Perceptual and Acoustic Characteristics of Speech Clarity with and Without a Facemask

Duy Duong Nguyen (✉ duong.nguyen@sydney.edu.au)

University of Sydney

Antonia Chacon

University of Sydney

Christopher Payten

University of Sydney

Rebecca Black

University of Sydney

Meet Sheth

University of Sydney

Patricia McCabe

University of Sydney

Daniel Novakovic

University of Sydney

Catherine Madill

University of Sydney

Research Article

Keywords: Facemask, speech clarity, fricative signal, vowel amplitude, formant amplitude.

Posted Date: July 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-731592/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Some studies have found that the speech of speakers wearing facemasks has reduced intelligibility. Although it has been found that facemasks attenuated high-frequency energy, no study has examined the effects of masks on spectral characteristics of vowels or voiceless fricative consonants. The present study investigated auditory perceptual rating of speech clarity and acoustic-phonetic measures of vowels and voiceless fricative consonant production in 16 health care workers who produced standardized voice tasks without and with wearing either a standard surgical mask or a KN95 mask. Voice samples were perceptually rated for speech clarity and were acoustically analysed for root-mean-square amplitude (A_{RMS}), spectral moments of two voiceless fricatives /s/ and /ʃ/, and A_{RMS} and amplitude of the first three formants (A1, A2, and A3). Speech produced whilst wearing either a surgical or KN95 mask was significantly less clear than without a mask, with KN95 showing greater impact than surgical masks. In both fricatives, A_{RMS} was lower in the surgical mask and KN95 mask conditions compared to the non-mask condition. None of the amplitude measures of vowels were affected by facemasks. Linear regression models indicated that perception of speech produced by mask users was mainly affected by modification of voiceless fricative consonant characteristics.

1. Introduction

The COVID -19 pandemic has instigated widespread use of facemasks both within the healthcare sector and more widely in the general population. Facemasks are recommended as a method of reducing droplet and/or aerosol-based transmission of infection ¹. There are three main types of facemasks offering differing levels of protection. Fluid resistant medical masks (surgical masks) are worn both by healthcare workers (HCW) and the community. They fit loosely on the face and are designed to reduce the spread of large droplets. Respirators including the N95 type are high performance filtering masks designed to be tightly attached and prevent against inhalation of small airborne particles ². WHO recommends use of surgical masks in lower risk situations and respirators in high risk contexts, such as frontline HCWs who may be directly involved in aerosol generating procedures ². Thirdly, cloth masks, the use of which has become widespread due to the scarcity of high-quality masks, are reported to provide some level of protection in studies of other respiratory diseases, but cannot provide the same level of effectiveness as medical masks ³.

Although masks are effective personal protective equipment (PPE) ⁴, wearing a mask may negatively affect the physiological and psychological performance of the wearer ⁵. From the users' perspective, self-perception of vocal effort, fatigue, discomfort and coordination of phonation and breathing have been reported by those wearing a facemask ⁶. Mask-wearers pause more frequently in connected speech as opposed to when not wearing a mask ⁷. From the listeners' perspective, speech produced whilst wearing a facemask can result in degraded signals with impacts on speech perception ⁸. When the acoustic signal is degraded, the frequency and amplitude of signal components are dampened, masked, or altered, and listeners must increase their cognitive work to interpret the signal which results in longer processing time

^{9,10}. The use of a mask becomes more problematic when listeners have hearing loss ¹¹, communication disorders ¹², English as a second language (ESL) ¹³, or communicate in background noise ^{14,15}.

Previous research has examined the impact of face masks upon speech intelligibility ¹⁶, perception ⁸ and understanding ¹⁷, and findings have been inconclusive. Radonovich et al showed that the N95 mask resulted in a mean (standard deviation, SD) Modified Rhyme Test score of 83 (16.2) % compared to 92 (5.8) % in non-mask controls ¹⁶. Bandaru et al ⁸ found an increased speech reception threshold and decreased speech discrimination score whilst the N95 mask was worn. Coyne et al ¹⁸ found that listener's percentage of accurate speech comprehension was lower in respirator conditions than non-respirator conditions for both a single-word comprehension test and a sentence comprehension test. They found that when listening at 9.15m, the percentage of correct comprehension was 0% for the respirator condition compared to 64.2% for the non-respirator in the single-word comprehension test ¹⁸. Mendel et al ¹⁷, however, showed that the use of a standard surgical mask did not have unfavourable effects upon speech understanding in both the normal-hearing and hearing-impaired listeners. Magee et al ¹⁹ also found no significant effects of masks (N95, surgical, and cloth mask) on intelligibility for both single word and sentences of the Assessment of Intelligibility of Dysarthria Speech ²⁰. Cohn et al ²¹ found that the intelligibility of speech produced with facemasks was determined by speaking style rather than the effects of a mask. They showed that in the clear speaking style, speech in the mask condition was more intelligible than that in the non-mask condition. This raises a question of how listeners perceive the clarity of the speech sounds produced with a facemask. Although the above-mentioned studies examined intelligibility in speech produced by mask users, they did not provide data on how listeners perceived clarity of speech. One study mentioned the impact of mask-wearing upon volume and speech clarity without providing any data ¹². Recommendations to improve communication/speech clarity in mask-wearing conditions have been made without evidence ¹². Given that speech clarity is a factor that determines intelligibility (see Uchanski ²²), it seems necessary to examine how this would be perceived in mask wearing conditions. The concept of speech clarity can be differentiated from speech intelligibility by the definition of clarity posited by Greisinger; "the ability of a space to transfer information from a source to a listener with minimal loss of data" ²³, p.3.

Several studies have attributed the reduced intelligibility of speech produced by mask users to the attenuation effects of masks on high-frequency spectral energies ²⁴⁻²⁶. In an experiment that presented the white noise signals via the mouth of a head and torso simulator, Goldin et al ²⁷ found that the surgical mask and the N95 mask attenuated the sound levels at frequency regions between 2kHz and 7 kHz. Magee et al ¹⁹ showed that different mask types affected different spectral regions with the N95 mask impacting above 3 kHz and surgical and cloth masks above 5 kHz. Nguyen et al ²⁸ showed that spectral levels of connected speech were significantly lower in the 1 - 8kHz frequency range in surgical and KN95 masks. However, there is a limited understanding of the effects of facemasks on specific phonemes that may help account for the reduced intelligibility. Llamas et al ²⁶ showed misperceptions of speech resulting from confusion of stop consonants (e.g. /t/~ /θ/), misperception of place of articulation

of stop consonants and place of articulation of fricatives e.g. /f/~θ/. They also found that speech produced with a surgical mask had lower amplitude of vowel formants and fricative energy above 3.2kHz. However, they only provided visual illustration of their observations and did not provide quantitative data regarding the impact of facemasks on speech spectral characteristics associated with the affected vowels and fricative consonants. Given that the high frequency regions play an essential role in discriminating consonants²⁹, speech recognition³⁰, and speech intelligibility³¹, it seems necessary to investigate how facemasks impact on particular fricative consonants and vowels. Examining the characteristics of fricatives in mask-wearing conditions has important implications to not only normal hearing but also hearing-impaired people who have difficulties recognizing voiceless fricatives³².

The speech spectrum conveys information about both the voice source and the vocal tract³³. There are two major components within the speech spectrum: harmonics and noise³⁴. Harmonics represent vibratory components as whole-number multiples of fundamental frequency (F0)³⁵, whilst noise components can include both glottal noise³⁶ and vocal tract noise produced by articulators in consonants such as fricatives³⁵. The formants represent the spectral energy associated with specific vocal tract configurations (length, shape, and size) of relevant vowels³⁷. Given that any vowel can be characterized spectrographically using frequency-based (formant frequency) and amplitude-based (formant amplitude) measures, measuring spectral amplitude of vowel formants can help understand whether those measures are attenuated by facemasks. Formant amplitude plays an important role in vowel perception^{38,39} may relate to the naturalness and phonetic quality of vowels⁴⁰. Fricatives can be examined using spectral amplitude e.g. root-mean-square (RMS) amplitude and spectral moment measures (centre of gravity, standard deviation, skewness, and kurtosis)^{41,42}. Quantifying spectral amplitude of fricative consonants would also give some insight into understanding more specifically how wearing a facemask affects the characteristics of fricatives. Amplitude and spectral moments of fricatives are important correlates of clearly spoken speech⁴².

Understanding how the speech signals are affected when wearing a mask seems necessary, as mask wearing is a standard recommendation for reducing the risk of respiratory pathogens. Evidence of which linguistic components of speech are affected by masks would be useful in suggesting strategies that help speakers overcome the impact of masks on their speech. Those findings would also help mask manufacturers in selecting suitable materials and designing mask styles that minimize filtering out of certain spectral information relevant to speech understanding in normal hearing and hearing-impaired listeners.

Surgical masks and KN95 masks (China GB2626-2006)⁴³ are commonly used worldwide during the COVID-19 pandemic. The major filtering and fitting characteristics of the KN95 mask as provided by 3M⁴⁴ were as follows: Filter performance $\geq 95\%$; Flow rate = 85 litres/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 leakage⁴⁴.

The aims of the present study were to 1) examine listener's perception of the clarity of speech produced without and with wearing a surgical or KN95 mask; 2) evaluate amplitude and spectral moment measures of voiceless fricative consonants in speech produced with and without wearing a facemask; 3) quantify vowel spectral amplitude in speech with and without a facemask; and 4) investigate the relationships between the acoustic measurements and speech clarity ratings. Our hypotheses were: 1) Wearing a facemask would impact on auditory perception of speech clarity; 2) Wearing a facemask would attenuate amplitude of voiceless fricative consonants and spectral amplitude of vowels of connected speech; 3) Spectral moment measures of voiceless fricative consonants would be affected in mask-wearing conditions; and 4) In the context of mask-wearing, there would be significant relationship between speech clarity ratings and the amplitude measures of fricatives and vowels.

2. Methods

2.1. Ethical approval

This study was approved by the Human Research Ethics Committee of The University of Sydney (project number: 2020/399). Informed consent was obtained from all participants to participate in this study. All methods used in the present study were performed in accordance with the relevant ethical guidelines and regulations. The measurement procedures conformed to the standards set by the latest revision of the Declaration of Helsinki.

The study required recruitment of two cohorts of participants: speakers and listeners. Speakers were recorded wearing and not wearing two types of masks, while listeners were required to listen to voice recordings and rate specific auditory-perceptual features.

2.2. Voice recording

2.2.1. Speakers

Sixteen speaker HCWs took part in this study (12 females, 4 males) with mean age = 43 years (range = 24 - 61), including two otolaryngologists, 13 practicing speech language pathologists, and one registered nurse working in an Ear Nose and Throat clinic. Inclusion criteria were English speakers, non-smokers, with no self-reported voice or hearing problems at the time of the study. Exclusion criteria included any voice or hearing problem at the time of the study.

2.2.2. Procedure

Voice recordings were performed in a quiet room or soundproof booth at the participants' respective clinics as social distancing measures during the COVID-19 pandemic prohibited the use of the same room. Participants used their habitual voice to read the Rainbow Passage⁴⁵ in three conditions with the speaker (1) not wearing a mask, (2) wearing a surgical mask, and (3) wearing a KN95 mask. 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. It has been known that in unfavourable/challenging speaking conditions, speakers may adapt a phonation style that helps improve clear phonation^{46,47}. Therefore, we required participants to maintain similar habitual voice in terms of pitch, loudness, and speaking style 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 cardioid ear-mounted microphone⁴⁸ placed at a constant distance of 6cm, 45° off the mouth axis with analog-to-digital conversion via a professional external sound card (Roland Quadcapture⁴⁹) at 44.1kHz and 16-bit resolution. The signals were processed and saved to a laptop computer using the Audacity sound editing software⁵⁰ in *.wav format. 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.

Given that 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.01.01⁵¹. Only samples with a SNR greater than 30dB were used for auditory-perceptual and acoustic analyses⁵².

2.3. Auditory-perceptual analyses

2.3.1. Listeners

Listeners were recruited via email advertisement sent to an international professional network of speech language pathologists and ENT specialists. Inclusion criteria included 1) working with voice patients as speech-language pathologists, voice specialists, or laryngologists; and 2) Normal hearing at the time of the study. Exclusion criteria: None of the above-mentioned occupations and self-reported hearing impairment. Twenty listeners initially signed in to complete the ratings. Seven raters demonstrated good intra-rater reliability (see below) and only data from these raters were included in the study. The remaining listeners had poor intra-rater reliability and were excluded from further analysis.

2.3.2. Stimuli

The Rainbow Passage⁴⁵ was used for listening tests (*'When the sunlight strikes raindrops in the air... the pot of gold at the end of the rainbow'*). The stimuli represented non-mask (n=16), surgical mask (n=12), and KN95 mask (n=12) conditions. Twelve samples were repeated for intra-rater reliability evaluation. In total, there were 52 samples which were coded and randomized throughout for presentation to the listeners. Given that auditory-perceptual rating of clear speech is affected by vocal intensity^{53,54}, all stimuli were normalized for intensity using the 'Normalize' command in Audacity with the checkbox

'Normalize peak amplitude to -3.0 dB' being checked. After normalization, the output intensity level of stimuli was between 72.0 to 75.0 decibel (dB) sound pressure level (SPL) as measured in Praat and was presented to listeners via a headphone. This level was used as it has been shown that intelligibility of speech produced at a range of vocal effort levels from 55dB to 78dB was constant whilst speech produced at louder and softer intensity levels resulted in decreased word intelligibility ⁵⁵.

2.3.3. Perceptual rating scale

In this study we were interested in the ratings of clarity of speech in surgical mask and KN95 mask conditions. The perception of speech clarity has been examined using different rating schemes ⁵⁶. Magnitude-estimation scaling, for example, is a reliable estimation for speech clarity ⁵⁷. Magnitude estimation can be implemented using both word identification tests and/or the degree of the attribute of interest ⁵⁸. Some authors have also used rating scales to quantify speech clarity. For example. Tasko and Greilick ⁵⁹ used a computer-based slider scale where raters compared clarity of word pairs and made judgment by moving the slider from the mid-point of the scale toward the clearer stimulus ⁵⁹. Reinhart and Souza ⁶⁰ used a 7-point Likert scale with 1 representing 'completely unclear' and 7 'completely clear'. In the present study, we used the VAS with a straight line containing 100 points (1-100) with 1 and 100 representing 'completely clear' and 'completely unclear', respectively, i.e. the higher the score, the less clear the speech sound.

2.3.4. Procedure

Listening tasks were conducted using an online auditory-perceptual rating tool called Bridge2practice, which is a free online education and research platform developed for perceptual learning and practice of allied health professionals and students ⁶¹. Listeners were required to listen to the speech stimuli as many times as they wished using headphones and make a judgment about speech sound clarity by changing the position of the slider on the VAS line described above. All stimuli were randomized and raters were not aware that a number of stimuli were produced with speakers wearing a facemask. Responses were registered in the rating platform and exported to an Excel spreadsheet.

2.3.5. Reliability of perceptual ratings

Intra- and inter-rater reliability were assessed using SPSS 24.0 (SPSS, Inc., Chicago, IL, USA). Intraclass correlation coefficients (ICC) ⁶² were used to determine the level of agreement between the first and second (repeated) ratings (intra-rater reliability) and across listeners (inter-rater reliability). ICC was calculated using a two-way mixed model, consistency type, and single measure analysis [ICC (3,1)]. To assess the level of correlation, ICC < 0.5 indicates poor correlation, 0.5 - 0.75 moderate, 0.75 - 0.9 good, and > 0.9 excellent correlation ⁶³. Intra-rater reliability ranged from ICC = 0.647 to 0.785 for Single Measures and from ICC = 0.785 to 0.880 for Average Measures. Inter-rater reliability amongst the seven raters was moderate based on average measures (ICC = 0.692, p = 0.003).

2.4. Acoustic analyses

2.4.1. Root-mean-square (RMS) amplitude (A_{RMS}) of fricatives

Two fricatives /s/ (alveolar) and /ʃ/ (palato-alveolar) were used for analysis for specific reasons. Firstly, these fricatives have higher amplitude than other voiceless fricatives e.g. /θ/ and /f/ ⁶⁴ which would make it easier to reliably identify and extract them from connected speech compared to other fricatives. Secondly, the duration of these two fricatives is longer than that of the other fricatives ⁶⁴, making it more likely to obtain stable and reliable identification of fricative boundaries and acoustic values. The /s/ and /ʃ/ are characterized by well-defined spectral shapes compared to labio-dental and dental fricatives /f/, /v/, /θ/ and /ð/ which have a relatively flat spectrum without a clear dominant peak ⁴¹. Lastly, /s/ and /ʃ/ differ in their spectral mean, representing different locations of spectral peaks in the spectrum ⁴¹, and allowing investigation of a wider range of high-frequency fricative signal assumed to be affected by facemasks.

The A_{RMS} was used as this has been frequently examined to characterize English voiceless fricatives ^{41,42}. The amplitude of the fricative signal has also been considered an important cue to perceive the place of articulation in voiceless fricatives and hence the accuracy of fricative consonant production ⁶⁵. Firstly, the signals were high-pass filtered at 1 kHz in Audacity ⁵⁰ to remove any potential trace of voicing due to the pre- and postvocalic environment of the fricative ⁶⁶ and to minimize low-frequency energy which could interfere with detection of zero-crossings due to the turbulent source ⁴². Audacity software was used to high-pass filter the sound files with 6-dB roll-off per octave. The fricative /ʃ/ was extracted from the word 'shape' in '*...These take the shape of a long round arch...*' and /s/ was extracted from the word 'say' in the sentence '*...his friends say he is looking for the pot of gold...*'. The boundaries of this consonant were identified visually using acoustic waveform and spectrograms in Praat 6.1.40 ⁶⁷ (Figure 1) and by listening to the sample. Fricative signals were defined as having the following criteria: 1) Characteristic waveform with zero-crossing; and 2) High-frequency noise energy in the narrow-band spectrogram. Koenig et al. ⁶⁸ used 25- millisecond (ms) fricative segments. In this study, the middle 50ms segment was extracted from the centre of these fricatives for acoustic analysis. Onset and offset segments were excluded as for these fricatives, the onset (immediately before voicing onset) and offset have lower amplitude than the middle ⁶⁴ hence extracting the middle segment would increase the probability of capturing the amplitude peaks of the fricative noise signals. The edited /s/ underwent a Fast Fourier transform and was analysed in Praat in the frequency range 1 - 10kHz.

The A_{RMS} over the time interval $t1 \leq t \leq t2$ was defined using the formula ⁶⁷:

$$A_{\text{RMS}} = \sqrt{\left[\frac{1}{(t2-t1)} \int_{t1}^{t2} x^2(t) dt \right]} \quad (1)$$

in which A is the amplitude of the sound. A_{RMS} was converted from Pascal (Pa) unit in Praat to sound pressure level (SPL) in decibels (dB) using the formula:

$$\text{dB SPL} = 20\log_{10} (P/P_0) \quad (2)$$

where P was A_{RMS} value and $P_0 = 20$ micropascals (μPa) which was the reference value.

2.4.2. Spectral moments of fricatives

Apart from quantifying amplitude of the two fricatives, we were interested in clarifying whether facemasks affected other attributes of these consonants. Centre of Gravity (Hz), Standard Deviation (SD, in Hz), Skewness, and Kurtosis have been used extensively in the literature to characterize voiceless fricatives in both conventional speech⁴¹ and clear speech contexts⁴². These measures were obtained from two fricatives, being /s/ and /ʃ/ in Praat.

2.4.3. Amplitude measures of vowels

2.4.3.1. Vowel root-mean-square amplitude

The A_{RMS} was measured as it has been used frequently to investigate spectral amplitude of vowels⁶⁹. The following vowels were edited from the Rainbow Passage: /æ/ in 'arch', /ɪ/ in 'many', and /ʊ/ in 'two', which represent primary cardinal vowels with the highest and most forward tongue position (/æ/), highest and most backward tongue position /ɪ/, and lowest tongue position (/ʊ/) ³⁵. A previous study has shown that the RMS amplitude was different across vowels when put in context (connected speech) with /æ/ presenting the highest RMS amplitude whilst /ɪ/ and /ʊ/ have the lowest RMS amplitude⁶⁹. Using these vowels would help clarify whether different vowel amplitudes were affected similarly or differently by the masks. In addition, these three vowels are also produced with different levels of lip rounding and protrusion³⁵ which might also be affected by the KN95 mask because of its tighter levels of fitting than a standard surgical mask. The vowels were extracted in Praat by listening and identifying waveform and spectrogram characteristics associated with the required vowel. For each vowel, the middle 50ms was extracted and analysed for A_{RMS} using the same protocols as mentioned above for fricatives.

2.4.3.2. Amplitude of the first three formants with formant frequency

Although amplitude of vowels in context can be quantified using both RMS and formant amplitude⁶⁹, the RMS gives an overall vowel amplitude rather than amplitude of specific formants. Therefore, this study measured amplitude of the first three formants (hence A1, A2, and A3) from the above-mentioned vowels using a MATLAB code called VoiceSauce^{70,71} employing the Snack Sound Toolkit⁷². Our previous observations indicated that facemasks impacted the high-frequency ranges above 1kHz²⁸, therefore, the first formant was not deemed to be affected. However, amplitude of this formant (A1) was included so that between-formant cross-reference could be made if formant amplitude is normalized to eliminate between-speaker variabilities. The signals were first down-sampled to 16kHz and all measurements were

implemented automatically at every 1 milliseconds (ms) for voiced segments with a window length of 25ms⁴⁰. The highest formant amplitude within this window length was obtained. Settings were as follows: min F0 = 75Hz and max F0 = 400Hz; pre-emphasis = 0.96; and Linear Predictive Coding (LPC) order = 12. Data points with zero values were deleted.

Frequencies of the first three formants (F1, F2, and F3) were measured in Praat to present the range of specific formant frequencies of the formants used in amplitude measurements. Formant frequencies were measured using Praat from the middle 30ms of the vowels where the vowel production was the most stable⁷³. Formant settings followed default in Praat including: Maximum formant = 5500Hz, number of formants = 4, window length = 25ms, dynamic range = 30dB, and dot size = 1.0mm⁷³.

2.4.4. Reliability of acoustic analyses

A co-author repeated the file editing and measurement process on both /s/ and /ʃ/. Table 1 shows results of intraclass correlation coefficients calculated for the two fricative consonants in the no-mask condition.

Table 1. Intraclass correlation coefficient (ICC) for inter-rater reliability of spectral measures for the two fricatives in no-mask (SM, Single Measures; AM, Average Measures). All p values were < 0.001

Spectral measures	Measures	/s/	/ʃ/
		ICC	ICC
Root-mean-square	SM	.997	.999
	AM	.998	1.000
Centre of gravity	SM	.998	.996
	AM	.999	.998
Standard Deviation	SM	.987	.966
	AM	.993	.983
Skewness	SM	.968	.991
	AM	.984	.995
Kurtosis	SM	.991	.998
	AM	.996	.999

A co-author also repeated measurements of formants F1 and F2 of three vowels mentioned above for all three conditions in 50% of the participants. In total, 72 repeats were implemented (n = 8 participants x 3 conditions x 3 vowels = 72). ICC was calculated for inter-rater reliability and the results are presented in Table 2, which showed good to excellent agreement between the two raters for these acoustic measures.

Table 2. Intraclass correlation coefficient (ICC) for inter-rater reliability of formant measurement (SM, Single Measures; AM, Average Measures)

Formants	Measures	ICC	p
F1	SM	0.832	< 0.001
	AM	0.908	< 0.001
F2	SM	0.899	< 0.001
	AM	0.947	< 0.001

2.5. Statistical analyses

Data were managed in Microsoft Excel 365⁷⁴ and analysed using IBM SPSS Statistics v.24.0⁷⁵ and Prism v8.1.2⁷⁶. One-way repeated-measures analysis of variance (ANOVA) was used to examine the effects of masks on acoustic measures. Significant main effects were further analysed using pairwise comparisons with Bonferroni-adjusted p values. Prior to analyses, normal distribution of the data was examined using Kolmogorov-Smirnov tests⁷⁷. 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, respectively⁷⁸. If normality assumption was not met, the non-parametric Friedman test was used to compare data across non-mask, KN95, and surgical mask conditions. Pearson's correlation coefficient (r) was used to calculate the correlation between acoustic data and perceptual rating of speech clarity in which r = 0.1, 0.3, and 0.5 indicated small, medium, and large effects, respectively⁷⁹. Multivariate linear regression was used to examine acoustic predictors of speech clarity. In all statistical calculations, a significance level of 0.05 was used (two-tailed). Where there were multiple calculations, Bonferroni adjustment was applied to p values.

3. Results

3.1. Speech clarity ratings

For each participant, rating of speech clarity was scored by raters on a 100-point scale with 1 = completely clear and 100 = completely unclear. Rating scores were then averaged from all raters and used to calculate the population's parameters and analysed statistically. Mean rating scores of speech clarity of the Rainbow Passage in all three conditions are presented in Figure 2, with the higher the scores representing less clear speech. Repeated-measures ANOVA showed significant effects of masks [$F(2, 22) = 11.702$, $p = 0.000$, partial $\eta^2 = 0.539$]. The post-hoc test showed that the surgical and KN95 masks deteriorated speech clarity scores by 6.1 (Bonferroni-adjusted $p = 0.127$) and 12.0 (Bonferroni-adjusted $p = 0.001$), respectively.

3.2. Root-mean-square (RMS) amplitude (A_{RMS}) of fricatives

This measure was obtained in Praat as Pascal units and converted to decibels using the formula (2). Figure 3 shows data of this measure obtained from the two fricatives. Two-way repeated-measures ANOVA [within (no-mask, surgical, KN95) x between (/s/, /ʃ/)] showed significant effects of masks on A_{RMS} [$F(2, 42) = 6.540, p = 0.003, \text{partial } \eta^2 = 0.237$]. Compared with the no-mask condition, this measure decreased by 3.4 dB in the surgical mask (Bonferroni-adjusted $p = 0.001$) and by 5.9 dB in the KN95 mask (Bonferroni-adjusted $p = 0.0002$) groups. This measure was not significantly different between surgical masks and KN95 masks ($p = 0.062$).

3.3. Spectral moments of fricatives

Table 3 shows descriptive statistics of the four spectral moments of the two fricatives in all conditions. For /s/, Centre of Gravity appeared to be higher in no-mask than in surgical and KN95 mask conditions, but this was not statistically significant ($p = 0.103, \text{partial } \eta^2 = 0.187$). No significant effects were found for Standard Deviation ($p = 0.725, \text{partial } \eta^2 = 0.019$); Skewness ($p = 0.204, \text{partial } \eta^2 = 0.134$); and Kurtosis ($p = 0.201, \text{partial } \eta^2 = 0.136$).

For /ʃ/, significant main effects were observed for Standard Deviation: [$F(2, 22) = 7.615, p = 0.003, \text{partial } \eta^2 = 0.409$]. The Standard Deviation in KN95 was 379.3Hz higher than that in no-mask ($p = 0.022$) and 342.9Hz higher than that in surgical mask conditions ($p = 0.039$). Standard Deviation was not significantly different between no-mask and the wearing of a surgical mask ($p = 1.000$). There was also significant main effects on Skewness [$F(2, 22) = 4.968, p = 0.017, \text{partial } \eta^2 = 0.311$]: This measure was lower in KN95 than that in the surgical mask group ($p = 0.027$) but was not significantly different from that in the no-mask condition ($p = 0.173$). Skewness was almost similar between no-mask and surgical mask conditions ($p = 1.000$). There was no significant effect of masks on Centre of Gravity ($p = 0.051, \text{partial } \eta^2 = 0.237$) and Kurtosis ($p = 0.056, \text{partial } \eta^2 = 0.231$).

Table 3. Mean (SD) of the spectral moments of the two fricatives in all conditions

Fricative and conditions		Centre of Gravity (Hz)		Standard Deviation (Hz)		Skewness		Kurtosis	
		Mean (SD)	Min-Max	Mean (SD)	Min-Max	Mean (SD)	Min-Max	Mean (SD)	Min-Max
/s/	No-mask	8693.5 (1238.8)	6141.1 - 10231.4	1748.1 (287.4)	1341.9 - 2330.5	0.5 (0.5)	-0.7 - 1.1	3.4 (2.1)	0.7 - 7.4
	Surgical	8272.2 (1395.4)	4834.2 - 9991.5	1831.6 (518.0)	1125.0 - 3083.9	0.5 (0.9)	-0.3 - 2.9	5.6 (6.0)	1.2 - 23.5
	KN95	8362.2 (1179.6)	6186.8 - 10283.8	1744.2 (433.9)	1022.4 - 2699.9	0.2 (0.7)	-0.8 - 1.2	6.9 (5.4)	1.7 - 22.2
/ʃ/	No-mask	4509.2 (717.1)	3157.6 - 5806.2	1633.1 (286.5)	982.9 - 1951.5	2.6 (1.1)	1.6 - 5.8	14.9 (20.4)	3.4 - 78.5
	Surgical	4430.6 (654.8)	3219.8 - 5336.3	1669.5 (381.6)	1077.4 - 2361.7	2.8 (1.1)	1.3 - 5.4	17.4 (17.3)	2.8 - 67.3
	KN95	4868.2 (1119.0)	3241.8 - 6795.8	2012.4 (436.1)	1352.9 - 2685.0	2.0 (1.1)	1.0 - 4.4	9.5 (10.8)	2.3 - 37.4

3.4. Amplitude measures of vowels

3.4.1. A_{RMS}

Figure 4 shows A_{RMS} of three vowels extracted from the Rainbow Passage in all experimental conditions. No significant main effect of masks was found on this measure for /ʃ/ ($p = 0.289$, partial $\eta^2 = 0.104$), / ʒ / ($p = 0.565$, partial $\eta^2 = 0.051$), and /s/ ($p = 0.210$, partial $\eta^2 = 0.132$).

3.4.2. A1, A2, and A3 with formant frequencies

Figure 5 presents the data for A1, A2, and A3 for the three extracted vowels. As there was great between-speaker variability in these measures across conditions, all formant amplitude data were normalized so that they fell within 0 and 1. Although A1 was not expected to change as the frequency of the first formant (F1) was lower than the frequency ranged affected by these masks (i.e. above 1kHz), A1 was included in the analyses and plots to allow cross-comparisons between formants.

Given the impact of vocal intensity on formant amplitude⁸⁰, vocal intensity data were collected and have been reported in our previous publication²⁸ in which mean (SD) of vocal intensity were 59.7 (5.0) dB, 61.3 (5.2) dB, and 60.9 (5.1) dB for no-mask, surgical, and KN95 mask, respectively ($p = 0.290$).

Results showed that, in /ʃ/, no significant effects were observed for A1 ($p = 0.461$, partial $\eta^2 = 0.068$), A2 ($p = 0.565$, partial $\eta^2 = 0.051$), and A3 ($p = 0.286$, partial $\eta^2 = 0.108$). For / ʒ /, no significant effects were found for A1 ($p = 0.630$, partial $\eta^2 = 0.041$), A2 ($p = 0.994$, partial $\eta^2 = 0.001$), and A3 ($p = 0.495$, partial $\eta^2 = 0.062$). For /s/, there were also no significant effects for A1 ($p = 0.490$, partial $\eta^2 = 0.063$), A2 ($p = 0.383$, partial $\eta^2 = 0.077$), and A3 ($p = 0.667$, partial $\eta^2 = 0.036$). Owing to non-significant effects, no further

post-hoc analyses were implemented to clarify the interaction between mask effects and vowel amplitude levels.

Frequencies of the first three formants were also measured and displayed in Figure 6 to illustrate the specific formant ranges associated with the amplitude measurements. Although formant frequency analyses were not the primary outcome measures in this study, they were also compared across the conditions. Results showed no significant main effects for F1 ($p = 0.647$), F2 ($p = 0.271$), and F3 ($p = 0.594$) of /æ/. No significant main effects were found for F1 ($p = 0.215$), F2 ($p = 0.174$), and F3 ($p = 0.582$) of /ɪ/.

For /ʊ/, no significant main effects were observed for F1 ($p = 0.209$). Although significant main effects were observed for F2 [$F(2,22) = 4.920$, $p = 0.017$, partial $\eta^2 = 0.309$], post-hoc tests only showed a trend that F2 of this vowel in KN95 was lower than that in surgical mask ($p = 0.054$) and not significantly different from that in the no-mask condition ($p = 0.206$); and no significant differences in F2 were present between non-mask and surgical mask conditions ($p = 0.773$). For F3, significant effects of masks were present [$F(2,22) = 5.993$, $p = 0.008$, partial $\eta^2 = 0.353$], F3 in KN95 was significantly lower than that in no-mask ($p = 0.025$) and surgical mask conditions ($p = 0.049$) whilst no statistically significant difference was observed between wearing no-mask and wearing a surgical mask ($p = 1.00$).

3.5. Relationship between speech clarity ratings and acoustic measures in the context of mask-wearing

Significant correlation was observed between clarity ratings and A_{RMS} of /s/ ($r = -0.481$, $p = 0.001$) and /ʃ/ ($r = -0.540$, $p = 0.0002$). There was no significant correlation between clarity ratings and A_{RMS} of /æ/ ($r = -0.044$, $p = 0.779$), /ɪ/ ($r = 0.095$, $p = 0.545$), and /ʊ/ ($r = 0.048$, $p = 0.761$). There was also no significant correlation between clarity ratings and all formant amplitudes measures A1, A2, and A3 for all three vowels /æ/, /ɪ/, and /ʊ/ ($p > 0.05$).

Three prediction models were calculated to estimate the prediction of speech clarity using acoustic measures.

3.5.1. Amplitude regression model

The A_{RMS} values of the two fricatives (/s/ and /ʃ/) and three vowels (/æ/, /ɪ/, and /ʊ/) were input as predictors into a multivariate regression model with clarity ratings being the dependent variable. This model accounted for 50.5% of the total variance [$F(5, 37) = 7.543$, $p = 0.00006$]. A_{RMS} of /ʃ/ was the single significant predictor for speech clarity rating in this model (Unstandardized Coefficients B = -1.034, $t = -3.336$, $p = 0.002$). A_{RMS} of /s/ was not a significant predictor ($t = -1.140$, $p = 0.261$). The A_{RMS} values of the three vowels were also not significant predictors of clarity rating ($p > 0.05$).

3.5.2. Fricative regression models

A_{RMS} value and the four spectral moments of /s/ were input into a multivariate regression model. This model accounted for 30.3% of the total variance [$F(5, 37) = 3.222, p = .016$] and A_{RMS} of /s/ was the significant predictor ($B = -.606, t = -2.482, p = .018$). Spectral moments of /s/ were not significant predictors of speech clarity ($p > 0.05$).

In the second fricative multivariate model to predict speech clarity using A_{RMS} and the four spectral moments of /ʃ/, the model accounted for 40.2% of the total variance [$F(5, 37) = 4.984, p = 0.001$]. In this model, A_{RMS} of /ʃ/ ($B = -0.707, t = -3.003, p = 0.005$) and Standard Deviation of /ʃ/ ($B = 0.008, t = 2.069, p = 0.046$) were significant predictors of speech clarity rating.

4. Discussion

Speaking with a facemask creates an adverse speaking condition for mask users and an adverse reception or perception condition for listeners, especially the hearing-impaired. Knowing how acoustic-phonetic characteristics of speech are presented in masked speech is necessary so that mask users can use strategies to make their speech clearer and more audible. These findings would also be of interest for mask manufacturers in their process of selecting materials and designing mask styles. In this study we hypothesized that mask wearing would impact listeners' perception of speech clarity and attenuate amplitude measures of voiceless fricatives and vowels as measured from connected speech. We also expected that spectral moments of the voiceless fricatives would be affected by the masks. Results confirmed that speech clarity was reduced in both surgical masks and KN95 masks with the latter having greater impact. The masks lowered A_{RMS} of both voiceless fricatives but not of the vowels. Amongst the measures of the fricatives, A_{RMS} was the main significant predictor of speech clarity.

Of the two mask types used in this study, the KN95 mask has high filtering levels⁴⁴, which means it outperforms a standard surgical mask in terms of protection effects. Those characteristics may also stipulate how speech sounds are transmitted, absorbed and reflected. Different materials have varying sound absorption effects due to varying transmission loss (a property of a material that relates to its sound attenuation characteristics)²⁶. It is likely that the effects observed in this study are partly due to the transmission loss which is probably greater in KN95 masks compared to surgical mask materials. In addition, the results on formant frequency (i.e. the decrease in F2 and F3 of /ʃ/ in KN95) appear to imply the effects of higher fitting levels and closer facial proximity of this mask on vowels with lip rounding or protruding. Fant³⁷ demonstrated that in the frequency range of the first four formants, lip-rounding has effects on all resonance frequencies.

Speech clarity in mask-wearing

Speech in the KN95 mask condition was significantly less clear than that in surgical mask and no-mask conditions, but it was not possible to infer that the intelligibility is reduced as the speech tasks and rating scale were not designed to test that dimension. Speech perception can be influenced by the distortions of the acoustic signals⁹, ambient noise levels¹⁰ and the use of other psycho-linguistic features in speech

production⁸¹. It is also established that in adverse communication environments, the speaker may use a particular phonation style as compensation or adaptation^{10,82}. We required participants to maintain a similar conversational speaking style across the mask conditions. Although changes in auditory feedback might factor in the actual phonation, we would expect minimal compensation given that 13 of the participants were practicing speech language pathologists who were trained to control phonation well as instructed to do so. We did not intentionally examine how different speaking styles would change whilst wearing a facemask. Instead, we attempted to minimise the impact of clear speech mechanisms on both perceptual and acoustic data to isolate the effects of the masks on speech. Previous research has found that the performance of a clear speech style was different between non-mask and mask conditions. In examining three different speaking styles i.e. casual, clear, and emotional speech with and without a fabric mask, Cohn et al²¹ found that speaking style outperformed the impact of facemasks on intelligibility. They showed that in the clear speaking style, the speech in mask condition was more intelligible than that in the non-mask condition. Note that Cohn et al²¹ conducted their experiments using a fabric (cloth) mask. It is likely that adaptation mechanisms would be different across different mask types i.e. compensation might be more successful behind a fabric mask than masks possessing tighter fitting and higher filtering characteristics such as surgical and KN95 masks. Although we might assume that a mechanism of clear speech was present unintentionally in our participants (less likely with trained participants and more likely with untrained participants) which might to some extent affect rating scores of speech clarity, the findings indicated that speech clarity was more likely to be impacted by both KN95 and surgical masks.

We asked raters to judge clarity of the speech sounds they heard based on a standard reading passage (Rainbow Passage); therefore, perceptual scores of speech clarity did not imply deteriorated speech intelligibility. Moreover, intelligibility and clarity have been regarded as separate concepts by some authors⁶⁰. Clear speech has an advantage over conversational speech in terms of intelligibility²². However, findings in the present study did not allow inference of how less clear speech would impact on intelligibility. Previous research by Magee et al¹⁹ has demonstrated that speech intelligibility at both word and sentence level was not influenced by mask wearing of N95 (similar filtering as KN95), surgical, and cloth masks when listeners were asked to transcribe audio recordings. Mendel et al¹⁷ have also shown a minimal impact of mask-wearing upon speech understanding. Mendel also minimised the implications of speech understanding when wearing a mask upon hearing-impaired listeners.

Amplitude of fricatives and vowels

Previous research on the impact of facemasks on the speech signal has found attenuated energy at high frequency ranges²⁴⁻²⁶ but they have not provided data on the type of phonemes/tasks that might be affected. In the present study, we isolated two types of voiceless fricatives (/s/ and /ʃ/) and three primary cardinal vowel types (/i/, /e/, and /a/) from connected speech to provide some preliminary data on how facemasks might interact with the linguistic tasks to produce different effects. These two fricatives were alveolar and palato-alveolar in place³⁵ with different central frequencies⁴¹, allowing investigation of a

broader range of frequencies given that the impact of masks may be different across different high-frequency ranges¹⁹. The three vowels were chosen as they have different formant structures³⁵ and represent cardinal vowels in speech. We found that the RMS amplitude of both fricatives was lower in mask conditions. However, the RMS amplitude and formant amplitude (A1, A2, and A3) of all the vowels were not significantly different across conditions. These findings agreed with our previous observations that in the 1 - 8 kHz region, vowel energy was not affected by these masks²⁸. The finding also explained our previous assumptions of possible phonemes affected when observing decreased spectral levels at 1-8kHz regions of connected speech²⁸. On the other hand, the findings implied the selective filtering of these masks on different components of the speech spectra. In other words, not all types of the speech signals are filtered out in the same way by facemasks. The selective filtering might be partly related to the amplitude levels of different phonemic types. In connected speech, vowels often have stronger power levels than voiceless fricatives e.g. /θ/ and /f/⁶⁵. That means weaker amplitudes would be more likely to be masked out. Nevertheless, some fricatives e.g. /s/ used in the present study can have equal or higher intensity than the vowel⁶⁵. As such, determining how the amplitude of these fricatives was lowered in mask conditions would need further investigation. The decreased voiceless fricative amplitude would add further levels of difficulty of understanding connected speech for people with hearing loss who already have difficulties perceiving voiceless fricatives in non-masked speech³².

Spectral moments of fricatives

No significant differences in the central frequency were found across conditions for both fricatives, although Table 3 shows that the Centre of Gravity of /s/ appeared to be lower for surgical mask and KN95 masks than for the no-mask condition. This study found greater Standard Deviation and lower Skewness of /ʃ/ in KN95 masks, and this demonstrated a greater variability in frequency distribution in this mask condition compared to the surgical and no-mask conditions. It is interesting to note that the Standard Deviation of /ʃ/ was a significant predictor of speech clarity in this study. As such, apart from reducing spectral energy of the fricatives, changing frequency distribution of /ʃ/ was another effect of KN95 mask on speech which also stipulated the perception of speech clarity.

It is likely that these two fricatives may not be the only segmental units to be modified by facemasks. Other fricatives and consonants, and vowel formant amplitude and harmonics may also be affected in the high frequency regions, however further experiments would be needed to confirm this. The non-significance in spectral moments across conditions of /s/ also meant that adaption towards clear speech mechanisms in mask conditions was probably not the case given previous reports that Center of Gravity of fricative /s/ were significantly higher in clear speech than in conversational speech⁴².

Centre of Gravity values were higher than that of 6133 Hz for English fricative /s/ as reported by Jongman et al⁴¹, which is likely due to differences in study populations and analysis methods. They measured spectral moments from /s/ occurring with different vowels whilst we only used /s/ in 'say'; they used a slightly shorter analysis window (40ms) than ours (50ms); and they sampled at different locations

of the fricative (onset, middle, end and centered over fricative offset) whilst we only measured the middle 50ms. The small sample size ($n = 16$) meant that our data should be interpreted with caution.

Determinants of speech clarity in mask-wearing conditions

Although various factors can stipulate speech clarity²², it was important to find out which speech signal features the raters relied upon to judge the level of clarity in the context of mask-wearing. This is important to determine the phonation/speaking strategies that speakers would need to implement to overcome the effects of the masks in making their speech more audible. Using acoustic and perceptual data, we calculated three linear regression models to estimate how speech clarity could be explained using the acoustic measures. In the amplitude model, all RMS values of the two fricatives and three vowels were used. This model demonstrated that only A_{RMS} of /ʃ/ was a significant predictor whilst none of the vowel amplitude measures successfully explained speech clarity. The two fricative multivariate models both confirmed that A_{RMS} of these fricatives was an important cue that stipulated the speech clarity rating in the present study. The Standard Deviation of /ʃ/ was another significant predictor of speech clarity, meaning that depending on fricative types, frequency distribution might also affect speech perception in mask-wearing conditions. Previous research has shown that both the amplitude and spectral properties of the fricative signal are amongst important factors that determine the perception of the place of articulation for fricative consonants, i.e. the accuracy of consonant production⁶⁵. The design of the present study did not allow us to infer much about how these amplitude measures affected listeners' perception of speech sound clarity in masked speech. It is necessary to acknowledge that the term "speech clarity" can have different meanings and implications and does not only imply a link with consonant production. However, the findings would suggest that improving fricative production might be a reasonable strategy for facemask users. In addition, in the manufacturing process of masks, it would be important to use materials that have minimal filtering effects on fricative signals.

The reduced quality of speech as observed in the present study was probably not significantly accounted for by the vowel amplitude. The amplitude was only measured for the first three formants and did not capture all of the frequency regions affected by these masks. Future studies should examine vowel amplitude in regions above the third formant given that Magee et al¹⁹ found that the N95 mask attenuated energy above 3 kHz and surgical and cloth masks affected the regions above 5 kHz. The amplitude model included amplitude of both fricatives; the central frequency of which was well above 4000Hz for /ʃ/ and above 8000Hz for /s/ (Table 3) and these performed well in capturing the frequency regions impacted by both the surgical mask and KN95 mask. Although we did not find formant amplitudes A1, A2, and A3 to be significant predictors of speech clarity, the role of these in vowel and speech perception is well recognized^{38,39}. The design of this study did not allow us to confirm that these amplitude measures were not important determinants of speech perception/intelligibility. Formant frequency might also play a role given the lower F2 and F3 in the KN95 condition, which needs further studies to clarify.

5. Conclusion

The following points can be made based on the findings of this study:

- 1) Wearing either a surgical mask or KN95 mask has a negative impact on auditory perception of speech clarity. Speech sound was perceived as significantly less clear in these mask conditions. Owing to the tasks used in the perceptual test, it was not possible to conclude that speech intelligibility was affected by these masks.
- 2) These facemasks attenuated spectral energy of both voiceless fricative consonants /s/ and /ʃ/. The RMS amplitude and the amplitude of the first three formants of the vowels /æ/, /i/, and /u/ were not modified by these masks. These implied a selective filtering of spectral energy by these masks. The reduced fricative information might impact not only the normal hearing but also the hearing-impaired listeners. The non-change in formant amplitude of these vowels did not imply that vowel spectral amplitude was not modified by these masks, given that only the first three formants were measured.
- 3) Amongst the spectral moments of the fricatives, only the Standard Deviation of /ʃ/ was affected by these masks. Central frequency of these fricatives was not affected. This demonstrated that these facemasks did not appear to modify the frequency structure of these two fricatives.
- 4) Speech clarity in the context of mask-wearing was significantly predicted by the reduced spectral amplitude of the fricatives rather than their frequency distribution and amplitude of the vowels.

Declarations

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

Conflicts of interests: Nil declared.

Author contribution statements

Duy Duong Nguyen was involved in research question identification, study planning, protocol preparation, data analysis, data interpretation, manuscript writing and editing, and graphic works. Antonia Chacon was involved in data collection, data analysis, and manuscript editing. Christopher Payten, Rebecca Black, and Meet Sheth were involved in problem identification and manuscript editing. Patricia McCabe was involved in study planning, result interpretation, and manuscript reviewing and editing. Daniel Novakovic interpreted findings and edited the manuscript. Catherine Madill 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.

Additional Information

Competing interests

The authors have no competing interests to declare in this study.

References

1. Long, Y. *et al.* Effectiveness of N95 respirators versus surgical masks against influenza: A systematic review and meta-analysis. *J Evid Based Med* **13**, 93-101 (2020).
2. Cook, T. M. Personal protective equipment during the coronavirus disease (COVID) 2019 pandemic - a narrative review. *Anaesthesia* **75**, 920-927 (2020).
3. Davies, A. *et al.* Testing the efficacy of homemade masks: would they protect in an influenza pandemic? *Disaster Med Public Health Prep* **7**, 413-418 (2013).
4. Leung, N. H. L. *et al.* Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med* **26**, 676-680 (2020).
5. Johnson, A. T. Respirator masks protect health but impact performance: a review. *J Biol Eng* **10**, 4 (2016).
6. Ribeiro, V. V. *et al.* Effect of Wearing a Face Mask on Vocal Self-Perception during a Pandemic. *J Voice* (2020).
7. Loukina, A., Evanini, K., Mulholland, M., Blood, I. & Zechner, K. Do Face Masks Introduce Bias in Speech Technologies? The Case of Automated Scoring of Speaking Proficiency. *INTERSPEECH*, 1942-1946 (2020).
8. Bandaru, S. V. *et al.* The effects of N95 mask and face shield on speech perception among healthcare workers in the coronavirus disease 2019 pandemic scenario. *J Laryngol Otol*, 1-4 (2020).
9. Evitts, P. M. *et al.* The Impact of Dysphonic Voices on Healthy Listeners: Listener Reaction Times, Speech Intelligibility, and Listener Comprehension. *Am J Speech Lang Pathol* **25**, 561-575 (2016).
10. Ishikawa, K. *et al.* The Effect of Background Noise on Intelligibility of Dysphonic Speech. *J Speech Lang Hear Res* **60**, 1919-1929 (2017).
11. 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).
12. Knollman-Porter, K. & Burshnic, V. L. Optimizing Effective Communication While Wearing a Mask During the COVID-19 Pandemic. *J Gerontol Nurs* **46**, 7-11 (2020).
13. Coniam, D. The Impact of Wearing a Face Mask in a High-Stakes Oral Examination: An Exploratory Post-SARS Study in Hong Kong. *Language Assessment Quarterly* **2**, 235-261 (2005).
14. Thomas, F. *et al.* Does wearing a surgical facemask or N95-respirator impair radio communication? *Air Med J* **30**, 97-102 (2011).
15. Hampton, T. *et al.* The negative impact of wearing personal protective equipment on communication during coronavirus disease 2019. *J Laryngol Otol* **134**, 577-581 (2020).
16. Radonovich, L. J., Jr., Yanke, R., Cheng, J. & Bender, B. Diminished speech intelligibility associated with certain types of respirators worn by healthcare workers. *J Occup Environ Hyg* **7**, 63-70 (2010).

17. 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).
18. Coyne, K. M., Johnson, A. T., Yeni-Komshian, G. H. & Dooly, C. R. Respirator performance ratings for speech intelligibility. *Am Ind Hyg Assoc J* **59**, 257-260 (1998).
19. Magee, M. *et al.* Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. *J Acoust Soc Am* **148**, 3562 (2020).
20. Yorkston, K. M., Beukelman, D. R. & Traynor, C. *Assessment of intelligibility of dysarthric speech*. (Proed Austin, TX, 1984).
21. Cohn, M., Pycha, A. & Zellou, G. Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. *Cognition* **210**, 104570 (2021).
22. Uchanski, R. M. in *The handbook of speech perception* (eds David B. Pisoni & Robert Ellis Remez) 207-235 (Blackwell Pub., 2005).
23. Griesinger, D. H. What is "clarity", and how it can be measured? *Proceedings of Meetings on Acoustics* **19**, 015003 (2013).
24. Corey, R. M., Jones, U. & Singer, A. C. Acoustic effects of medical, cloth, and transparent face masks on speech signals. *J Acoust Soc Am* **148**, 2371 (2020).
25. Palmiero, A. J., Symons, D., Morgan, J. W. & Shaffer, R. E. Speech intelligibility assessment of protective facemasks and air-purifying respirators. *J Occup Environ Hyg* **13**, 960-968 (2016).
26. Llamas, C., Harrison, P., Donnelly, D. & Watt, D. Effects of different types of face coverings on speech acoustics and intelligibility. *York Papers Ling.* **9**, 80-104 (2009).
27. Goldin, A., Weinstein, B. & Shiman, N. How do medical masks degrade speech perception? . *Hearing Review* **27**, 8-9 (2020).
28. Nguyen, D. D. *et al.* Acoustic voice characteristics with and without wearing a facemask. *Sci Rep* **11**, 5651 (2021).
29. Vitela, A. D., Monson, B. B. & Lotto, A. J. Phoneme categorization relying solely on high-frequency energy. *J Acoust Soc Am* **137**, EL65-70 (2015).
30. 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).
31. Hazan, V. & Markham, D. Acoustic-phonetic correlates of talker intelligibility for adults and children. *J Acoust Soc Am* **116**, 3108-3118 (2004).
32. Zeng, F. G. & Turner, C. W. Recognition of voiceless fricatives by normal and hearing-impaired subjects. *J Speech Hear Res* **33**, 440-449 (1990).
33. Baken, R. J. & Orlikoff, R. F. *Clinical measurement of speech and voice*. 2nd edn, (Singular Thomson Learning, 2000).
34. 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).

35. Ladefoged, P. *A course in phonetics*. 6th edn, (Wadsworth, Cengage, 2011).
36. Sousa, R., Ferreira, A. & Alku, P. The harmonic and noise information of the glottal pulses in speech. *Biomed Signal Process Control* **10**, 137–143 (2014).
37. Fant, G. *Acoustic theory of speech production : with calculations based on X-ray studies of Russian articulations*. 2nd edn, (Mouton, 1970).
38. Kieffe, M., Enright, T. & Marshall, L. The role of formant amplitude in the perception of /i/ and /u/. *J Acoust Soc Am* **127**, 2611-2621 (2010).
39. Ito, M., Ohara, K., Ito, A. & Yano, M. in *INTERSPEECH* 2490-2493 (Chiba, Japan, 2010).
40. Jacewicz, E. Listener sensitivity to variations in the relative amplitude of vowel formants. *Acoust Res Lett Online* **6**, 118-124 (2005).
41. Jongman, A., Wayland, R. & Wong, S. Acoustic characteristics of English fricatives. *J Acoust Soc Am* **108**, 1252-1263 (2000).
42. Maniwa, K., Jongman, A. & Wade, T. Acoustic characteristics of clearly spoken English fricatives. *J Acoust Soc Am* **125**, 3962-3973 (2009).
43. 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).
44. M. *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).
45. Fairbanks, G. *Voice and articulation drillbook*. 2nd edn, (Harper & Row, 1960).
46. 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).
47. Hazan, V. *et al*. Clear speech adaptations in spontaneous speech produced by young and older adults. *J Acoust Soc Am* **144**, 1331 (2018).
48. AKG Acoustics. *C520*, <<https://www.akg.com/Microphones/Headset%20Microphones/C520.html>> (2018).
49. Roland Corp. *Quad-capture - USB 2.0 Audio Interface*, <<https://www.roland.com/au/products/quad-capture/>> (2019).
50. Audacity Team. *Audacity(R): Free Audio Editor and Recorder [Computer application]*, <<https://www.audacityteam.org/>> (2019).
51. Maryn, Y. *Recording quality: Speech-to-noise ratio and Voice-to-noise ratio*, <<https://www.phonanium.com/product/recording-quality/>> (2020).
52. Deliyski, D. D., Shaw, H. S. & Evans, M. K. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice* **19**, 15-28 (2005).
53. Neel, A. T. Effects of loud and amplified speech on sentence and word intelligibility in Parkinson disease. *J Speech Lang Hear Res* **52**, 1021-1033 (2009).

54. Tjaden, K., Sussman, J. E. & Wilding, G. E. Impact of clear, loud, and slow speech on scaled intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *J Speech Lang Hear Res* **57**, 779-792 (2014).
55. Pickett, J. M. Effects of Vocal Force on the Intelligibility of Speech Sounds. *J Acoust Soc Am* **28**, 902-905 (1956).
56. Whitfield, J. A. & Goberman, A. M. Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech. *Int J Speech Lang Pathol* **19**, 184-194 (2017).
57. Fucci, D., Ellis, L. & Petrosino, L. Speech clarity/intelligibility: test-retest reliability of magnitude-estimation scaling. *Percept Mot Skills* **70**, 232-234 (1990).
58. Schiavetti, N. in *Intelligibility in speech disorders : theory, measurement, and management* (ed Raymond D. Kent) 11-34 (John Benjamins, 1992).
59. Tasko, S. M. & Greilick, K. Acoustic and articulatory features of diphthong production: a speech clarity study. *J Speech Lang Hear Res* **53**, 84-99 (2010).
60. Reinhart, P. N. & Souza, P. E. Intelligibility and Clarity of Reverberant Speech: Effects of Wide Dynamic Range Compression Release Time and Working Memory. *J Speech Lang Hear Res* **59**, 1543-1554 (2016).
61. Madill, C., So, T. & Corcoran, S. *Bridge2practice: Translating theory into practice*, <<https://bridge2practice.com/>> (2019).
62. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* **86**, 420-428 (1979).
63. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **15**, 155-163 (2016).
64. Behrens, S. J. & Blumstein, S. E. Acoustic characteristics of English voiceless fricatives: a descriptive analysis. *J Phon* **16**, 295-298 (1988).
65. Behrens, S. & Blumstein, S. E. On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *J Acoust Soc Am* **84**, 861-867 (1988).
66. Martel-Sauvageau, V., Breton, M., Chabot, A. & Langlois, M. The Impact of Clear Speech on the Perceptual and Acoustic Properties of Fricative-Vowel Sequences in Speakers With Dysarthria. *Am J Speech Lang Pathol*, 1-19 (2021).
67. Boersma, P. & Weenink, D. *Praat: doing phonetics by computer*, <<http://www.fon.hum.uva.nl/praat/>> (2018).
68. Koenig, L. L., Shadle, C. H., Preston, J. L. & Mooshammer, C. R. Toward improved spectral measures of /s/: results from adolescents. *J Speech Lang Hear Res* **56**, 1175-1189 (2013).
69. Jacewicz, E. & Fox, R. A. Amplitude variations in coarticulated vowels. *J Acoust Soc Am* **123**, 2750-2768 (2008).
70. Shue, Y. L., Keating, P. & Vicens, C. *VOICESAUCE*, <<http://www.seas.ucla.edu/spapl/voicesauce/>> (2009).

71. Shue, Y.-L., Keating, P., Vicens, C. & Yu, K. VoiceSauce: A program for voice analysis. *J Acoust Soc Am* **126** (2009).
72. Sjölander, K. *The Snack Sound Toolkit*, <<https://www.speech.kth.se/snack/>> (2004).
73. Caverle, M. W. J. & Vogel, A. P. Stability, reliability, and sensitivity of acoustic measures of vowel space: A comparison of vowel space area, formant centralization ratio, and vowel articulation index. *J Acoust Soc Am* **148**, 1436 (2020).
74. Microsoft. *Microsoft Excel*, <<https://www.microsoft.com/en-us/microsoft-365/excel>> (2020).
75. IBM Corp. *IBM SPSS Software*, <<https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software>> (2018).
76. GraphPad Software. *Prism 8*, <<https://www.graphpad.com/scientific-software/prism/>> (2018).
77. Massey, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46**, 68-78 (1951).
78. Murphy, K. R. *Statistical power analysis : a simple and general model for traditional and modern hypothesis tests*. Fourth edn, (Routledge, 2014).
79. Cohen, J. A power primer. *Psychol Bull* **112**, 155-159 (1992).
80. Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A. & Johnson, K. Formants of children, women, and men: the effects of vocal intensity variation. *J Acoust Soc Am* **106**, 1532-1542 (1999).
81. Bradlow, A. R., Torretta, G. M. & Pisoni, D. B. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun* **20**, 255-272 (1996).
82. Mattys, S. L., Davis, M. H., Bradlow, A. R. & Scott, S. K. Speech recognition in adverse conditions: A review. *Lang Cogn Process* **27**, 953-978 (2012).

Figures

Figure 1

Wide-band spectrogram illustrating the extraction of the fricative [ʃ] in Praat. Vertical dashed lines indicate boundaries of this consonant.

Figure 2

Rating score of speech clarity (rating scale 1-100). Error bars indicate standard deviation.

Figure 3

Root-mean-square amplitude measured from middle 50ms of the two fricatives @ 6cm from the mouth.
Note: The data was not calibrated to real sound pressure level. NM, no-mask; SM, surgical mask.

Figure 4

Root-mean-square amplitude measured from middle 50ms of the edited vowels @ 6cm from the mouth.
Note: The data was not calibrated to real sound pressure level. NM, no-mask; SM, surgical mask.

Figure 5

Normalized formant amplitude of three extracted vowels. Error bars indicate standard deviation. Dots represent individual values. NM, no mask; SM, surgical mask.

Figure 6

Formant frequency of the three extracted vowels