How Face Masks Affect Acoustic and Auditory Perceptual Characteristics of the Singing Voice

*Liran Oren, †Michael Rollins, ‡Ephraim Gutmark, and *Rebecca Howell, *†‡Cincinnati, Ohio

Summary: Wearing a face mask has been accepted as one of the most effective ways for slowing the spread of COVID-19. Yet information regarding the degree to which masks affect acoustics and perception associated with voice performers is scarce. This study examines these effects with common face masks, namely a neck gaiter, disposable surgical mask, and N95 mask, as well as a novel material that could be used as a mask (acoustic foam). A recorded excerpt from the "Star-Spangled Banner" was played through a miniature speaker placed inside the mouth of a masked manikin. Experienced listeners were asked to rate perceptual qualities of these singing stimuli by blindly comparing them with the same recording captured without a mask. Acoustic analysis showed that face masks affected the sound by enhancing or suppressing different frequency bands compared to no mask. Acoustic energy around the singer's formant was reduced when using surgical and N95 masks, which matches observations that these masks are more detrimental to the perceptions of singing voice compared with neck gaiter or acoustic foam. It suggests that singers can benefit from masks designed for minimal impact on auditory perception of the singing voice while maintaining reasonable efficacy of filtering efficiency.

Key Words: Singing voice−Face mask−Acoustics−Perceptual characteristics.

INTRODUCTION

Broadway theatre in New York City is a billion-dollar industry that employs thousands and contributes billions more to the local economy in tourism.[1] On March 17, 2020 the Skagit County, Washington choir informed public health officials of ill members, leading to 53 of 122 testing positive for SARS-CoV-2 virus. Since this report, the act of singing has been identified as a contributor of transmission (emission of aerosols by loudness of vocalization) or super-emitters (those that release more aerosols compared to others).[2]

The mode of transmission for SARS-CoV-2 virus is exposure to infected respiratory droplets or aerosolized particles.[3] Both types of particles can be expelled by infected individuals through coughing, sneezing, speech, singing, or breathing.[4] Droplets are particles that, once emitted by a person, will follow a trajectory that is affected predominantly by gravity and depends on their size and ejection force. Aerosolized particles are generally <5 μm and can remain suspended in the air for longer periods after being expelled.[4, 5] The transmission of SARS-CoV-2 occurs when healthy individuals come into contact with or inhale particles that carry these pathogens.[6]

Until herd immunity can be achieved (likely by vaccination), public health efforts to combat COVID-19 have focused on prevention strategies. The Center for Disease Control has focused on reducing transmission via droplet and aerosolized particles by encouraging social distancing, avoiding crowds, hand washing, and mask wearing. Wearing a facial covering has become a common practice to mitigate transmission.

All face masks, whether home-made or well-fitted (eg, medical grade), can effectively block the larger droplet size particles. However, the filtration efficiency, defined as the percentage of aerosolized particles stopped by the mask from spreading away from the source, varies with the filter material. For example, the filter efficiency of woven or non-woven polypropylene, such as a disposable surgical mask or N95 mask, is about 50% and 95%, respectively, compared with about 10% efficiency for fabric material such cotton.[7−9] Despite these differences, it is well documented that using any type of face covering is critical to reducing the number of hospitalizations and deaths related to COVID-19.[10−12]

The factors affecting the filtration efficiency of masks are well documented, but their impact on speech and especially singing has not been sufficiently quantified. Studies have shown that masks can reduce both intelligibility and loudness of speech[13−15] and that speakers wearing a facial covering tend to make their normal speech louder.[16] Similar information regarding the degree to which face masks affect acoustics and perception associated with professional voice users such as singers is scarce. With live entertainment largely on hold, wearing face masks will likely be recommended once these activities resume in efforts to minimize risks of transmitting COVID-19, especially considering it is known that loud speech increases emission of aerosolized particles.[17, 18] Because vocal performers rely on the perceptual qualities of their voice[19] (eg, loudness, clarity, and ring), there is a need to improve the understanding regarding the acoustic effects of masks and auditory perception of the singing voice.

As a first step, this study examines the effects of common face masks, namely a neck gaiter, disposable surgical mask,

Accepted for publication February 22, 2021.
The authors have no financial or personal interests that could inappropriately influence (bias) the work in this manuscript.
From the *Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati, Cincinnati, Ohio; †Department of Biomedical Engineering, University of Cincinnati, Cincinnati, Ohio; and the ‡Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Cincinnati, Ohio.
Address correspondence and reprint requests to Liran Oren, Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati, 231 Albert Sabin Way Room 6113, MSB PO Box 670528, Cincinnati, OH 45267. E-mail: liran.oren@uc.edu
Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■
0892-1997

https://doi.org/10.1016/j.jvoice.2021.02.028

FIGURE 1. Setup used for audio recordings. Miniature speaker is placed inside the manikin's mouth and the same audio sample is played with five types of masks. (a) Baseline recording with no face mask, (b) neck gaiter, (c) disposable surgical mask, (d) N95 mask, (e) acoustic foam. Single- (shown) and double-layers of acoustic foam were tested.

and N95 mask, as well as a novel material that can be used as a mask (acoustic foam), on acoustics measures and perceptual characteristics of listening to a masked singing voice. A recorded singing voice was used as the acoustic source in order to isolate the direct acoustic effects caused by the mask from the influences of using a mask on the act of singing. This study therefore does not evaluate the singer's perspective, only the listener's.

METHODS

Setup and recording of singing voice stimuli

The audio sample used for the study consisted of the phrase "O'er the land of the free" extracted from a recording of a soprano singing "The Star-Spangled Banner" a cappella. It was played from a miniature speaker (Intsun, Tomtop Electronic) that was placed inside the mouth of an airway simulator manikin (AirSim Combo Bronchi X, TruCorp). The manikin's external features such as head circumference, nose and ears matched an average human adult. The miniature speaker was aligned with the manikin's lips and was connected to a laptop by an audio cable that was threaded through the airway (Figure 1a). This setup intended to mimic a reproducible singing voice emanating from a singer and also allowed for comparable evaluation of several types of masks.

Recordings of the audio sample were captured with the manikin's face covered with one of three commonly used face masks (Figures 1b-d). These include a neck gaiter (single-layer cotton fabric material), a disposable surgical mask, and an N95 mask. Reference data were obtained by recording a sample without the mask. In addition to the three standard masks and the reference, the manikin's face was covered with acoustic foam of a single- or double-layer (7.5 and 15 mm, respectively) thickness (Figure 1e). Overall, the audio recordings were made for six mask configurations.

Audio recordings

The manikin was placed near the center of an anechoic chamber (7.6 by 7.2 m). A 1/2-inch microphone (Type 2671, Brüel & Kjær) was positioned 30 cm directly in front of the mouth. In testing directivity effects of this setup, there was a <3 dB change when the microphone was offset by 30° from the sagittal plane. All audio recordings were captured at 51.2 kHz using a National Instruments data acquisition system (NI 9234).

Singing audio samples were captured for 6.6 seconds with each mask. Recordings with each mask type also included a chirp signal with a tone that increased logarithmically over the frequency range from 20 to 20 kHz over 1.5 seconds. The chirp sound was repeated three times (total of 4.5 seconds) in each recording.

Acoustical data analysis

The acoustic analysis was based on the spectrum for the audio data captured with the chirp signal. The analysis also included extracting 0.5 seconds of the sustained /æ/ and /i/ vowel segments in each singing sample and computing the acoustic energy of the singer's formant[20] by integrating the

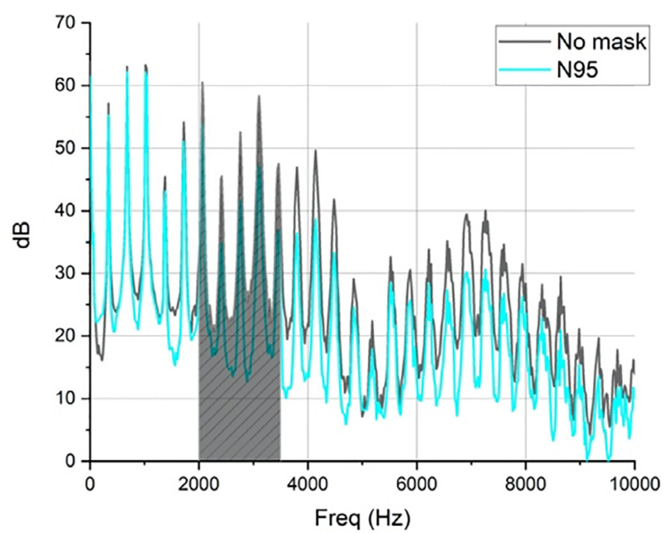

FIGURE 2. Spectra of /æ/ vowel segment extracted from the recorded samples of baseline (no mask) and N95 mask. Acoustic energy in the singer's formant energy is computed by integrating the area under the curve between 2 and 3.5 kHz (shown for the no-mask case as the shaded area).

FIGURE 3. Screen capture of the Alvin program used for perceptual ratings. Baseline recording was always included as one of the pair. Rating was done by having the listener moving the slider toward the side with better "Ring" and "Overall Singing Quality."

spectral energy in the frequency band of 2–3.5 kHz (Figure 2).

Perceptual evaluations

A panel of four listeners (R1-R4) rated the recorded singing samples for each of the six mask configurations. All four have personal backgrounds in vocal performance and speech-language pathology training (range of 0–7 years clinical experience). Each listener completed the evaluation alone seated in a quiet office in front of a laptop that presented the audio samples through its internal loudspeakers.

Perceptual ratings were done using a paired-stimulus presentation method. That is, each listener rated a series of trials, each trial containing a pair of singing stimuli. The presentation of audio recordings and capture of the perceptual response were done using the Alvin program.[21] Listeners could control the pace of presentation and their response time for each trial. Although they were allowed to repeat each pair as often as needed, they had to make their selection before advancing to the next pair. The samples in each pair were separated by 0.4 second pause.

The paired-stimulus had listeners compare a baseline recording (no mask) with another recording in a blinded fashion. Rating the singing stimulus of each mask was repeated 8 times, with the baseline recording presented 4 times as the first of the pair and 4 times as the second sample. The paired samples also included comparison of the baseline with itself; this configuration was done to aid in assessing rater reliability and only repeated 4 times. In summary, each listener rated a total of 44 singing samples (five masks for 8 times + no mask for 4 times). The order of presentation was randomized and the experiment lasted 10–15 minutes for each listener.

The perceptual evaluations included comparison of the "ring" and "overall singing quality" for each pair. The computer displayed a "slider scale," with the slider button positioned at the midpoint (Figure 3). The ends of the slider scale were labeled as Sample 1 and Sample 2, respectively. The following instructions were read to listeners before the experiment: "You will be presented with a series of paired audio segments sung by a woman wearing a face mask. For each presentation, the segments should have differing singing qualities. Your task is to (1) indicate whether Sample 1 or Sample 2 has the better Ring and (2) by how much. Use the mouse to move the slider from the middle of the scale toward the presentation that has the better Ring. Indicate the degree of Ring difference by how close you move the marker to the speech sample; the closer to either end, the larger the difference in Ring. You will then perform a similar judgment of the Overall Singing Quality on the second slider."

The Alvin program recorded the listener's rating for each pair as an integer that ranged from −500 to +500, where "minus" scores indicate better "ring" or "overall singing quality" of Sample 1 while "plus" scores related to Sample 2. For data analysis, the score's absolute value was used as it was always referenced to the "no mask" baseline. A rating of zero would indicate that there was no difference between the two samples. Hence, lower ratings indicated that the singing recording with a mask was perceived to be more like the singing without a mask (i.e., lower score means less impact of the mask).

Sample pairs where listeners judged the baseline sample (no mask) to have worse ring or worse overall singing quality than the singing sample with a mask were removed from the data set (27 of the total 176 samples, 15%). The statistical mean (and its 95% confidence interval) of the ratings of the singing samples with each mask of the 4 listeners was then computed. Rater reliability was assessed using the intraclass correlation coefficient, ICC(2,k).[22]

RESULTS

Acoustic analysis showed that the face masks affected the sound by enhancing or suppressing different frequency bands compared to no mask. For example, the difference between N95 mask and the baseline spectra are shown in Figure 4a. The difference between each mask and baseline was assessed by subtracting the spectrum of the chirp signal for baseline from the spectrum of each mask (Figure 4b). Because the chirp signal includes all audible frequencies (within the 20 Hz to 20 kHz range), the differences between the spectra showed which frequencies were amplified and which were suppressed for each mask. Compared with baseline findings, the neck gaiter had a small effect on the spectrum with some amplification (<2.5 dB) of frequencies below 4 kHz and minimal differences for frequencies >4 kHz. Amplification of frequencies below 2 kHz was observed for all masks. The greatest amplification (~5 dB) was observed with the disposable surgical mask and with two layers of acoustic foam. The surgical mask also showed significant amplification also between 4.5 and 7.0 kHz.

The face masks caused suppression in other frequency bands. Most notably, the N95 mask suppressed frequencies between 2 and 5 kHz and above 6 kHz. The surgical mask showed some suppression between 3-5 kHz. The neck gaiter yielded minimal change in the spectrum. Similar spectral trends for each mask were also found using the long time averaging spectrum of the singing audio samples.

Perceptual testing shows a direct relationship between the type of mask and the perception of the singing samples (Figure 5). Overall, the perception of "ring" seemed to be more affected when using a mask than the perception of 'overall singing quality'. Three out of four listeners showed excellent intra-rater reliability (0.98 for R2, 0.93 for R3, and 0.91 for R4). Intrarater reliability for R1 was 0.71, which is moderate, perhaps because of lack of experience in performing perceptual ratings. The reliability between listeners was moderate at 0.67, which was similar to the

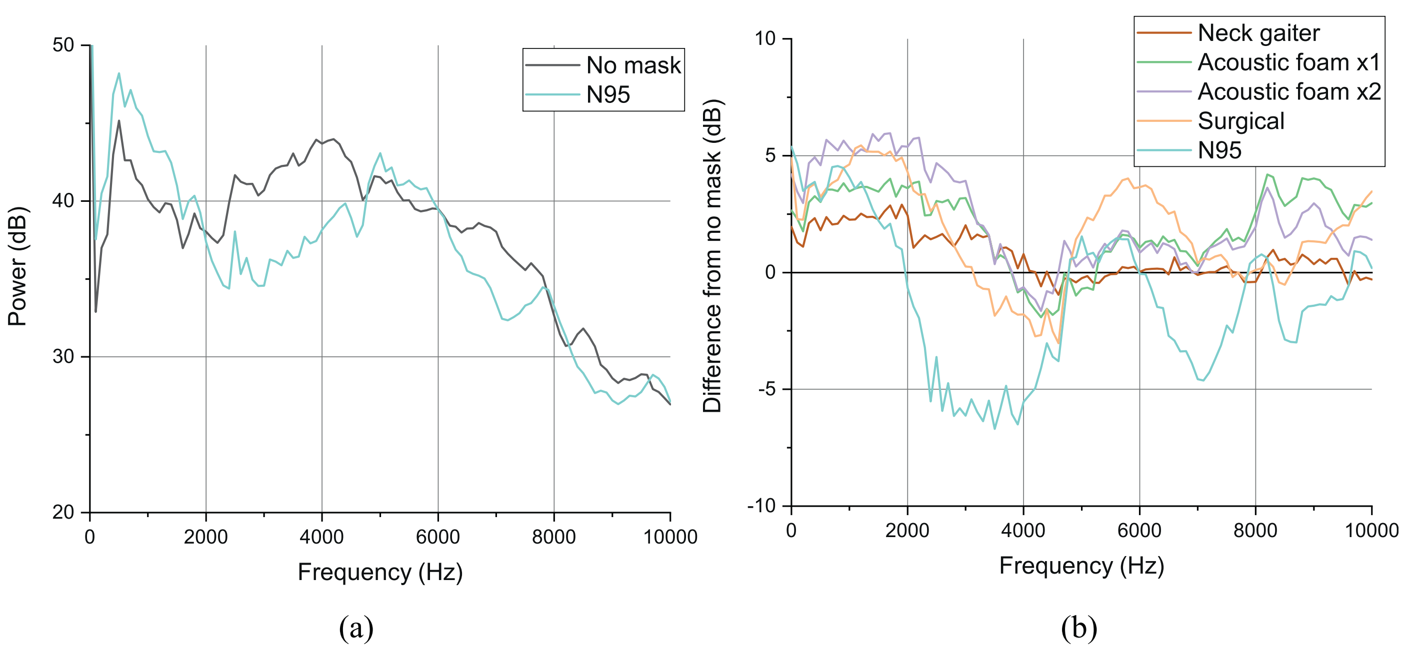

FIGURE 4. Face masks amplify and supressed different frequencies. (a) Spectra of the audio sample using chirp singal; only the baseline and N95 cases are shown. (b) Difference of each mask's spectra relative to baseline (no mask) using a chirp signal.

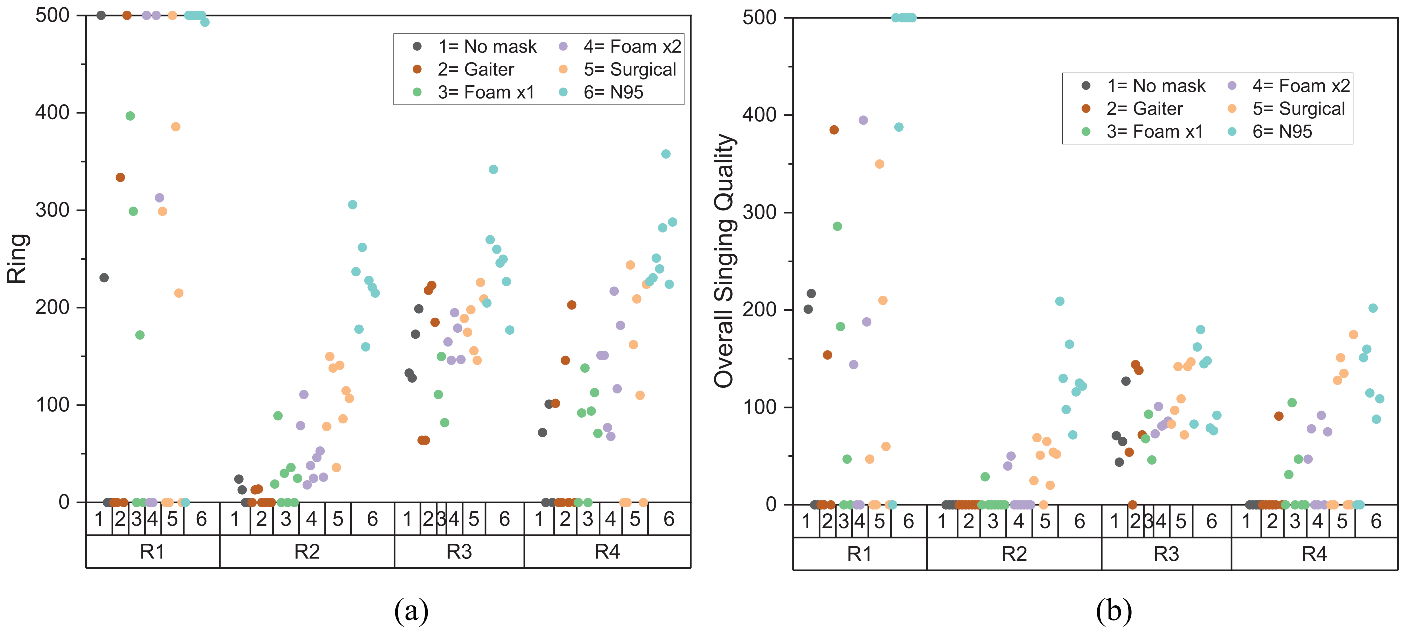

FIGURE 5. Perceptual ratings of singing samples by mask type. Lower scores indicate less perceptual difference relative to baseline (no mask). R1−R4 correspond to each listener (a) Ring. (b) Overall singing quality.

current literature using perceptual ratings of the singing voice.[23-25]

The statistical mean (with its 95% confidence interval) for the listener's ratings of each mask is shown in Figure 6 (left ordinate). The figure also shows the corresponding acoustic energy calculated for the 2−3.5 kHz frequency band (the singer's formant) with each mask (right ordinate). Overall, there was minimal difference between no mask and a neck gaiter while a disposable surgical mask or an N95 mask was more detrimental to ring and overall singing quality perceptions. The N95 mask had the greatest impact on reducing the ring quality. Single layer of acoustic foam had minimum impact on perception compared to both surgical and N95 masks.

DISCUSSION

This study provides useful information regarding the impact of different face masks on the acoustics and perceptual measures of the singing voice for the listener. This will be of particular interest to voice professionals such as singers, actors, and vocal coaches. A mask creates a barrier to sound propagation, acting as an acoustic filter where material and thickness determine its response characteristics. Changes in

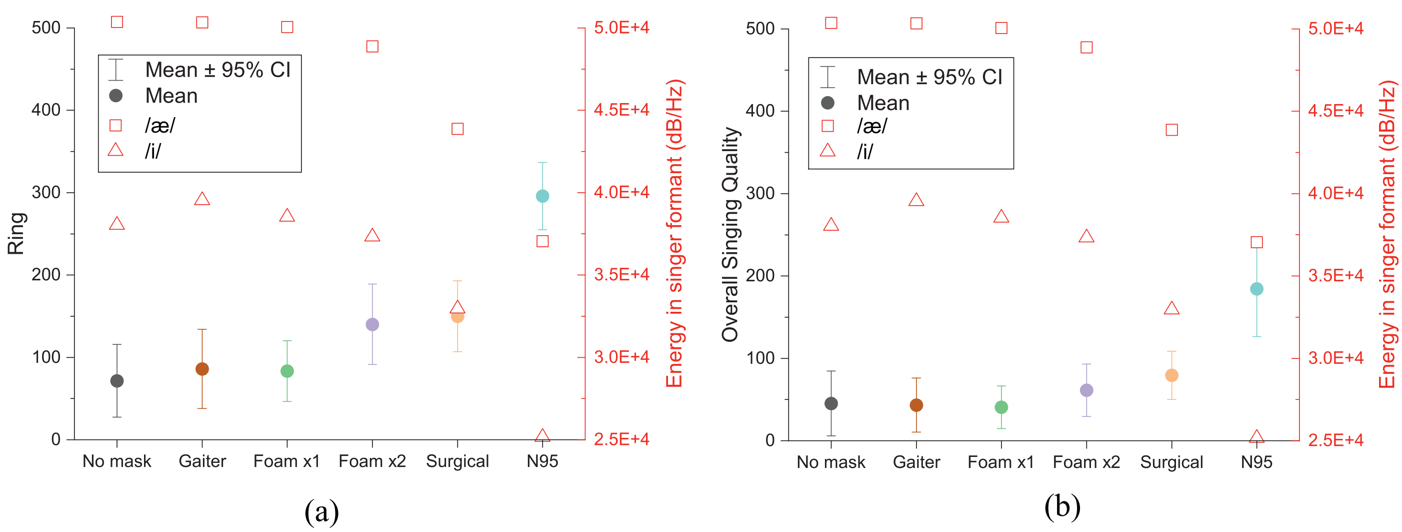

FIGURE 6. Comparing perceptual and acoustic characterisitcs of all mask configurations. The statisical mean (with 95% confidance interval) based on all listeners' ratings for each mask is shown on the left ordinate for the perception of (a) "Ring" and (b) "Overall singing quality." The acoustic energy within the 2−3.5 kHz frequency band (the singer's formant) during sustained /æ/ and /i/ vowel segments in each singing sample is shown on the right ordinate of each plot.

these factors can amplify or attenuate certain frequency bands, thus altering the formants and affecting the perception of radiated sound.

The purpose of the face mask is to maintain a reasonable efficacy in filtering aerosolized particles. An ideal singing mask should be designed to avoid frequency suppression around the signer's formant and be made from material(s) with the highest filter efficiency.

A neck gaiter has minimal effect compared to no face mask. However, a single-layer cotton material common to neck gaiter design offers minimal filtering efficiency, specifically for aerosolized particles. While the N95 mask is highly effective in filtering air particles, this study suggests it is far more disruptive to the auditory perceptual characteristics of the singing voice, possibly because it suppresses the singer's formant.

The magnitude of acoustic energy in the singer's formant frequency band could be an indicator of the perceptual data trends. There was minimal difference in acoustic energy between the no-mask baseline, neck gaiter, and acoustic foams. The bigger differences for the surgical and N95 masks matched the drop in their perceptual ratings. Changes to the acoustic energy in the singer's formant are expected by looking at the chirp signal spectra. The difference from baseline (Figure 4b) shows both surgical and N95 masks suppressed some of the frequencies within the singer's formant, with the latter to a larger extent. This comparison shows that the amplification of frequencies below 2 kHz, which occurred with all masks, did not affect the perception. Although the highest amplification occurred for the double layer foam, the perceptual difference relative to the baseline was not as pronounced as for the N95 mask.

Acoustic foam, which is not a material currently used for face covering, showed little effect on the perception of singing voice. Its single- and double-layer filtration efficiency is about 30% and 50%, respectively.[26] These values are similar to disposable surgical masks but far lower than N95. Using acoustic foam as face covering material would require further characterization such as its breathing resistance[27] (ie, difficulty of breathing through the material). An ideal singing mask should be designed to avoid frequency suppression around the signer's formant and be made from material(s) with the highest filter efficiency.

Limitations

There are several limitations to our study. First, we only considered a limited number of masks and only certain types. For example, neck gaiter with multiple layers, thicker N95 material, and surgical mask with medical grade can absorb the sound differently and therefore will have different perception and/or acoustic characteristics. Second, the sample recording was a soprano voice of a native-English female singer. A singer's formant is known to depend on the singing range (alto/tenor/baritone/bass).[28] Third, the acoustic propagation from the computer's internal speakers can limit listeners' ability to notice perceptual differences. This could also explain why nearly 15% of the ratings were "misjudged" (as if wearing these specific masks is better compared with no mask) and needed to be removed from the analysis. It is reasonable to expect different perceptual results if the listeners were using high-quality headphones. Lastly, the study did not consider perception from the singer using a mask; this would require a different analysis on the aerodynamics and compensatory breathing patterns of a singer wearing a mask. However, it is possible that professional singers could adapt their singing to the mask, allowing their voice to be perceived as the same quality by a listener, regardless of the mask type used. This requires further evaluation.

ACKNOWLEDGMENTS

The authors would like to thank Sergey Grinshpun, PhD for measuring the filter efficiency of the acoustic foam material.

REFERENCES

1. Grosses - Broadway in NYC. The Broadway League. www.broadwayleague.com/research. Accessed December 09, 2020.
2. Hamner L, Dubbel P, Capron I, et al. High SARS-CoV-2 Attack Rate Following Exposure at a Choir Practice — Skagit County, Washington, March 2020. *Morbid Mort Wkl Rep (MMWR)*. 2020;69:606–610. https://doi.org/10.15585/mmwr.mm6919e6.
3. Scientific Brief: SARS-CoV-2 and Potential Airborne Transmission. Coronavirus Disease 2019 (COVID-19). Vol 2020: Centers for Disease Control.
4. Morawska L, Johnson G, Ristovski Z, et al. Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities. *J Aerosol Science*. 2009;40(3):256–269. https://doi.org/10.1016/j.jaerosci.2008.11.002.
5. Duguid JP. The size and the duration of air-carriage of respiratory droplets and droplet-nuclei. *Epidemiol Infect*. 1946;44(6):471–479. https://doi.org/10.1017/S0022172400019288.
6. Van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med*. 2020;382:1564–1567. https://doi.org/10.1056/NEJMc2004973.
7. Pan J, Harb C, Leng W, et al. Inward and outward effectiveness of cloth masks, a surgical mask, and a face shield. *Aerosol Sci Tech*. 2021:1–16. https://doi.org/10.1080/02786826.2021.1890687.
8. Tcharkhtchi A, Abbasnezhad N, Seydani MZ, et al. An overview of filtration efficiency through the masks: mechanisms of the aerosols penetration. *Bioactive Mat*. 2021;6(1):106–122. https://doi.org/10.1016/j.bioactmat.2020.08.002.
9. Asadi S, Cappa CD, Barreda S, et al. Efficacy of masks and face coverings in controlling outward aerosol particle emission from expiratory activities. *Sci Rep*. 2020;10(1):1–13. https://doi.org/10.1038/s41598-020-72798-7. 15665.
10. Eikenberry SE, Mancuso M, Iboi E, et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infect Dis Model*. 2020;5:293–308. https://doi.org/10.1016/j.idm.2020.04.001.
11. Liang M, Gao L, Cheng C, et al. Efficacy of face mask in preventing respiratory virus transmission: a systematic review and meta-analysis. *Travel Med Infect Dis*. 2020;36: 101751. https://doi.org/10.1016/j.tmaid.2020.101751.
12. Gandhi M, Beyrer C, Goosby E. Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer. *J Gen Intern Med*. 2020;35:3063–3066. https://doi.org/10.1007/s11606-020-06067-8.
13. Palmiero AJ, Symons D, Morgan III JW, et al. Speech intelligibility assessment of protective facemasks and air-purifying respirators. *J*

Occup Environ Hyg. 2016;13:960–968. https://doi.org/10.1080/15459624.2016.1200723.
14. Mendel LL, Gardino JA, Atcherson SR. Speech understanding using surgical masks: a problem in health care? *J Am Acad Audiol*. 2008;19(9):686–695. https://doi.org/10.3766/jaaa.19.9.4.
15. Saeidi R, Huhtakallio I, Alku P. Analysis of face mask effect on speaker recognition. *Interspeech*. 2016:1800–1804. 2016. https://www.isca-speech.org/archive/Interspeech_2016/pdfs/0518.PDF. Accessed March 2021.
16. Fecher N. *Effects of Forensically-Relevant Facial Concealment on Acoustic and Perceptual Properties of Consonants*. University of York; 2014. Doctoral dissertation; 2014. https://etheses.whiterose.ac.uk/7397/1/Natalie_Fecher_PhD_2014.pdf. Accessed March 2021.
17. Asadi S, Wexler AS, Cappa CD, et al. Effect of voicing and articulation manner on aerosol particle emission during human speech. *PLoS One*. 2020;15:(1) e0227699. https://doi.org/10.1371/journal.pone.0227699.
18. Asadi S, Wexler AS, Cappa CD, et al. Aerosol emission and superemission during human speech increase with voice loudness. *Sci Rep*. 2019;9:1–10. https://doi.org/10.1038/s41598-019-38808-z. 2348.
19. Ekholm E, Papagiannis GC, Chagnon FP. Relating objective measurements to expert evaluation of voice quality in western classical singing: critical perceptual parameters. *J Voice*. 1998;12(2):182–196. https://doi.org/10.1016/S0892-1997(98)80038-6.
20. Sundberg J, Rossing TD. *The Science of Singing Voice*. Northern Illinois University Press; 1987.
21. Hillenbrand JM, Gayvert RT. Open source software for experiment design and control. *J Speech Lang Hear Res*. 2005;48(1):45–60. https://doi.org/10.1044/1092-4388(2005/005).
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428. https://doi.org/10.1037/0033-2909.86.2.420.
23. Eckert DJ, White DP, Jordan AS, et al. Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. *Am J Respir Crit Care Med*. 2013;188:996–1004. https://doi.org/10.1164/rccm.201303-0448OC.
24. Gupta C, Li H, Wang Y. Perceptual evaluation of singing quality. *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE; 2017:577–586. https://doi.org/10.1109/APSIPA.2017.8282110.
25. Oates JM, Bain B, Davis P, et al. Development of an auditory-perceptual rating instrument for the operatic singing voice. *J Voice*. 2006;20(1):71–81. https://doi.org/10.1016/j.jvoice.2005.01.006.
26. Filtration efficiency of acoustic foam materials. Unpublished results.
27. O'Kelly E, Pirog S, Ward J, et al. Ability of fabric face mask materials to filter ultrafine particles at coughing velocity. *BMJ Open*. 2020;10:(9) e039424. https://doi.org/10.1136/bmjopen-2020-039424.
28. Sundberg J. Level and center frequency of the singer's formant. *J Voice*. 2001;15(2):176–186. https://doi.org/10.1016/S0892-1997(01)00019-4.