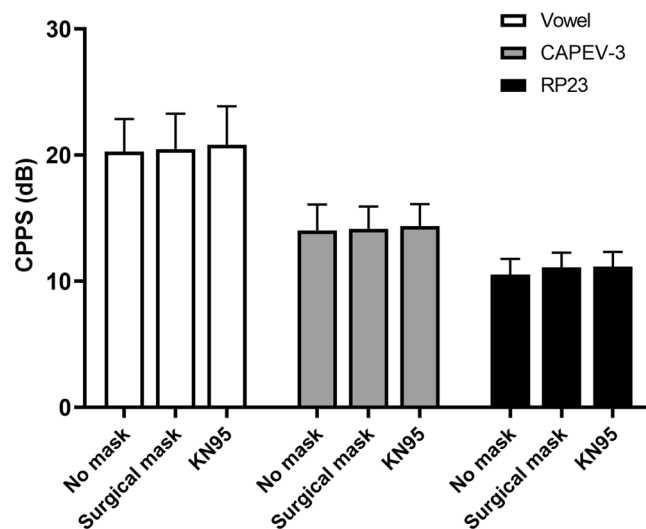


Figure 5. Cepstral peak prominence smoothed (CPPS) for all conditions. CAPEV-3 = The third CAPE-V phrase. RP23 = The 2nd and 3rd sentences of the Rainbow Passage.

across non-mask, surgical mask, and KN95 mask. For the sustained vowel, no significant main effects were observed: $F(2, 22) = 0.949$, $p = 0.402$, partial $\eta^2 = 0.079$. In Fig. 3, LH1000 for the vowel produced did not change significantly across non-mask and mask conditions.

For RP23, a significant main effect was present: $F(1.279, 14.073) = 84.346$, $p = 0.000$, partial $\eta^2 = 0.885$. Figure 3 shows that LH1000 of RP23 was the lowest for the non-mask [mean (SD) = 23.0 (1.7) dB], higher for the surgical mask [mean (SD) = 25.5 (2.2) dB], and highest for the KN95 mask condition [mean (SD) = 28.2 (1.7) dB]. Pairwise Bonferroni-adjusted comparisons showed that wearing a KN95 mask increased the LH1000 of RP23 by 5.2 dB ($p = 0.000$) and wearing a surgical mask increased the LH1000 of RP by almost 2.5 dB ($p = 0.000$).

Harmonics-to-noise ratio. Figure 4 shows HNR in all experimental conditions. HNR was compared across non-mask, surgical mask, and KN95 mask conditions using one-way repeated-measures ANOVA; significant main effects were found: $F(2, 22) = 14.749$, $p = 0.000$, partial $\eta^2 = 0.573$. Post-hoc Bonferroni-adjusted tests showed that HNR significantly increased with wearing either a surgical mask or KN95 mask. In the non-mask condition, the mean (SD) of HNR was 25.0 (3.5) dB. HNR values increased significantly to 27.3 (4.5) dB ($p = 0.004$) and 28.4 (4.1) dB ($p = 0.000$) when wearing a surgical mask and KN95 mask, respectively.

Cepstral peak prominence smoothed. Figure 5 shows that CPPS did not change across all conditions for all three vocal tasks. One-way repeated-measures ANOVA comparing no-mask, surgical mask, and KN95 mask showed no statistically significant effects of mask-wearing on CPPS of vowel [$F(2, 22) = 0.695$, $p = 0.51$, partial $\eta^2 = 0.059$], CAPEV-3 [$F(1.326, 13.260) = 0.013$, $p = 0.954$, partial $\eta^2 = 0.001$], nor RP23 [$F(1.218, 13.393) = 4.313$, $p = 0.051$, partial $\eta^2 = 0.282$].

Tasks	Non-mask	Surgical mask	KN95	<i>p</i>
Vowel	64.0 (7.0)	65.8 (7.2)	66.7 (7.5)	0.188
CAPEV-3	63.9 (4.9)	64.5 (5.7)	64.8 (6.1)	0.595
RP23	59.7 (5.0)	61.3 (5.2)	60.9 (5.1)	0.290

Table 2. Mean (SD) of vocal intensity (dB) in each condition and *p* values from repeated-measures ANOVA. CAPEV-3, The third CAPEV phrase; RP23, The 2nd and 3rd sentences of the Rainbow Passage.

Vocal intensity. The intensity of the vowel increased slightly in both mask conditions however the changes were not statistically significant (Table 2). Similarly, there were no main effects of mask-wearing on vocal intensity of the CAPEV-3 and RP23.

Data from all three conditions were used to calculate the Pearson's correlation coefficient (*r*) between vocal intensity and CPPS. There were moderate correlations between vocal intensity and CPPS (Vowel: $r = 0.595$, $p = 0.000$; CAPEV-3: $r = 0.522$, $p = 0.000$; and RP23: $r = 0.366$, $p = 0.014$).

Discussion

In this study we hypothesized that voice quality measures (e.g. spectral levels at low- and high-frequency, HNR and CPPS) and audibility measure (vocal intensity) would change during wearing either a surgical mask or KN95 mask. The data confirmed our hypotheses for spectral characteristics, showing a significant decrease in mean spectral levels at high frequency regions (1–8 kHz) and an increase in LH1000 (implying steeper spectral slope) when wearing either a surgical mask or KN95 mask. HNR improved significantly in both mask-wearing conditions. However, CPPS and vocal intensity did not change.

The decreased spectral levels in the 1–8 kHz region was in agreement with a previous modelling experiment. In Goldin et al.'s study¹⁴, sound levels between 2 and 7 kHz was attenuated by 3–4 dB with the surgical mask and by approximately 12 dB with the N95 compared with the non-mask condition. In the present study we found that the spectral levels at the 1–8 kHz region was attenuated by 2.0 dB by surgical mask and 5.2 dB by KN95 mask. The findings may be explained as degrading/attenuating effects of mask on spectral levels. These appeared to depend upon the filter performance and the level of fitting of the masks in which the KN95 mask outperforms standard surgical masks. Although both our and Goldin et al.'s studies found greater impact of masks with higher barrier levels (N95/KN95), levels of fitting may have varied in Goldin et al.'s study as they used a model in which face contour and skin characteristics were not similar to human subjects. Perceptual studies have shown that masks with higher filtering characteristics impact more on speech i.e. respirators decreased speech intelligibility scores while surgical masks did not show significant differences in speech intelligibility compared with no-mask condition⁸. Mendel et al.¹⁵ also found that surgical masks did not have a negative effect on speech understanding in both normal hearing and hearing loss groups. Taken together, these data showed that KN95 masks differed from standard surgical masks in level of degradation of the signals, hence listener's mechanisms of adaptation in speech perception might be different when listening to speech produced whilst wearing these masks. This would imply that surgical masks seem to be a more appropriate choice over the KN95 (similarly to N95 masks in specifications⁵²), given that there is insufficient evidence for selecting N95 masks over surgical masks for protecting HCWs against infectious disease transmission⁸⁰.

The spectral energy in this study was calculated using data combined from both genders because of the within-subject study design. Between-speaker and between-gender variability in the level of changes in spectral characteristics as a result of the mask was therefore not examined. It is well known that speech spectra carry information of both the larynx (voice source) and the vocal tract (filter)⁸¹ and reflect gender characteristics³¹. Apart from common spectral features across speakers⁸², there may also be variabilities in characteristics of the speech spectrum across speakers. Between-speaker variability has been observed in source spectral shape and spectral noise in both genders, F3 and F4 and formant dispersion in female voices; and spectral slope in the higher frequencies (from the fourth harmonic to the harmonic closest to 2 kHz and from the harmonic closest to 2 kHz to the harmonic closest to 5 kHz) for male voices⁸². These are within the 1–8 kHz range investigation in the present study. As such, the impact of mask on phonation may not be the same for all speakers. For these reasons spectral measures should be analysed separately for each gender. However, the limited sample size (12 females and 4 males) did not allow sufficient statistical power for each group. Testing spectral measures that characterize between-speaker variability can help evaluate the impact more specifically for each gender. Speaker-specific and gender-specific source-filter characteristics and phonation and articulation strategy with and without wearing a mask were therefore not examined. It is not known whether male or female voices were affected similarly by wearing a mask.

It is possible that the speakers in this study used an individual strategy to adapt their phonation style, which may also account for the findings. Although the participants were instructed to keep their phonation style constant, it was impossible to control for this. Adaptation in mask-wearing condition may include unconsciously increasing vocal projection to compensate for presence of the mask. Increased HNR has been observed in speakers with increased vocal intensity⁸³. However, in the present study vocal intensity was not significantly different across the three conditions. It was therefore not possible to confirm whether the improved HNR resulted from vocal adaptation. In addition, why HNR was higher in mask-wearing conditions whilst CPPS remained unchanged was unclear. Although HNR and CPPS actually reflect different phenomena in voice quality⁴⁹, the non-significant changes in CPPS and vocal intensity across conditions did not appear to support

the compensation assumption. In addition, phonation compensation in mask wearing may include increased vocal effort, as reported in a previous study¹³. Increased vocal effort seems associated with a different trend of spectral change. In vocally healthy speakers McKenna and Stepp⁸⁴ observed that typical phonation style had the highest L/H ratio (a ratio of low to high spectral energy with cut-off at 4 kHz) and this measure decreased steadily from mild to maximal effort (i.e. decreased spectral slope in vocal effort). Meanwhile, the two mask-wearing conditions in the present study showed an opposite trend in spectral slope as expressed in the LH1000 findings.

Regardless of the actual mechanisms, the findings implied that several important high-frequency components might change during wearing these masks. These include the higher formants that are necessary to identify linguistic contents of segmental units. Previous studies have shown that high frequency energy is important to recognize vowels²⁶. As the second formant for the vast majority of English vowels exist above 1 kHz⁸⁵, it is likely that correct identification of vowels will be compromised if frequencies above 1 kHz are attenuated by the wearing of a facemask. The recognition of voiceless consonants (e.g. /s/ and /f/) also requires detection of high frequency regions above 1 kHz⁶¹. For example, the noise component produced in /s/ is centred between 5 and 6 kHz⁸⁶. The changes in higher frequency regions also imply that the quality of the speech signal will be affected as the frequencies of voiceless plosives, fricatives and affricates occur within 2–4 kHz, as does place of articulation⁸⁶. Although the findings were not specific to any specific segmental acoustic measures of speech recognition, accurate recognition of speech may be compromised when the speaker is wearing a mask, as demonstrated in some previous studies^{8,9}. It is important to note that the data in this study was recorded in sound-proofed or quiet conditions, while in reality, people are wearing masks in noisy environments, which will most likely further reduce the audibility of the speech signal. The degradation of speech spectral signals, associated with background noise and with acoustically reflective surfaces (e.g. hard uncarpeted floors) will further reduce audibility of the signal and speech perception, making it difficult for listeners, especially those with hearing loss, to perceive and understand the speech signal in such high noise level environments. Further research using landmarks for speech recognition⁸⁷ would help clarify the specific impact of wearing a mask on important linguistic content e.g. vowels and consonants.

The improved HNR in mask-wearing condition agreed with previous findings that people with voice disorders sound less dysphonic when their mouth is covered. For example, hand-over-mouth has been believed to reduce vocal constriction and increase frontal/oral resonance, facilitating more effective voice production⁸⁸. Wearing a semi-occluded ventilation mask (SOVM)⁸⁹ in people with normal voices and functional dysphonia also showed significant differences between SOVM and control in aerodynamic, electroglottographic, and acoustic measures. The improvement in HNR can also be explained using similar mechanisms as for spectral findings i.e. filtering out/attenuation of glottal noise from the signals by the masks. Given that glottal noise components in voice typically appear in high frequency range^{90,91}, noise attenuation by the masks may make spectral energy at low frequency more dominant relative to high frequency, hence improving HNR. In addition, the masks might affect vocal tract resonance, leading to changes in nasal-oral coupling with more nasal resonance involved because of increased impedance at the mouth opening whilst wearing the masks. A previous study has shown that when phonation changed from vowel to nasal, HNR increased by 1.5 dB⁴⁹.

We did not observe significant changes in vocal intensity for all vocal tasks in mask conditions. This agreed with our non-significant findings on CPPS and was contrary to our expectation that wearing a mask would attenuate intensity. This is possibly due to the calculation of intensity by the Praat program, as default settings of intensity are calculated on frequency settings. In this study, intensity calculations were conducted on frequencies between 75 and 500 Hz. Additionally, although the participants were asked to maintain approximately similar vocal intensity across conditions, they might inadvertently increase their volume slightly in mask conditions to compensate for the altered feedback created by the mask. In addition, the intensity values were obtained from the signals recorded using a cardioid (directional) microphone at distance of 6 cm from the mouth. Whether this applied to perception/audibility of loudness/vocal intensity at real communication distances is not known.

Several issues need consideration in future studies. The main limitation of this study was that it did not evaluate speech intelligibility as the vocal tasks used were standardized for assessing voice quality and not for speech intelligibility. Future studies should compare speech perception across people with normal hearing and hearing impairments using speech materials produced with and without wearing different mask types. This study did not calibrate sound level as its aim was to test within-subject factors; hence, real intensity level was deemed unnecessary. Future studies may measure real sound pressure levels in non-mask and mask conditions to further clarify the impact of mask-wearing on speech sound audibility. Findings presented were for standard surgical mask and KN95 mask only and may not be generalizable to other types of masks. In light of the use of cloth masks amongst members of the community, investigating the effect of fabric masks on both acoustic and auditory-perceptual measures is recommended.

Conclusion

This study showed that the recorded acoustic voice signal changed whilst wearing either a standard surgical mask or a KN95 mask. Low/high spectral ratio increased i.e. spectral slope was steeper, which resulted from an attenuation of mean spectral levels in the 1–8 kHz regions relative to the 0–1 Hz regions. These spectral measures changed to a greater extent for KN95 mask than for surgical mask, suggesting that filtering and fitting characteristics of masks might determine the level of the voice and speech signal degradation. Alternatively, findings may reflect that the degree to which a speaker wearing a mask adjusts their phonation style is different across masks differing in characteristics. The findings appeared to imply that surgical masks might be a more relevant choice over the KN95 in COVID-19 pandemic to minimize the impact on communication.

HNR was improved in both surgical mask and KN95 conditions, implying possible filtering effects of these masks on spectral/glottal noise. This had implications in voice assessment in situations where patients are wearing a mask.

This study did not find significant changes in both CPPS and vocal intensity in the mask condition, implying that the design (e.g. recording setup) of the present study did not allow detection of changes in these measures, or that participants may have altered their vocal production in response to the mask.

Received: 17 September 2020; Accepted: 19 February 2021

Published online: 11 March 2021

References

- Institute of Medicine. *Reusability of Facemasks During an Influenza Pandemic: Facing the Flu* (National Academies Press, 2006).
- Gralton, J. & McLaws, M. L. Protecting healthcare workers from pandemic influenza: N95 or surgical masks?. *Crit. Care Med.* **38**, 657–667 (2010).
- Bartoszko, J. J. *et al.* Medical masks vs N95 respirators for preventing COVID-19 in healthcare workers: A systematic review and meta-analysis of randomized trials. *Influenza Other Respir. Viruses* **14**, 365–373 (2020).
- Leung, N. H. L. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat. Med.* **26**, 676–680 (2020).
- World Health Organization. *Coronavirus Disease (COVID-19) Advice for the Public: When and How to Use Masks* <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/when-and-how-to-use-masks> (2020).
- Johnson, A. T. Respirator masks protect health but impact performance: A review. *J. Biol. Eng.* **10**, 4 (2016).
- Palmiero, A. J. *et al.* Speech intelligibility assessment of protective facemasks and air-purifying respirators. *J. Occup. Environ. Hyg.* **13**, 960–968 (2016).
- Radonovich, L. J. Jr. *et al.* Diminished speech intelligibility associated with certain types of respirators worn by healthcare workers. *J. Occup. Environ. Hyg.* **7**, 63–70 (2010).
- Thomas, F. *et al.* Does wearing a surgical facemask or N95-respirator impair radio communication?. *Air Med. J.* **30**, 97–102 (2011).
- Coyne, K. M. *et al.* Respirator performance ratings for speech intelligibility. *Am. Ind. Hyg. Assoc. J.* **59**, 257–260 (1998).
- Atcherson, S. R. *et al.* The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss. *J. Am. Acad. Audiol.* **28**, 58–67 (2017).
- Chodosh, J., Weinstein, B. E. & Blustein, J. Face masks can be devastating for people with hearing loss. *BMJ* **370**, m2683 (2020).
- Ribeiro, V. V. *et al.* Effect of wearing a face mask on vocal self-perception during a pandemic. *J. Voice* <https://doi.org/10.1016/j.jvoice.2020.09.006> (2020).
- Goldin, A., Weinstein, B. & Shiman, N. How do medical masks degrade speech perception?. *Hear. Rev.* **27**, 8–9 (2020).
- Mendel, L. L., Gardino, J. A. & Atcherson, S. R. Speech understanding using surgical masks: A problem in health care?. *J. Am. Acad. Audiol.* **19**, 686–695 (2008).
- Wild, C. J. *et al.* Effortful listening: The processing of degraded speech depends critically on attention. *J. Neurosci.* **32**, 14010–14021 (2012).
- Hillenbrand, J. M., Houde, R. A. & Gayvert, R. T. Speech perception based on spectral peaks versus spectral shape. *J. Acoust. Soc. Am.* **119**, 4041–4054 (2006).
- Hillenbrand, J. & Houde, R. A. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J. Speech Hear. Res.* **39**, 311–321 (1996).
- Healy, E. W. & Bacon, S. P. Measuring the critical band for speech. *J. Acoust. Soc. Am.* **119**, 1083–1091 (2006).
- Apoux, F. & Bacon, S. P. Differential contribution of envelope fluctuations across frequency to consonant identification in quiet. *J. Acoust. Soc. Am.* **123**, 2792 (2008).
- Dubno, J. R. & Dorman, M. F. Effects of spectral flattening on vowel identification. *J. Acoust. Soc. Am.* **82**, 1503–1511 (1987).
- Sakayori, S. *et al.* Critical spectral regions for vowel identification. *Neurosci. Res.* **43**, 155–162 (2002).
- Hughes, G. W. & Halle, M. Spectral properties of fricative consonants. *J. Acoust. Soc. Am.* **28**, 303–310 (1956).
- MacDonald, E. N., Pichora-Fuller, M. K. & Schneider, B. A. Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech. *Hear. Res.* **261**, 63–66 (2010).
- Hazan, V. & Markham, D. Acoustic-phonetic correlates of talker intelligibility for adults and children. *J. Acoust. Soc. Am.* **116**, 3108–3118 (2004).
- Donai, J. J. & Paschall, D. D. Identification of high-pass filtered male, female, and child vowels: The use of high-frequency cues. *J. Acoust. Soc. Am.* **137**, 1971–1982 (2015).
- Vitela, A. D., Monson, B. B. & Lotto, A. J. Phoneme categorization relying solely on high-frequency energy. *J. Acoust. Soc. Am.* **137**, EL65–EL70 (2015).
- Monson, B. B., Lotto, A. J. & Story, B. H. Gender and vocal production mode discrimination using the high frequencies for speech and singing. *Front. Psychol.* **5**, 1239 (2014).
- Macho, D. & Cheng, Y. M. On the use of wideband signal for noise robust ASR. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing* 109–112.
- Boyd-Pratt, H. A. & Donai, J. J. The perception and use of high-frequency speech energy: Clinical and research implications. *Perspect. ASHA Spec. Interest Groups* **5**, 1–9 (2020).
- Donai, J. J. & Halbritter, R. M. Gender identification using high-frequency speech energy: Effects of increasing the low-frequency limit. *Ear Hear.* **38**, 65–73 (2017).
- Maniwa, K., Jongman, A. & Wade, T. Acoustic correlates of intelligibility enhancements in clearly produced fricatives. In *8th Annual Conference of the International Speech Communication Association, Interspeech 2007* 1581–1584.
- Awan, S. N. *et al.* Effects of a variably occluded face mask on the aerodynamic and acoustic characteristics of connected speech in patients with and without voice disorders. *J. Voice* **33**(809), e801–809 (2019).
- Porcaro, C. K. *et al.* Effect of dysphonia and cognitive-perceptual listener strategies on speech intelligibility. *J. Voice* **34**(5), 806.e7–806.e18 (2019).
- Ishikawa, K. *et al.* The effect of background noise on intelligibility of dysphonic speech. *J. Speech Lang. Hear. Res.* **60**, 1919–1929 (2017).
- Qi, Y. & Hillman, R. E. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *J. Acoust. Soc. Am.* **102**, 537–543 (1997).
- Baken, R. J. & Orlikoff, R. F. *Clinical Measurement of Speech and Voice* 2nd edn. (Singular Thomson Learning, 2000).
- Mathew, M. M. & Bhat, J. S. Soft phonation index—A sensitive parameter?. *Indian J. Otolaryngol. Head Neck Surg.* **61**, 127–130 (2009).
- Akif Kiliç, M. *et al.* The effects of vowels on voice perturbation measures. *J. Voice* **18**, 318–324 (2004).
- Godino-Llorente, J. I. *et al.* Acoustic analysis of voice using WPCVox: A comparative study with Multi Dimensional Voice Program. *Eur. Arch. Otorhinolaryngol.* **265**, 465–476 (2008).

41. Heman-Ackah, Y. D., Michael, D. D. & Goding, G. S. Jr. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice* **16**, 20–27 (2002).
42. Warhurst, S. *et al.* The vocal clarity of female speech-language pathology students: An exploratory study. *J. Voice* **26**, 63–68 (2012).
43. Awan, S. N., Roy, N. & Dromey, C. Estimating dysphonia severity in continuous speech: Application of a multi-parameter spectral/cepstral model. *Clin. Linguist. Phon.* **23**, 825–841 (2009).
44. Maryn, Y. *et al.* Acoustic measurement of overall voice quality: A meta-analysis. *J. Acoust. Soc. Am.* **126**, 2619–2634 (2009).
45. Hillenbrand, J., Cleveland, R. A. & Erickson, R. L. Acoustic correlates of breathy vocal quality. *J. Speech Hear. Res.* **37**, 769–778 (1994).
46. Awan, S. N. & Roy, N. Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. *Clin. Linguist. Phon.* **20**, 35–49 (2006).
47. Awan, S. N., Giovinco, A. & Owens, J. Effects of vocal intensity and vowel type on cepstral analysis of voice. *J. Voice* **26**(670), e615–620 (2012).
48. Madill, C. *et al.* Comparison of cepstral peak prominence measures using the ADSV, SpeechTool and VoiceSauce acoustic analysis programs. *Acoust. Aust.* **46**, 215–226 (2018).
49. Madill, C. *et al.* The impact of nasalance on Cepstral peak prominence and harmonics-to-noise ratio. *Laryngoscope* **129**, E299–E304 (2019).
50. Phadke, K. V. Selected Topics in Laryngeal, Perceptual and Acoustic Assessments of Human Voice: Videokymographic Evaluations of Vocal Folds and Investigations of Teachers' Voices. PhD thesis, Palacký University Olomouc (2018).
51. Australian Government Department of Health. *The Use of Face Masks and Respirators in the Context of Covid-19* <https://www.health.gov.au/sites/default/files/documents/2020/05/the-use-of-face-masks-and-respirators-in-the-context-of-covid-19.pdf> (2020).
52. 3M. *Comparison of FFP2, KN95, and N95 Filtering Facepiece Respirator Classes* <https://multimedia.3m.com/mws/media/17915000/comparison-ffp2-kn95-n95-filtering-facepiece-respirator-classes-tb.pdf> (2020).
53. Kempster, G. B. *et al.* Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *Am. J. Speech Lang. Pathol.* **18**, 124–132 (2009).
54. Fairbanks, G. *Voice and Articulation DRILLBOOK* 2nd edn. (Harper & Row, 1960).
55. Krause, J. C. & Braida, L. D. Acoustic properties of naturally produced clear speech at normal speaking rates. *J. Acoust. Soc. Am.* **115**, 362–378 (2004).
56. Hazan, V. *et al.* Clear speech adaptations in spontaneous speech produced by young and older adults. *J. Acoust. Soc. Am.* **144**, 1331 (2018).
57. AKG Acoustics. *C520* <https://www.ake.com/Microphones/Headset%20Microphones/C520.html> (2018).
58. Roland Corp. *Quad-Capture—USB 2.0 Audio Interface* <https://www.roland.com/au/products/quad-capture/> (2019).
59. Audacity Team. *Audacity(R): Free Audio Editor and Recorder [Computer application]*, <https://www.audacityteam.org/> (2019).
60. Boersma, P. & Weenink, D. *Praat: Doing Phonetics by Computer* <http://www.fon.hum.uva.nl/praat/> (2018).
61. Strevens, P. Spectra of fricative noise in human speech. *Lang. Speech* **3**, 32–49 (1960).
62. Frøkjær-Jensen, B. & Prytz, S. Registration of voice quality. *Brüel and Kjær – Tech. Rev.* **3**, 3–17 (1976).
63. Kitzing, P. LTAS criteria pertinent to the measurement of voice quality. *J. Phon.* **14**, 477–482 (1986).
64. Rantala, L. *et al.* Working-day effects on the spectral characteristics of teaching voice. *Folia Phoniatr Logop* **50**, 205–211 (1998).
65. Monson, B. B. & Caravello, J. The maximum audible low-pass cutoff frequency for speech. *J. Acoust. Soc. Am.* **146**, EL496 (2019).
66. Kakouros, S., Räsänen, O. & Alku, P. Comparison of spectral tilt measures for sentence prominence in speech—Effects of dimensionality and adverse noise conditions. *Speech Commun.* **103**, 11–26 (2018).
67. Bahmanbiglu, S. A., Mojiri, F. & Abnavi, F. The Impact Of Language On Voice: An LTAS study. *J. Voice* **31**, 249.e249–249.e212 (2017).
68. Titze, I. R. *Workshop on Acoustic Voice Analysis: Summary Statement* (National Center for Voice and Speech, 1995).
69. Sprecher, A. *et al.* Updating signal typing in voice: Addition of type 4 signals. *J. Acoust. Soc. Am.* **127**, 3710–3716 (2010).
70. Noll, A. M. Cepstrum pitch determination. *J. Acoust. Soc. Am.* **41**, 293–309 (1967).
71. Watts, C. R., Awan, S. N. & Maryn, Y. A comparison of cepstral peak prominence measures from two acoustic analysis programs. *J. Voice* **31**, 387.e381–387.e310 (2017).
72. Phadke, K. V. *et al.* Cepstral and perceptual investigations in female teachers with functionally healthy voice. *J. Voice* **34**, 485.e433–485.e443 (2018).
73. Maryn, Y. *Recording Quality: Speech-to-Noise Ratio and Voice-to-Noise Ratio* <https://www.phonanium.com/product/recording-quality/> (2020).
74. Deliyiski, D. D., Shaw, H. S. & Evans, M. K. Adverse effects of environmental noise on acoustic voice quality measurements. *J. Voice* **19**, 15–28 (2005).
75. Microsoft. *Microsoft Excel* <https://www.microsoft.com/en-us/microsoft-365/excel> (2020).
76. IBM Corp. *IBM SPSS Software* <https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software> (2018).
77. GraphPad Software. *Prism 8*, <<https://www.graphpad.com/scientific-software/prism/>> (2018).
78. Massey, F. J. The Kolmogorov–Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
79. Murphy, K. R. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests* (Routledge, 2014).
80. Smith, J. D. *et al.* Effectiveness of N95 respirators versus surgical masks in protecting health care workers from acute respiratory infection: A systematic review and meta-analysis. *CMAJ* **188**, 567–574 (2016).
81. Fant, G. Glottal source and excitation analysis. *Q. Progress Status Rep.* **20**, 85–107 (1979).
82. Lee, Y., Keating, P. & Kreiman, J. Acoustic voice variation within and between speakers. *J. Acoust. Soc. Am.* **146**, 1568 (2019).
83. Sampaio, M. *et al.* Effects of fundamental frequency, vocal intensity, sample duration, and vowel context in cepstral and spectral measures of dysphonic voices. *J. Speech Lang. Hear. Res.* **63**, 1326–1339 (2020).
84. McKenna, V. S. & Stepp, C. E. The relationship between acoustical and perceptual measures of vocal effort. *J. Acoust. Soc. Am.* **144**, 1643 (2018).
85. Kent, R. D. & Vorperian, H. K. Static measurements of vowel formant frequencies and bandwidths: A review. *J. Commun. Disord.* **74**, 74–97 (2018).
86. Ladefoged, P. *A Course in Phonetics* 6th edn. (Wadsworth, Cengage, 2011).
87. Chen, F. & Loizou, P. C. Contribution of consonant landmarks to speech recognition in simulated acoustic-electric hearing. *Ear Hear.* **31**, 259–267 (2010).
88. Behrman, A. & Haskell, J. *Exercises for Voice Therapy* 3rd edn. (Plural Publishing Inc, 2020).
89. Frisancho, K. *et al.* Immediate effects of the semi-occluded ventilation mask on subjects diagnosed with functional dysphonia and subjects with normal voices. *J. Voice* **34**, 398–409 (2020).
90. Yanagihara, N. Significance of harmonic changes and noise components in hoarseness. *J. Speech Hear. Res.* **10**, 531–541 (1967).
91. Shoji, K. *et al.* High-frequency components of normal voice. *J. Voice* **5**, 29–35 (1991).

Author contributions

D.D.N. was involved in research question identification, study planning, protocol preparation, data analysis, data interpretation, manuscript writing and editing, and graphic works. P.M. was involved in study planning, result interpretation, and manuscript reviewing and editing. Donna Thomas was involved in problem identification and manuscript editing. A.P. was involved in problem identification, result interpretation, and manuscript editing. M.D. was involved in problem identification and manuscript reviewing and editing. D.N. interpreted findings and edited the manuscript. A.C. was involved in data collection, data analysis, and manuscript editing. C.M. was involved in problem identification, study planning, protocol preparation, data collection, and manuscript writing and editing. All authors reviewed and approved the manuscript prior submission.

Funding

This study was supported by Doctor Liang Voice Program of The University of Sydney.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021