The Journal of the Acoustical Society of America

MAY 20 2022

The impact of face masks on spectral acoustics of speech: Effect of clear and loud speech styles[a] FREE

Special Collection: COVID-19 Pandemic Acoustic Effects

Thea Knowles ; Gursharan Badh

J. Acoust. Soc. Am. 151, 3359–3368 (2022)
https://doi.org/10.1121/10.0011400

JASA ARTICLE

The impact of face masks on spectral acoustics of speech: Effect of clear and loud speech styles[a]

Thea Knowles[b] and Gursharan Badh
Department of Communicative Disorders and Sciences, University at Buffalo, Buffalo, New York 14215, USA

ABSTRACT:
This study quantified the effects of face masks on spectral speech acoustics in healthy talkers using habitual, loud, and clear speaking styles. Harvard sentence lists were read aloud by 17 healthy talkers in each of the 3 speech styles without wearing a mask, when wearing a surgical mask, and when wearing a KN95 mask. Outcome measures included speech intensity, spectral moments, and spectral tilt and energy in mid-range frequencies which were measured at the utterance level. Masks were associated with alterations in spectral density characteristics consistent with a low-pass filtering effect, although the effect sizes varied. Larger effects were observed for center of gravity and spectral variability (in habitual speech) and spectral tilt (across all speech styles). KN95 masks demonstrated a greater effect on speech acoustics than surgical masks. The overall pattern of the changes in speech acoustics was consistent across all three speech styles. Loud speech, followed by clear speech, was effective in remediating the filtering effects of the masks compared to habitual speech. © *2022 Acoustical Society of America*. https://doi.org/10.1121/10.0011400

(Received 4 January 2022; revised 2 May 2022; accepted 4 May 2022; published online 20 May 2022)
[Editor: Charles C. Church] Pages: 3359–3368

I. INTRODUCTION

In light of the COVID-19 pandemic, the United States Center for Disease Control (CDC) recommended that individuals wear face masks to prevent the spread of airborne viral particles and reduce disease transmission (CDC, 2020a). Face masks have been shown to act as a low-pass filter on speech, presumably because they act as a barrier to the acoustic signal. Many types of face masks attenuate acoustic energy above approximately 1–2 kHZ (e.g., Palmiero *et al.*, 2016; Corey *et al.*, 2020). Some types of face masks have also been shown to negatively affect speech intelligibility in healthy talkers (e.g., Bandaru *et al.*, 2020; Caniato *et al.*, 2021; Randazzo *et al.*, 2020; Toscano and Toscano, 2021).

Modifying our speaking style may be one way to overcome the effects of masks on speech. Although there is mounting evidence that speaking clearly improves intelligibility while wearing masks (Cohn *et al.*, 2021; Gutz *et al.*, 2021; Smiljanic *et al.*, 2021; Yi *et al.*, 2021), little is known about the acoustic characteristics of altered speech in masks. Furthermore, there is limited information of how other behavioral speech strategies, such as speaking loudly, impact speech production in masks. The current study quantified the effects of two face masks on spectral speech acoustics in young, healthy talkers across three speech styles: habitual, clear, and loud.

A. Face masks and spectral attenuation

In the spring of 2020, the CDC recommended several different types of masks that could be worn by the general public as a means of reducing transmission of COVID-19 (CDC, 2020b). Of these, two examples of widely available, disposable masks that meet a medical-grade standard include surgical masks and KN95 masks. Surgical masks (also known as medical procedure masks) are commonly made from nonwoven polypropylene fabric constructed of three layers (Chua *et al.*, 2020). KN95 masks are a type of disposable respirator that meets an international standard of quality regarding their effectiveness in filtering out very small particles. KN95 masks are similar in construction to N95 masks with the difference being that KN95 masks are not approved by the National Institute for Occupational Safety and Health (CDC, 2021).

Recent research has characterized a consistent pattern of a low-pass filter effect of masks in spite of methodological differences, including recording distance. This effect exists regardless of the type of material used for the masks, although attenuation is greater for thicker, more tightly woven materials compared to others (Corey *et al.*, 2020). Greater attenuation has been observed for KN95 masks compared to surgical masks (Atcherson *et al.*, 2020; Atcherson *et al.*, 2021; Nguyen *et al.*, 2021; Pörschmann *et al.*, 2020).

The attenuation of higher frequency acoustic information may directly or indirectly impact a listener's ability to understand what is being said when a talker wears a mask. Acoustic information that listener's use to distinguish individual speech sounds typically ranges between 300 Hz (e.g., for high vowels; Hillenbrand *et al.*, 1995) and 7000–8000 Hz for high frequency sounds such as /s/ (Jongman *et al.*, 2000). Lower energy in these frequency ranges may also make it difficult to identify certain sound classes. Indirectly, an attenuated signal may also simply make it more *difficult* for listeners to comprehend or recall what they are hearing because

---

[a] This paper is part of a special issue on COVID-19 Pandemic Acoustic Effects.

[b] Electronic mail: thea.knowles@gmail.com

they have to expend more effort to understand (Brown *et al.*, 2021; Truong *et al.*, 2021).

Nguyen *et al.* (2021) compared the effects of a surgical mask and KN95 mask on speech in 16 healthy talkers and found that both masks attenuated spectral levels between 1 and 8 kHz. The KN95 mask had a more detrimental effect with an attenuation of an average 5.2 dB compared to 2 dB from the surgical mask (recorded 6 cm from the mouth). Neither mask attenuated spectral information below 1 kHz, a finding consistent with previous research (Atcherson *et al.*, 2020; Atcherson *et al.*, 2021; Corey *et al.*, 2020; Goldin *et al.*, 2020). Pörschmann *et al.* (2020) reported peak attenuation between 3 and 5 kHz of an emphasized sine wave sweep to be approximately 7 dB and 15 dB with the surgical and KN95 masks, respectively, at a 2-m (6.6-ft) microphone distance. Atcherson *et al.* (2021) found similar degrees of attenuation at a 3-ft distance as well.

B. Face masks and speech intensity

While masks attenuate higher frequencies, generally, the overall vocal intensity appears to be less impacted. Fiorella *et al.* (2021) found that in 60 healthy talkers, wearing a surgical mask was not associated with a significant reduction in speech intensity of a sustained vowel. At an individual level, however, 65% of talkers demonstrated reduced speech intensity with the surgical mask on, whereas 35% demonstrated an increase. The authors suggested that some speakers may be unconsciously producing greater vocal effort to compensate for the filtering effects of the masks. Maryn *et al.* (2021) controlled for behavioral adjustments to masks by taking acoustic measures of prerecorded speech reproduced through a mannequin fitted in three distinct mask conditions as well as with no mask. Compared to no mask, they found no significant changes in intensity for standard surgical masks but did find reduced intensity for speech produced with a FFP2 mask (which are similar in filtration properties to N95 and KN95 masks) and a transparent window face mask on the order of 1.3 and 1.5 dB sound pressure level (SPL), respectively. Cohn *et al.* (2021) reported *higher* descriptive mean speech intensities on the order of 0.1–2 dB SPL for sentences produced with rather than without a fabric mask in three different speech styles (habitual, clear, and emotional) produced by two trained speakers. The authors suggested this was evidence that masks do not show an across-the-board pattern of intensity which distinguished face masks from no face masks. Overall, it appears that while masks may attenuate higher frequency components of the signal, they do not uniformly result in lower overall speech intensity.

C. Modified speech styles and spectral acoustics

To compensate for the filtration effects of face masks, speakers may need to adopt strategies to modify their speech to be better understood when wearing a face covering. Two strategies include speaking more clearly and/or loudly. Both of the clear and loud speaking styles have been shown to result in similar but not identical spectral changes to the speech signal. The changes across these two styles mirror those of and may be attributable to increased vocal effort (Rosenthal *et al.*, 2014).

Loud speech may refer to noise-adapted Lombard speech, in which talkers reflexively increase their speech intensity in response to background noise, or a modified speech style, in which talkers intentionally speak at a higher volume. It is often elicited by introducing background noise to a talker or instructing them to speak at a volume that feels louder to them. *Clear speech*, which tends to be produced in adverse listening scenarios (Smiljanić and Bradlow, 2009), is typically elicited by instructing a talker to speak more clearly, although specific instructions vary and have been shown to have a systematic impact on the resultant speech alterations (e.g., Lam *et al.*, 2012). In general, both clear and loud speech are produced with greater speech intensity, relative to habitual speech, with a greater increase observed for loud speech (Tjaden *et al.*, 2013b). Both of these styles are also associated with an increase in energy in higher frequency ranges of speech, leading to a flatter (less negative) spectral slope. Flatter spectral slopes in loud speech have been attributed to greater energy in the first formant range (Fant, 1960; Ternström *et al.*, 2006). This is likely, in part, due to jaw lowering that occurs, and the result is a lower rate of spectral roll-off. Clear speech has been associated with an increase in energy in mid-range frequencies (i.e., 1–3 kHz; Krause and Braida, 2004, 2009; Gilbert *et al.*, 2014; Hazan *et al.*, 2018; Hazan and Baker, 2011; Smiljanic, 2021).

D. Modified speech styles and face masks

In addition to acting as a low-pass filter, face masks have also been shown to negatively impact speech intelligibility, especially in adverse listening conditions. This also appears to differ by mask type with surgical masks demonstrating little to no effect for listeners with typical hearing (Atcherson *et al.*, 2017; Fecher and Watt, 2013; Mendel *et al.*, 2008) and thicker or more tightly woven masks, such as N95 masks, being more detrimental (Caniato *et al.*, 2021; Randazzo *et al.*, 2020). Recent work has found that speech produced using clear or loud speaking strategies yields improvements in intelligibility of speech produced with face masks (Cohn *et al.*, 2021; Gutz *et al.*, 2021; Smiljanic *et al.*, 2021; Yi *et al.*, 2021). Talkers may also be subconsciously altering their speech style in response to wearing masks. Cohn *et al.* (2021) found no significant effect of face masks on speech intelligiblity when talkers were speaking in a habitual, conversational manner. However, when talkers were instructed to speak clearly with and without a face mask, listeners were actually *more* accurate in understanding their speech when the mask was on. The opposite was true when speakers were instructed to speak "emotionally," suggesting that speakers conform to a *targeted adaptation approach* in which when the goal is increased clarity, talkers may further and, in fact, overcompensate for the presence of an additional adverse variable, namely, a face mask.

What is not known at present is the nature of the relationship between the filtering effect of face masks on speech and the adjustments of speech styles on spectral acoustics. To understand the intelligibility benefit of altered speech styles in the presence of face masks and make adequate recommendations, a better understanding of the acoustic outcomes of altered speech styles in the presence of masks is needed.

E. Purpose

In summary, the primary acoustic impact of face masks is attenuation of higher frequency components of speech. More effortful speech, achieved through either clear or loud speaking styles, is associated with increased spectral energy in higher frequency components. The purpose of this study was to quantify acoustic spectral characteristics of speech produced by live talkers with and without face masks in clear and loud altered speech styles. Two research questions were of interest:

(1) What is the impact of face masks on spectral acoustics of speech in unaltered (habitual) speech?, and
(2) what is the relationship between face masks and altered speech styles (clear and loud) on spectral acoustics of speech?

This study builds on existing work of the acoustic and perceptual consequences of face masks on speech by investigating the effects of masks on speech produced in ways that talkers might use to compensate for the effects of masks: speaking more clearly or loudly.

II. METHODS

This study was approved by the Institutional Review Board at the University at Buffalo. Seventeen healthy adults with no history of speech, language, hearing, or neurological concerns (16 females and 1 male; mean age, 24 years old; age range, 20–42 years old) read aloud sentences from the Harvard sentence corpus (IEEE, 1969) in 3 face mask conditions and 3 speech style conditions. The face mask conditions included no mask, a standard disposable surgical mask, and a disposable KN95 mask. The speaking styles included habitual, loud, and clear.

All of the speakers began with the habitual style. The order of clear and loud speech conditions was counterbalanced across participants. The orders of face masks within and across each condition, as well as the order of Harvard sentence lists, were randomized for each participant to avoid order effects. All of the three mask types were worn for each of the three speech conditions, resulting in nine total conditions per participant. Within each condition, speakers read aloud two Harvard sentence lists (lists 1–18 were included for this study; IEEE, 1969).

The instructions for the clear speech condition were "speak clearly by overarticulating your speech, similar to how you might speak to someone who is having difficulty hearing you, or someone who is learning English and is having difficulty understanding you." The instructions for loud speech were to "speak at a volume that feels two times louder than your normal speaking voice." For both of the conditions, participants were given the opportunity to practice reading an additional subset of sentences aloud (not included in the stimuli) before beginning the block.

Participants were recorded in a sound-treated room and positioned 6 in. from a table top microphone (Shure SM58, Niles, IL). A second microphone (also a Shure SM58) was positioned at a 2-m distance. The results presented are from recordings made at the 6-in. distance. Prior to the experiment, a 1000 Hz tone of a fixed intensity was played via a small loudspeaker positioned under the chin of the participant. This tone was played and recorded three times and its intensity was measured via a sound level meter (Galaxy Audio CM-170, Wichita, KS) positioned adjacent to the microphone. The average intensity of this tone was used to calibrate the speech signal intensity for each participant.

A. Acoustic measures

The acoustic measures of interest included spectral measures known to be sensitive to the potential filtering characteristics of the masks (i.e., measures of spectral tilt; Nguyen *et al.*, 2021; Corey *et al.*, 2020) as well as measures known to be sensitive to speaking style (i.e., 1–3 kHz; Krause and Braida, 2004, 2009; Gilbert *et al.*, 2014; Hazan *et al.*, 2018; Hazan and Baker, 2011; Smiljanic, 2021). To address research question 1, this included overall speech intensity as well as four spectral moments (center of gravity, standard deviation of center of gravity, skewness, and kurtosis). The acoustic measures were taken from utterances produced in the habitual speech condition. The mean intensity was measured at the utterance level, and spectral moments were extracted from the long-term average spectrum (LTAS) of each utterance, characterizing the central tendency and shape of the speech frequency distribution in Praat (Boersma and Weenink, 2021).

To address research question 2, two measures related to spectral tilt were of interest: the total mean energy in the 1–3 kHz range and the difference in energy between 0 and 1 kHz and 1 and 10 kHz. Higher amounts of mean energy in the 1–3 kHz range are representative of increased vocal effort and have been associated with increased intelligibility (Hazan and Markham, 2004; Krause and Braida, 2004). A lower amount of energy in the higher frequency range (>1 kHz) is captured by a steeper or more negative spectral tilt. Steeper tilt has been associated with lower perceived loudness, effort, and intelligibility (Lu and Cooke, 2009).

B. Statistical analysis

All measures of interest were modelled as a function of the mask condition and, in the case of research question 2, speaking style, as well as the mask-by-speech style interaction, using linear mixed effects regression. To test whether observed patterns persisted at close and far recording distances, two sets of models were run for research question 2:

a main set of models on recordings made at the 6-in. distance, and a secondary set of models at the 2-m distance. All models included random by-participant and by-item intercepts. Models addressing research question 2 also included by-participant random slopes for speaking style, although the 2-m recording distance models required a simplified random slopes structure to prevent model non-convergence. Face mask and speaking style were both contrast coded using reverse Helmert contrasts with three levels. Baseline levels were set to no mask and habitual speech, respectively. This contrast scheme permits the mean of the baseline level to be compared to the overall mean of the subsequent levels and the means of the other two levels to be compared to each other. The interpretation is as follows for the mask: (a) no mask vs mask (i.e., the overall mean of the surgical and KN95 masks) and (b) surgical mask vs KN95 mask, and for thespeaking style: (a) habitual vs altered speech (i.e., overall mean of clear and loud speech) and (b) clear vs loud speech. For example, a positive model estimate for no mask vs mask would indicate a lower overall mean value for a given outcome when talkers were not wearing a mask compared to when wearing a mask, which is averaged across the mask types. A negative beta estimate for, e.g., clear vs loud, speaking styles would indicate a lower mean value for clear speech compared to loud speech, and so on.

The effect sizes were calculated for each model predictor by dividing the estimate by the square root of the total variance of the random effects (i.e., the sum of the variance for each random effects term in the model and total residual variance; Westfall *et al.*, 2014). Here, we refer to our effect sizes using traditional Cohen's *d* cutoffs (Cohen, 1962) as a means of comparing effects within this study, keeping in mind caveats when computing effects sizes for mixed models.[1] Cohen's *d* cutoffs suggest the following effect size interpretation for small, medium, and large effect sizes, respectively: 0.2, 0.5, and 0.8. The effect sizes less than 0.2 are considered negligible, and large effect sizes may exceed the value of one.

III. RESULTS

A. Research question 1: Effect of masks in habitual speech

The results for research question 1 are reported in Table I. In habitual speech, compared to baseline (no mask), wearing a mask was associated with lower speech intensity, higher center of gravity (COG) and COG variability, and lower skewness and kurtosis. These effects can be seen in Table I for the contrast "no mask vs mask." All of the effects significantly differed at $p < 0.001$, although the size of each effect varied. The large effect sizes (>0.8) were observed for COG variability ($\hat{\beta} = -393.853$, $p < 0.001$). The medium effect sizes (0.5–0.8) were observed for COG, skewness, kurtosis, and spectral tilt (estimates: COG, $\hat{\beta} = -169.896$, $p < 0.001$; skewness, $\hat{\beta} = 1.051$, $p < 0.001$; kurtosis, $\hat{\beta} = 30.744$, $p < 0.001$; tilt, $\hat{\beta} = -1.009$, $p < 0.001$). The negligible effect sizes (< 0.2) were found for intensity, which was estimated to differ by approximately 0.6 dB SPL ($\hat{\beta} = -0.623$, $p < 0.001$), and mid-range frequencies ($\hat{\beta} = -0.913$, $p < 0.001$).

The same general direction of results was found when comparing the two masks ("SM vs KN"), suggesting a greater filtering effect of the KN95 mask compared to the surgical mask. The spectral moments were all significantly altered when the talker wore a KN95 mask compared to the surgical

TABLE I. The model results for research question 1, showing the effects of masks in habitual speech, and the model estimates for each outcome measure are grouped by fixed effects terms.

Contrast	Measure	Estimate	Standard error	*t*	*p*	Effect size parameter
(Intercept)	Mid-range	10.251	0.980	10.459	<0.001	2.205
	COG	754.679	42.439	17.783	<0.001	3.357
	COG SD	909.761	60.711	14.985	<0.001	2.568
	Intensity	76.166	0.714	106.623	<0.001	24.038
	Kurtosis	59.545	7.093	8.394	<0.001	1.458
	Skewness	5.691	0.349	16.323	<0.001	2.950
	Tilt	–16.170	0.587	–27.547	<0.001	5.237
NM vs Mask	Mid-range	–0.913	0.157	–5.820	<0.001	0.196
	COG	–169.896	9.671	–17.568	<0.001	0.756
	COG SD	–393.853	17.251	–22.831	<0.001	1.112
	Intensity	–0.623	0.080	–7.765	<0.001	0.196
	Kurtosis	30.744	1.950	15.765	<0.001	0.753
	Skewness	1.051	0.088	11.949	<0.001	0.545
	Tilt	–1.009	0.132	–7.667	<0.001	0.327
SM vs KN	Mid-range	0.052	0.181	0.288	0.774	0.011
	COG	–48.734	11.155	–4.369	<0.001	0.217
	COG SD	–159.207	19.899	–8.001	<0.001	0.449
	Intensity	0.090	0.092	0.978	0.328	0.029
	Kurtosis	21.234	2.250	9.439	<0.001	0.520
	Skewness	0.473	0.101	4.657	<0.001	0.245
	Tilt	–0.297	0.152	–1.955	0.051	0.096

mask (estimates: COG, $\hat{\beta} = -48.734$, $p < 0.001$; COG variability, $\hat{\beta} = -159.207$, $p < 0.001$; skewness, $\hat{\beta} = 0.473$, $p < 0.001$; kurtosis, $\hat{\beta} = 21.234$, $p < 0.001$). The effect sizes were overall smaller between the two masks with a medium effect size found for kurtosis and small effect sizes found for COG, COG variability, and skewness. No significant differences were found for intensity ($\hat{\beta} = 0.09$, $p = 0.328$), mid-range frequencies ($\hat{\beta} = 0.052$, $p = 0.774$), or spectral tilt ($\hat{\beta} = -0.297$, $p = 0.051$).

B. Research question 2: Effect of masks and altered speech styles

The results for the 6-in. recording distance are pictured in Figs. 1 and 2 and summarized in Table II. The results for the 2-m distance are reported later in the text and summarized in Table III. The presence of masks demonstrated a systematic, significant effect on all spectral measures compared to not wearing a mask when the speaking condition was held constant. In Tables II and III, the no mask vs mask contrast ("NM vs mask") captures the overall pooled effect of the two mask types, and thes mask vs KN95 mask contrast ("SM vs KN") captures the differences between the two types. Both comparisons account for the effects when outcomes for the different speech styles are set to their average values.

To reiterate, three of the outcome measures from research question 1 were used in the models to address research question 2: mid-range frequency energy (1–3 kHz), spectral tilt, and speech intensity. All three of the measures were found to be sensitive to the speaking style and presence and type of face mask ($p < 0.001$ for all main effects of style and mask across all three of the models). Overall, the patterns observed across the altered speech styles mirrored those of habitual speech. A significant main effect of mask was found for all three of the measures, that is, when all speech styles were held at their average values. Masks, compared to no mask, were associated with less energy in mid-range frequencies ($\hat{\beta} = -0.98$, $p < 0.001$), lower (more negative) spectral tilt ($\hat{\beta} = -1.192$, $p < 0.001$), and lower speech intensity ($\hat{\beta} = -0.574$, $p < 0.001$). Masks, compared to no mask, were associated with less energy in mid-range frequencies and lower (more negative) spectral tilt. Changes in spectral tilt showed a medium effect size while the effects for speech intensity and mid-range frequency energy were negligible. Even with the two altered speech styles held at their average values, the intensity differences for the masks were on the order of 0.5 dB SPL. Compared to the KN95 mask, the surgical mask was associated with flatter tilt ($\hat{\beta} = -0.494$, $p < 0.001$, negligible effect size) but did not significantly differ for mid-range frequencies ($\hat{\beta} = -0.172$, $p = 0.216$) or speech intensity ($\hat{\beta} = -0.044$, $p = 0.575$).

Compared to habitual speech, clear and loud speech together were associated with higher intensity ($\hat{\beta} = 5.284$, $p < 0.001$), greater mid-range frequency energy ($\hat{\beta} = 7.686$, $p < 0.001$), and flatter spectral tilt ($\hat{\beta} = 3.252$, $p < 0.001$), all of which constituted large effects. Loud speech, compared to clear speech, demonstrated this same pattern and was reflected by large effect sizes for all of the outcomes

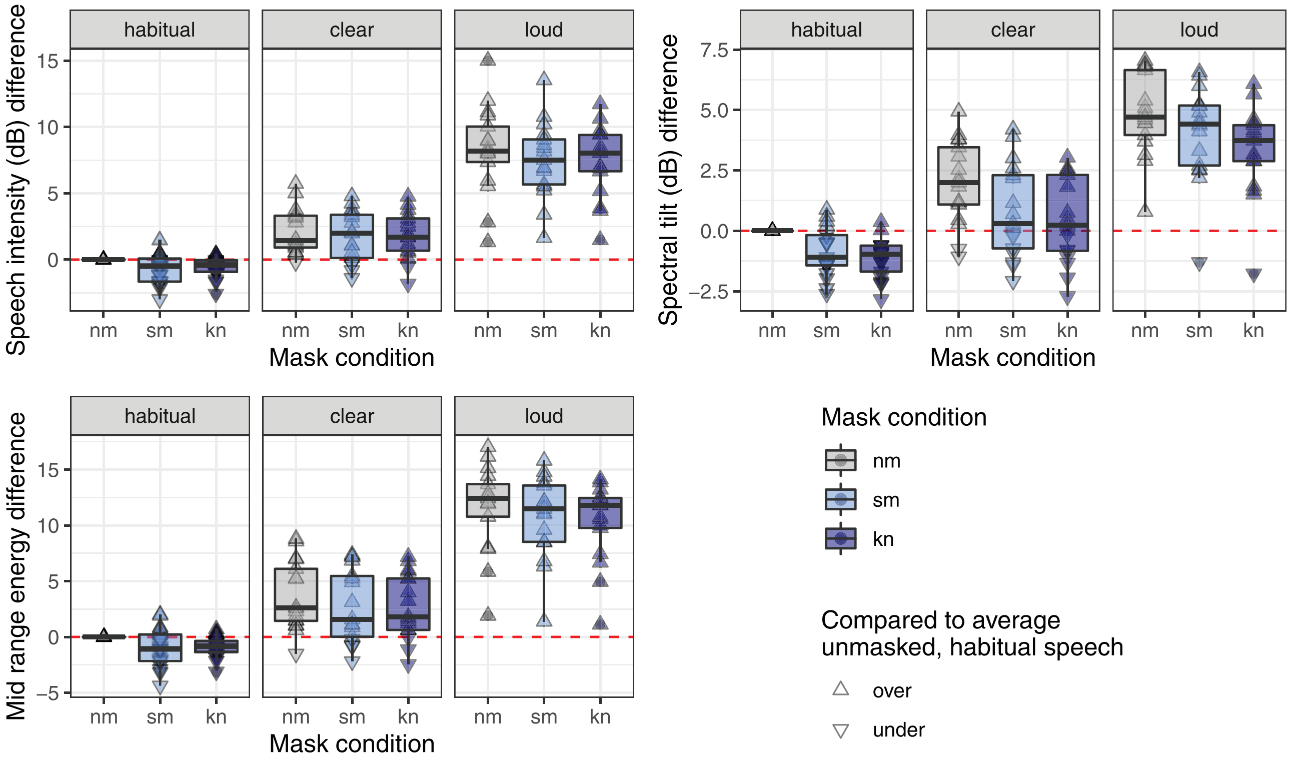

FIG. 1. (Color online) The acoustic measures of interest by speech style (habitual, clear, and loud) and mask type (no mask, surgical mask, and KN95 mask). The horizontal dashed line reflects the individual participants' baseline (no mask and habitual speech condition).

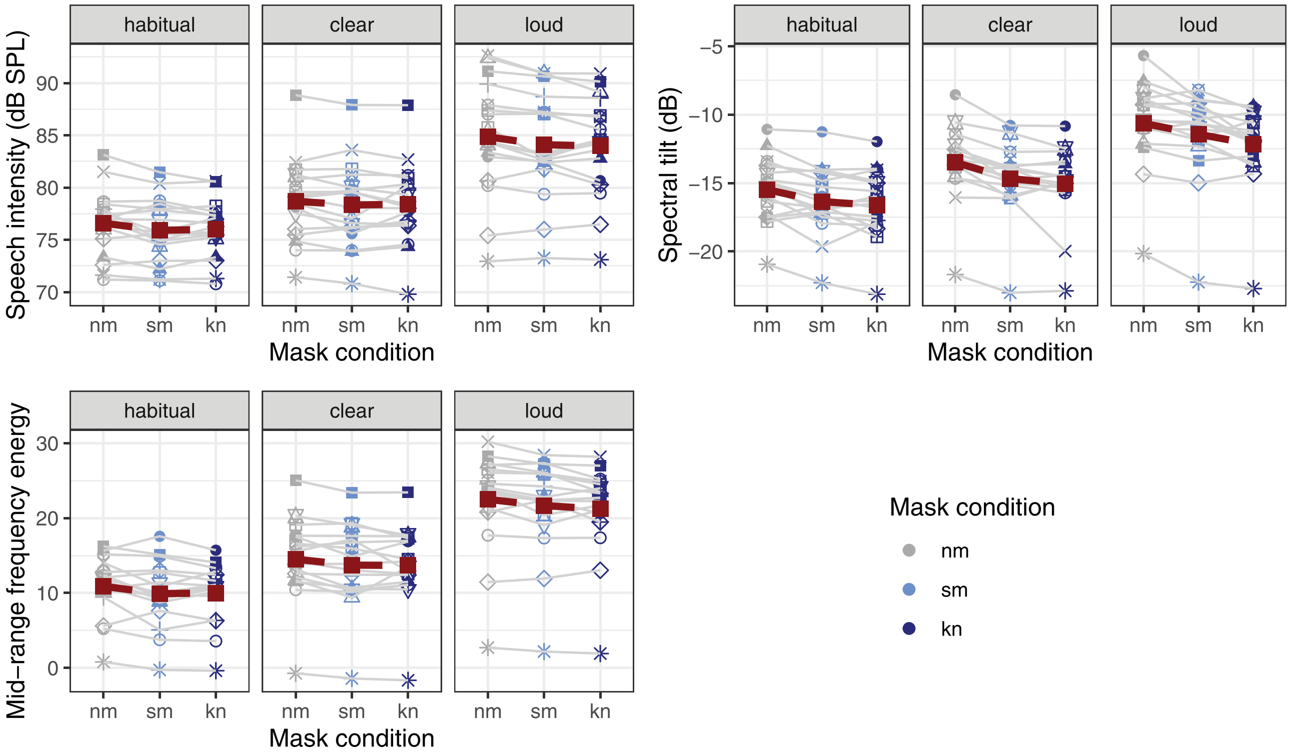

FIG. 2. (Color online) The differences in acoustic measures of interest for each individual speaker compared to the baseline (habitual speech without a face mask) by speech style (clear and loud) and mask type (surgical mask and KN95 mask). The red dashed line reflects the group mean.

TABLE II. The model results for research question 2, showing the effects of masks across habitual, clear, and loud speech styles (6-in. microphone distance). The model estimates for each outcome measure are grouped by fixed effects and interaction terms.

Contrast	Measure (6-in. distance)	Estimate	Standard error	*t*	*p*	Effect size parameter
(Intercept)	Intensity	79.688	0.922	86.417	<0.001	14.956
	Mid-range	15.382	1.204	12.779	<0.001	2.634
	Tilt	–14.002	0.631	–22.192	<0.001	3.304
NM vs mask	Intensity	–0.574	0.068	–8.435	<0.001	0.108
	Mid-range	–0.980	0.119	–8.215	<0.001	0.168
	Tilt	–1.192	0.075	–15.897	<0.001	0.281
SM vs KN	Intensity	–0.044	0.079	–0.561	0.575	0.008
	Mid-range	–0.172	0.139	–1.238	0.216	0.029
	Tilt	–0.494	0.087	–5.666	<0.001	0.117
Clear vs loud	Intensity	5.723	0.079	72.526	<0.001	1.074
	Mid-range	7.711	0.138	55.803	<0.001	1.321
	Tilt	2.986	0.411	7.272	<0.001	0.705
Clear vs loud:NM vs mask	Intensity	–0.459	0.166	–2.762	0.006	0.086
	Mid-range	–0.139	0.291	–0.477	0.633	0.024
	Tilt	0.194	0.183	1.059	0.29	0.046
Clear vs loud:SM vs KN	Intensity	–0.172	0.194	–0.889	0.374	0.032
	Mid-range	–0.341	0.340	–1.003	0.316	0.058
	Tilt	–0.379	0.214	–1.772	0.076	0.089
Habit vs altered	Intensity	5.284	0.539	9.798	<0.001	0.992
	Mid-range	7.686	0.120	64.085	<0.001	1.316
	Tilt	3.252	0.341	9.535	<0.001	0.767
Habit vs altered:NM vs mask	Intensity	0.072	0.145	0.496	0.62	0.013
	Mid-range	–0.008	0.254	–0.031	0.975	0.001
	Tilt	–0.271	0.159	–1.697	0.09	0.064
Habit vs altered:SM vs KN	Intensity	–0.201	0.168	–1.197	0.231	0.038
	Mid-range	–0.521	0.294	–1.774	0.076	0.089
	Tilt	–0.303	0.185	–1.641	0.101	0.072

TABLE III. The model results for research question 2, showing the effects of masks across habitual, clear, and loud speech styles (2-m microphone distance). The model estimates for each outcome measure are grouped by fixed effects and interaction terms.

Contrast	Measure (2-m distance)	Estimate	Standard error	*t*	*p*	Effect size parameter
(Intercept)	Intensity	60.968	0.749	81.397	<0.001	13.755
	Mid-range	–6.013	1.175	–5.119	<0.001	0.868
	Tilt	–16.232	0.709	–22.882	<0.001	4.091
NM vs mask	Intensity	–0.414	0.062	–6.659	<0.001	0.093
	Mid-range	–1.038	0.108	–9.650	<0.001	0.150
	Tilt	–2.158	0.079	–27.388	<0.001	0.544
SM vs KN	Intensity	0.157	0.072	2.167	0.03	0.035
	Mid-range	–0.240	0.125	–1.919	0.055	0.035
	Tilt	–0.930	0.092	–10.149	<0.001	0.234
Clear vs loud	Intensity	5.743	0.072	79.600	<0.001	1.296
	Mid-range	7.768	0.125	62.300	<0.001	1.122
	Tilt	2.866	0.091	31.387	<0.001	0.722
Clear vs loud:NM vs mask	Intensity	–0.310	0.152	–2.043	0.041	0.070
	Mid-range	0.036	0.263	0.138	0.89	0.005
	Tilt	0.106	0.192	0.553	0.58	0.027
Clear vs loud:SM vs KN	Intensity	–0.088	0.177	–0.496	0.62	0.020
	Mid-range	–0.201	0.307	–0.654	0.513	0.029
	Tilt	–0.237	0.225	–1.056	0.291	0.060
Habit vs altered	Intensity	5.237	0.472	11.090	<0.001	1.181
	Mid-range	7.974	0.683	11.675	<0.001	1.152
	Tilt	3.316	0.342	9.706	<0.001	0.836
Habit vs altered:NM vs mask	Intensity	–0.062	0.132	–0.469	0.639	0.014
	Mid-range	–0.032	0.229	–0.140	0.888	0.005
	Tilt	–0.148	0.168	–0.882	0.378	0.037
Habit vs altered:SM vs KN	Intensity	0.093	0.153	0.608	0.543	0.021
	Mid-range	0.128	0.265	0.484	0.628	0.019
	Tilt	–0.019	0.194	–0.097	0.923	0.005

(intensity, $\hat{\beta} = 5.723$, $p < 0.001$; mid-range frequencies, $\hat{\beta} = 7.711$, $p < 0.001$; spectral tilt, $\hat{\beta} = 2.986$, $p < 0.001$). No significant mask-by-speech-style interactions were found for any of the measures with the exception of speech intensity. For the spectral measures, this indicates that the general effects of the masks persisted across the three speaking styles. A two-way interaction ($p = 0.006$, negligible effect size) for intensity was found for the clear vs loud and no mask vs mask comparisons on the order of <0.5 dB SPL ($\hat{\beta} = –0.459$, $p = 0.006$). Further visual inspection of the data revealed that in loud speech, talkers produced greater speech intensity without a mask than with one, but in clear speech, the differences between masked and unmasked speech intensity were much smaller.

C. Effect of microphone distance

Lower values were found for speech intensity, mid-range frequency energy, and spectral tilt at the 2-m compared to at the 6-in. recording distance. This is reflected in the intercept values (value when all fixed effects are held at their constant value) in Table III. The patterns of the effects of masks and speaking style, however, were very similar to those identified at the 6-in. distance with some minor differences. Specifically, effect sizes for the mask comparisons were larger for spectral tilt but not for mid-range frequencies, although the overall pattern of results did not change for either outcome. As can be seen in Fig. 3, this is reflected by a steeper drop in spectral tilt across the masks in the 2-m distance. Higher speech intensity in surgical vs KN95 masks was found, and this was established to be significant at $p < 0.05$ in the 2-m distance model. However, effect sizes remained negligible in this model and reflected a difference of <0.2 dB SPL ($\hat{\beta} = 0.157$, $p = 0.03$).

IV. DISCUSSION

Consistent with previous literature, the face masks in this study provided further evidence of a low-pass filtering effect of masks, demonstrated by a systematic effect of masks on spectral density and tilt characteristics. The magnitude of this effect was greater for the KN95 mask compared to the surgical mask. The overall pattern of the masks on speech acoustics was preserved across all three of the speaking styles. However, as predicted, speaking clearly and/or loudly resulted in increased spectral tilt measures, which had the effect of amplifying the mid-range to high frequencies that were attenuated by the masks. In other words, while wearing a mask was consistently found to filter out higher frequency components of the speech signal, regardless of the style in which speech is spoken, speaking loudly or clearly while wearing a mask was found to compensate for this filtering effect compared to speaking in a conversational style with a mask.

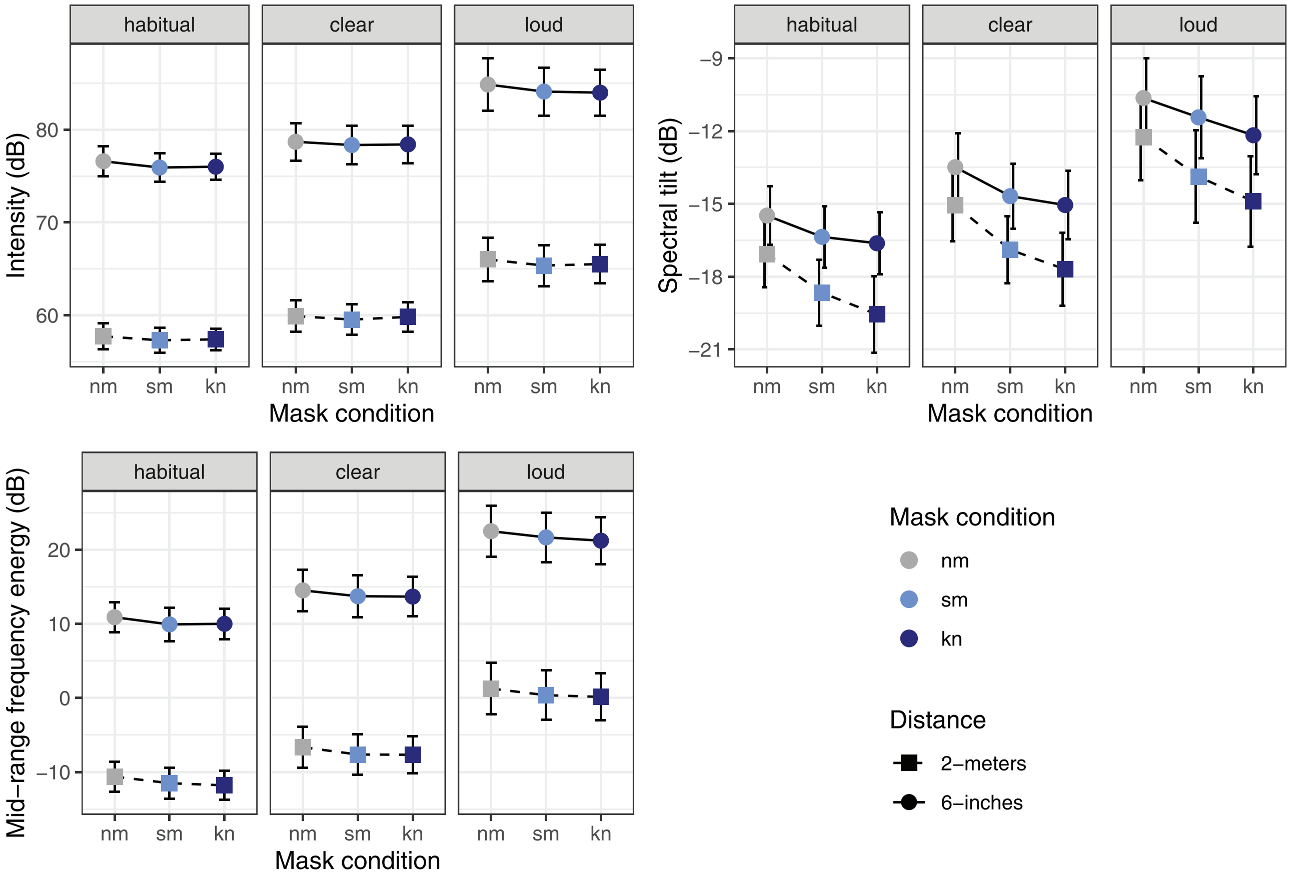

FIG. 3. (Color online) The differences in acoustic outcomes by recording distance (6 in., 2 m), speech style (habitual, clear, and loud), and mask type (no mask, surgical mask, and KN95 mask). The points represent mean values aggregated over the speaker means. The error bars represent the standard errors.

Averaged across all of the speech conditions, there was a systematic, predictable effect of masks on spectral acoustics. Compared to speech without a mask, masks were associated with significantly steeper spectral tilt and, to a lesser extent, lower energy in mid-range frequencies and a small reduction in speech intensity. This is consistent with previous findings of spectral tilt (Nguyen *et al.*, 2021). The present study also found medium to large effects of the masks on the center of gravity and center of gravity variability. This is inconsistent with the findings of Maryn *et al.* (2021), who reported no significant effects of masks on these spectral moments of prerecorded vowel prolongations. The differences in this study could be attributable to the speech stimuli; the spectral moments of the LTAS of connected speech samples may be more sensitive to capturing the filtering effects of masks. This study also included the speech of live talkers, rather than prerecorded speakers, who could be making additional compensatory or maladaptive changes in response to wearing a mask.

Averaged across all mask conditions, loud, followed by clear speech, had the opposite effect of the masks: significant flattening of spectral tilt, greater energy in mid-range frequencies, and increased speech intensity. These patterns of altered speech styles persisted across the different mask conditions for the acoustic measures of interest, captured by an absence of two-way interactions between mask and speaking style conditions. The observed interactions reflected differences in the magnitude of change across the masks rather than a difference in the general direction of the results. For example, no significant two-way mask-style interactions were found for spectral tilt. A two-way interaction was observed for COG for the habitual vs altered contrast and the no mask vs mask contrast. In Fig. 1, this is evident as a greater difference for the two face masks in loud speech. The general pattern, however, is maintained. Loud speech, rather than clear speech, was associated with the greatest change (flatter tilt, higher COG, lower skewness and kurtosis). In essence, the removal (or absence) of a face mask had the same overall pattern of effects on spectral density characteristics of speech as did speaking more loudly or clearly. The effect sizes, however, were much larger for altered speech styles compared to the presence or absence of a face mask.

A secondary finding of this research was that while greater distance was predictably associated with lower speech intensity, spectral tilt, and mid-range frequency energy, the pattern of effects was preserved across masks and speech styles. The larger effects, however, were observed for spectral tilt, which likely represents greater acoustic attenuation at greater distances. This is consistent with previous research reporting greater attenuation from masks recorded at a 6-ft compared to 3-f distance, on the order of 5 dB between 2 and 8 kHz (Atcherson *et al.*, 2021). Compared to no mask, Atcherson *et al.* (2021) reported only a 1–2 dB attenuation at a greater distance though, which is consistent with the results of the present study: The pattern holds with only a slight increase in the magnitude of effects for spectral tilt. The degree to which this increased distance and subsequent signal attenuation in combination with masks affects a listeners' ability to understand the speech remains an open question.

While perceptual outcomes were not included in the present study, findings may help identify causal relationships between speech acoustics and auditory-perceptual consequences of speech produced in masks. Gutz *et al.* (2021) found that while both of the loud and clear speech styles were associated with increases in automatic speech recognition accuracy for talkers wearing KN95 masks, larger effects were observed for clear speech. Clear speech in masks was also associated with larger increases in vowel space, which is consistent with previous studies of clear speaking characteristics (Tjaden *et al.*, 2013a). That is, while loud compared to clear speech is associated with greater increases in mid-range frequencies and spectral tilt, which are attenuated by the face masks, it may be the case that other segmental adjustments unrelated to the filtering effects of the masks are still responsible for maximizing intelligibility in masks.

Attenuation from masks may also simply make it more *difficult* for listeners to comprehend or recall what they are hearing because they have to expend more effort to understand a degraded signal (Brown *et al.*, 2021; Truong *et al.*, 2021). The attenuation imposed by masks may impact segmental speech perception. Previous research has shown that face coverings do impact consonant perception, although in ideal listening conditions, this effect tends to be small, especially for surgical masks (Fecher and Watt, 2013; Llamas *et al.*, 2008). Clear and loud speech have been shown to increase consonant and vowel distinctiveness for healthy talkers and talkers with dysarthria (Tjaden *et al.*, 2013a; Tjaden and Martel-Sauvageau, 2017). An open question remains as to whether these acoustic alterations aid in improved intelligibility at the word and/or phoneme level when talkers don masks and whether these relationships persist for degraded listening conditions, such as the presence of background noise, or for talkers with speech disorders.

In conclusion, this study provided further evidence of the damping effect of face masks on speech. Speaking more loudly, followed by more clearly, enhances spectral characteristics of speech that are degraded by the presence of face masks. The findings may have implications for talkers with degraded voice quality due to disordered speech or voice production. The results from the present study will inform future research regarding potential underlying causes of changes in perceptual speech outcomes as a result of wearing masks.

---

[1] Brysbaert and Stevens (2018) caution that the approach proposed by Westfall *et al.* (2014), which is designed for simple mixed effects model structures, may provide inflated measures of effect sizes and may not be directly comparable to classic Cohen's d. In their paper, Westfall *et al.* (2014) suggest that this approach in theory could be applied to more complex model designs, but acknowledge that this remains an open issue.

Atcherson, S., Finley, E., McDowell, B., and Watson, C. (**2020**). "More speech degradations and considerations in the search for transparent face coverings during the COVID-19 pandemic," available at https://www.audiology.org/audiology-today-novemberdecember-2020/more-speech-degradations-and-considerations-search-transparent (Last viewed December 18, 2021).

Atcherson, S. R., McDowell, B. R., and Howard, M. P. (**2021**). "Acoustic effects of non-transparent and transparent face coverings," J. Acoust. Soc. Am. **149**(4), 2249–2254.

Atcherson, S. R., Mendel, L. L., Baltimore, W. J., Patro, C., Lee, S., Pousson, M., and Spann, M. J. (**2017**). "The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss," J. Am. Acad. Audiol. **28**(1), 058–067.

Bandaru, S. V., Augustine, A. M., Lepcha, A., Sebastian, S., Gowri, M., Philip, A., and Mammen, M. D. (**2020**). "The effects of N95 mask and face shield on speech perception among healthcare workers in the coronavirus disease 2019 pandemic scenario," J. Laryngol. Otol. **134**(10), 895–898.

Boersma, P., and Weenink, D. (**2021**). "Praat: Doing phonetics by computer (version 6.1.35) [computer program]," http://www.praat.org/ (Last viewed December 18, 2021).

Brown, V. A., Van Engen, K. J., and Peelle, J. E. (**2021**). "Face mask type affects audiovisual speech intelligibility and subjective listening effort in young and older adults," Cogn. Res. **6**, 49.

Brysbaert, M., and Stevens, M. (**2018**). "Power analysis and effect size in mixed effects models: A tutorial," J. Cogn. **1**(1), 9.

Caniato, M., Marzi, A., and Gasparella, A. (**2021**). "How much COVID-19 face protections influence speech intelligibility in classrooms?," Appl. Acoust. **178**, 108051.

CDC (**2020a**). "Guidance for wearing masks," Centers for Disease Control and Prevention, Atlanta, GA, available at https://www.cdc.gov/coronavirus/2019-nCoV/index.html (Last viewed February 17, 2022).

CDC (**2020b**). "Types of masks and respirators," Centers for Disease Control and Prevention, Atlanta, GA, available at https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/types-of-masks.html (Last viewed February 17, 2022).

CDC (**2021**). "National Institute for Occupational Safety and Health (NIOSH)," Atlanta, GA, available at https://www.cdc.gov/niosh/index.htm (Last viewed February 17, 2022).

Chua, M. H., Cheng, W., Goh, S. S., Kong, J., Li, B., Lim, J. Y. C., Mao, L., Wang, S., Xue, K., Yang, L., Ye, E., Zhang, K., Cheong, W. C. D., Tan, B. H., Li, Z., Tan, B. H., and Loh, X. J. (**2020**). "Face masks in the new COVID-19 normal: Materials, testing, and perspectives," Research **2020**, 1–40.

Cohen, J. (**1962**). "The statistical power of abnormal-social psychological research: A review," J. Abnorm. Soc. Psychol. **65**(3), 145–153.

Cohn, M., Pycha, A., and Zellou, G. (**2021**). "Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech," Cognition **210**, 104570.

Corey, R. M., Jones, U., and Singer, A. C. (**2020**). "Acoustic effects of medical, cloth, and transparent face masks on speech signals," J. Acoust. Soc. Am. **148**(4), 2371–2375.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton de Gruyter, Berlin).

Fecher, N., and Watt, D. (**2013**). "Effects of forensically-realistic facial concealment on auditory-visual consonant recognition in quiet and noise conditions," in *Auditory-Visual Speech Processing (AVSP), 2013*.

Fiorella, M. L., Cavallaro, G., Di Nicola, V., and Quaranta, N. (**2021**). "Voice differences when wearing and not wearing a surgical mask," J. Voice (published online).

Gilbert, R. C., Chandrasekaran, B., and Smiljanic, R. (**2014**). "Recognition memory in noise for speech of varying intelligibility," J. Acoust. Soc. Am. **135**(1), 389–399.

Goldin, A., Weinstein, B., and Shiman, N. (**2020**). "How do medical masks degrade speech reception?," Hear. Rev. **27**(5), 8–9.

Gutz, S., Rowe, H., and Green, J. (**2021**). "Speaking with a KN95 face mask: ASR performance and speaker compensation," in *Proceedings of Interspeech 2021*, pp. 4798–4802.

Hazan, V., and Baker, R. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**(4), 2139–2152.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**(5), 3108–3118.

Hazan, V., Tuomainen, O., Kim, J., Davis, C., Sheffield, B., and Brungart, D. (**2018**). "Clear speech adaptations in spontaneous speech produced by young and older adults," J. Acoust. Soc. Am. **144**(3), 1331–1346.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**(3), 1252–1263.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**(1), 362–378.

Krause, J. C., and Braida, L. D. (**2009**). "Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech," J. Acoust. Soc. Am. **125**(5), 3346–3357.

Lam, J., Tjaden, K., and Wilding, G. (**2012**). "Acoustics of clear speech: Effect of instruction," J. Speech. Lang. Hear. Res. **55**(6), 1807–1821.

Llamas, C., Harrison, P., Donnelly, D., and Watt, D. (**2008**). "Effects of different types of face coverings on speech acoustics and intelligibility," York Papers Ling. Ser. **2**(9), 80–104.

Lu, Y., and Cooke, M. (**2009**). "The contribution of changes in $F0$ and spectral tilt to increased intelligibility of speech produced in noise," Speech Commun. **51**(12), 1253–1262.

Maryn, Y., Wuyts, F. L., and Zarowski, A. (**2021**). "Are acoustic markers of voice and speech signals affected by nose-and-mouth-covering respiratory protective masks?," J. Voice (published online).

Mendel, L. L., Gardino, J. A., and Atcherson, S. R. (**2008**). "Speech understanding using surgical masks: A problem in health care?," J. Am. Acad. Audiol. **19**(9), 686–695.

Nguyen, D. D., McCabe, P., Thomas, D., Purcell, A., Doble, M., Novakovic, D., Chacon, A., and Madill, C. (**2021**). "Acoustic voice characteristics with and without wearing a facemask," Sci. Rep. **11**(1), 5651.

Palmiero, A. J., Symons, D., Morgan, J. W. III, and Shaffer, R. E. (**2016**). "Speech intelligibility assessment of protective facemasks and air-purifying respirators," J. Occup. Environ. Hyg. **13**(12), 960–968.

Pörschmann, C., Lübeck, T., and Arend, J. M. (**2020**). "Impact of face masks on voice radiation," J. Acoust. Soc. Am. **148**(6), 3663–3760.

Randazzo, M., Koenig, L. L., and Priefer, R. (**2020**). "The effect of face masks on the intelligibility of unpredictable sentences," Proc. Mtgs. Acoust. **42**, 032001.

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

Smiljanic, R. (**2021**). "Clear speech perception," in *The Handbook of Speech Perception* (Wiley, New York), pp. 177–205.

Smiljanić, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Lang. Linguist. Compass **3**(1), 236–264.

Smiljanic, R., Keerstock, S., Meemann, K., and Ransom, S. M. (**2021**). "Face masks and speaking style affect audio-visual word recognition and memory of native and non-native speech," J. Acoust. Soc. Am. **149**(6), 4013–4023.

Ternström, S., Bohman, M., and Södersten, M. (**2006**). "Loud speech over noise: Some spectral attributes, with gender differences," J. Acoust. Soc. Am. **119**(3), 1648–1665.

Tjaden, K., Lam, J., and Wilding, G. (**2013a**). "Vowel acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear, loud, and slow speaking conditions," J. Speech. Lang. Hear. Res. **56**(5), 1485–1502.

Tjaden, K., and Martel-Sauvageau, V. (**2017**). "Consonant acoustics in Parkinson's disease and multiple sclerosis: Comparison of clear and loud speaking conditions," Am. J. Speech. Lang. Pathol. **26**(2S), 569–582.

Tjaden, K., Richards, E., Kuo, C., Wilding, G., and Sussman, J. (**2013b**). "Acoustic and perceptual consequences of clear and loud speech," Folia Phoniatr. Logop. **65**(4), 214–220.

Toscano, J. C., and Toscano, C. M. (**2021**). "Effects of face masks on speech recognition in multi-talker babble noise," PLoS One **16**(2), e0246842.

Truong, T. L., Beck, S. D., and Weber, A. (**2021**). "The impact of face masks on the recall of spoken sentences," J. Acoust. Soc. Am. **149**(1), 142–144.

Westfall, J., Kenny, D. A., and Judd, C. M. (**2014**). "Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli," J. Exp. Psychol. Gen. **143**(5), 2020–2045.

Yi, H., Pingsterhaus, A., and Song, W. (**2021**). "Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic," Front. Psychol. **12**, 682677.