Journal of the
American Academy of Audiology

The Effect of Conventional and Transparent Surgical Masks on Speech Understanding in Individuals With and Without Hearing Loss

Journal:	*Journal of the American Academy of Audiology*
Manuscript ID	15-151.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	16-Mar-2016
Complete List of Authors:	Atcherson, Samuel; University of Arkansas at Little Rock, Audiology & Speech Pathology; Mendel, Lisa; University of Memphis, School of Communication Sciences and Disorders Baltimore, Wesley; University of Arkansas at Little Rock, Audiology & Speech Pathology Patro, Chhatakant; University of Memphis, School of Communication Sciences and Disorders Lee, Sungmin; University of Memphis, School of Communication Sciences and Disorders Pousson, Monique; The University of Memphis, School of Communication Sciences and Disorders Spann, M.; University of Arkansas at Little Rock, Audiology & Speech Pathology
Keywords:	background noise, Speech perception, surgical masks

Date of Submission: December 20, 2015

NOTE: This paper has been presented at the American Speech-Language-Hearing Association convention, November 19-22, 2014, Orlando, FL

The Effect of Conventional and Transparent Surgical Masks on Speech Understanding in Individuals With and Without Hearing Loss

Samuel R. Atcherson[1], Lisa Lucks Mendel[2], Wesley J. Baltimore[1], Chhayakant Patro[2], Sungmin Lee[2], Monique Pousson[2], M. Joshua Spann[1]

[1]Department of Audiology and Speech Pathology
University of Arkansas at Little Rock, University of Arkansas for Medical Sciences
[2]School of Communication Sciences and Disorders
The University of Memphis

Author correspondence: Samuel R. Atcherson, Ph.D., Department of Audiology and Speech Pathology, University of Arkansas at Little Rock, 2801 S. University Avenue, Little Rock, AR 72204. Phone: (501) 569-3155; Fax: (501) 569-3157; E-mail: sratcherson@ualr.edu

Acknowledgments: Appreciation is expressed to Jessica Magro for her assistance in earlier phases of this project. Portions of this paper were presented at the American Speech-Language-Hearing Association convention, November 19-22, 2014, Orlando, FL.

Abstract

Background: It is generally well known that speech perception is often improved with integrated audiovisual input whether in quiet or in noise. In many healthcare environments, however, conventional surgical masks block visual access to the mouth and obscure other potential facial cues. In addition, these environments can be noisy. Although these masks may not alter the acoustic properties, the presence of noise in addition to the lack of visual input can have a deleterious effect on speech understanding (Mendel et al, 2008). A transparent (“see through”) surgical mask may help to overcome this issue.

Purpose: To compare the effect of noise and various visual input conditions on speech understanding for listeners with normal hearing and hearing impairment using different surgical masks.

Research Design: Participants were assigned to one of three groups based on hearing sensitivity in this quasi-experimental, cross-sectional study.

Study Sample: A total of 31 adults participated in this study: 1 talker, 10 listeners with normal hearing, 10 listeners with moderate sensorineural hearing loss, and 10 listeners with severe-to-profound hearing loss.

Data Collection and Analysis: Selected lists from the Connected Speech Test (CST) were digitally-recorded with and without surgical masks and then presented to the listeners at 65 dB HL in five conditions against a background of 4-talker babble (+10 dB SNR): without a mask (auditory only), without a mask (auditory and visual), with a transparent mask (auditory only), with a transparent mask (auditory and visual), and with a paper mask (auditory only).

Results: A significant difference was found in the spectral analyses of the speech stimuli with and without the masks; however, no more than ~2 dB (RMS). Listeners with normal hearing performed consistently well across all conditions. Both groups of listeners with hearing impairment benefitted from visual input from the transparent mask. The magnitude of improvement in speech perception in noise was greatest for the severe-to-profound group.

Conclusions: Findings confirm improved speech perception performance in noise in listeners with hearing impairment when visual input is provided using a transparent surgical mask. Most importantly, the use of the transparent mask did not negatively affect speech perception performance in noise.

Key Words: Background noise, speech perception, surgical masks

Abbreviations: ANOVA = analysis of variance; CST = Connected Speech Test; RMS = root mean square; SNR= signal-to-noise ratio

Introduction

Clear communication is an important fundamental component to successful relationships, whether it is used in one's personal or professional environment. In the United States, over 31 million people have a hearing impairment (Kochkin, 2005). When an individual has usable but impaired hearing, speech perception can be difficult. These individuals must make necessary adjustments to try and prevent a communication breakdown to achieve clear communication.

Communication in healthcare environments is absolutely crucial. Communication in such settings usually involves a variety of conversations from scheduling appointments over the phone, instructions to complete forms by receptionists, intake by nurses or healthcare assistants, and care by health professionals. It is important that all forms of communication and conversations involving one's health be clearly transmitted and received. According to Feldman-Steward, Brundage, and Tishelman (2005), it is important to understand patient-provider communication in order to evaluate health goals (or outcomes), for each goal is an expression of one or more of the patient's needs. These goals are the objective of a participant's communication effort. The effort of communicating healthcare needs must be as concise as possible for individuals with hearing loss, whether they are the professionals or the patients, in order to have information relayed successfully. Speech perception is more difficult for those with hearing loss than for individuals with normal hearing. Thus, when individuals who are deaf or hard-of-hearing engage in conversation about healthcare, it is critical that they have as much auditory and visual information available as possible to achieve successful communication.

One of the most prominent techniques deaf people can use to overcome a communication barrier is speechreading. Speechreading, also known as lipreading, is an important strategy that is used to improve speech understanding by utilizing visual cues when observing the speaker's face

(Wieczorek, 2013). Speechreading was described by Jeffers and Barley (1971) as "the art of understanding a speaker's thought by watching the movements of its mouth and his facial expression" (p. 4). Another definition is offered by Campbell (1997) as "the extraction of speech information from the seen action of the lower face, especially the jaws, lips, tongue and teeth, a natural skill in hearing people" (p. 1793). Other parts of the face, such as the cheeks, nose, and eyes, are also considered contributors as well (Thomas and Jordan, 2004). The "McGurk Effect" is a great example, and visual illusion, of how visual cues influence what is heard. McGurk and McDonald (1976) showed that under certain circumstances if one stop consonant was presented auditorily and a second consonant differing in only place was presented at the same time visually, a third consonant would be perceived. For example, /da/ would be perceived due to a simultaneous combination of an auditory /ba/ and a visual /ga/ (McGurk and McDonald, 1976).

While some deaf people rely on a speechreading method for speech perception, it is not always effective. In one of the classic papers on speechreading, Erber (1974) suggested that the lips and inside of the mouth (e.g., tongue and teeth) are important factors to consider, yet they are influenced by both observation angle and illumination conditions. Specifically, observation angles greater than 45°and overhead lighting can reduce speechreading performance. However, speechreading extends beyond just the lips and the inside of the mouth. Preminger, Payen, and Levitt (1988) showed that when the mouth and lower part of the face were masked, overall viseme recognition was poor despite the fact that some subjects were able to distinguish between different consonant visemes. Specifically, visemes /p/ and /f/ could identified with 96% accuracy when paired with /a/ and /i/ even when the mouth was obscured. However, visemes /t/ and /k/ were almost always confused when the mouth was obscured. Taken together, speechreading does not rely just on the lips and parts of the mouth, and that visual capture can be realized from other parts of the face, yet influenced by incidence angle and illumination. In the real world, these visual conditions change rapidly. Within healthcare environments, the use of conventional paper surgical face masks causes a very distinct communication barrier. However, if visual cues are accessible, it is possible that speechreading and viseme recognition can be improved.

Wearing surgical face masks is a necessary procedure that is followed in health-related settings, especially in a physician's or dentist's office. Although paper face masks do not appear to alter the acoustics of speech (Mendel, Gardino, and Atcherson, 2008), these surgical masks cause a direct issue for the deaf and hard of hearing population because they cover the mouth area of the health professionals using them. Because many individuals with hearing loss rely on visual cues from the mouth for speech comprehension, the masks may alter the intelligibility of speech communication. Surgical masks are usually composed of paper, which is not "see-through" material. Such paper masks may also act as a barrier to sound, causing an auditory signal that is subdued. Along with this communication barrier, there is the added difficulty that comes into effect when there is background noise present within the health-care environment. According to Champion and Holt (2000), nearly two-thirds of children with hearing impairment experienced a communication barrier with the dentist because of the mask the dentist was wearing and the music and noise being heard in the office. Thus, both noise and the restricted visual access by the mask would be considered obstacles to communication.

Noise, in general, can be defined as any sound that is unwanted and interferes with normal hearing (Way, Long, Weihing, Ritchie, Jones, Bush, and Shinn, 2013). Noise found within a healthcare setting can be grouped into two categories: equipment-related noise and staff-created noise. Some examples of equipment-related noise include anesthesia alarms, suction devices, and surgical instruments. Some examples of staff-created noise include conversation, door opening and closing, and even background music. According to Way et al. (2013), participants' task performance during a speech perception test presented in an operating room in quiet was superior to that of task performance in a noisy environment. With the combination of background noise and barely audible speech, it is important to realize that these masks may stifle the communication process within healthcare environments, an environment where quality communication is essential. In terms of counteracting the negative auditory effect of noise, Grant, Walden, and Seitz (1998) presented consonant-vowel segments, consonant-vowel-consonant segments, and low-context sentences in noise to listeners with hearing loss in order to study their ability to integrate both auditory and visual information. These materials were presented in an already difficult listening condition with 0 dB SNR. Although there was considerable variability among listeners with most benefitting in the auditory-visual condition, it was clear that some listeners were better at integrating both auditory and visual cues with as much as 26% improvement over the auditory condition alone.

Providing more visual information is a crucial factor when it comes to improving the quality of communication within the healthcare environment. Deaf and hard-of-hearing individuals rely on visual cues from the speaker's face, especially their mouth and nose area. If these two areas are subtracted out of the communication equation, then individuals with hearing loss can no longer rely on speechreading. The use of conventional paper surgical masks obstructs visual cues needed for successful speechreading, and may degrade the auditory signal, making communication more difficult. It has been suggested that to obtain speech understanding of 90% accuracy, the signal must be presented at 10 to 15 dB above the noise source (Way et al, 2013). Thus, with an average background noise level around 65 dB SPL, personnel would have to speak at levels around 80 dB SPL to be understood with 90% accuracy (Way et al, 2013).

This puts extreme stress on staff and patients, normal hearing and hearing impaired, in such a demanding environment. Transparent surgical masks may serve to improve the quality of communication for individuals with normal hearing and hearing loss because the clearness of the mask will allow necessary visual cues from the mouth to be available for speechreading. It is also possible that such transparent masks will improve one's ability to understand speech when background noise is present. It is important to compare the effects of a see-through transparent mask with a conventional paper mask to determine if the transparent mask will benefit both patients and health professionals with hearing loss while making them feel more comfortable interacting in such settings. Better access to the lips and mouth for those who must wear face masks may result in better healthcare outcomes.

The purpose of this study was to compare a conventional paper surgical face mask with a transparent ("see-through") prototype surgical face mask on speech perception performance in listeners with normal hearing and hearing loss (moderate and severe-to-profound sensorineural hearing loss). All participants listened to audio-only recordings made by a male talker speaking sentence passages while wearing either a paper mask, a transparent mask, or no mask at all. In addition, stimuli were presented with audiovisual cues for the transparent mask and no mask conditions. In order to make the testing challenging and more realistic, the test stimuli were presented in the presence of background noise. It was hypothesized that listening and watching the talker wearing a transparent face mask would result in improved speech understanding compared to the paper face mask, not only for deaf and hard-of-hearing individuals, but also for those with normal hearing.

Method

Participants

Thirty adult participants, aged 19 to 74 years (M = 44.4), were assigned equally (10 participants each) to one of three groups based on their hearing thresholds. Participants with normal hearing (thresholds better than 25 dB HL) were assigned to the control group (NH). NH participants (5 males, 5 females) ranged in age from 19 to 64 years (M = 28.5). Participants with moderate-to-moderately-severe sensorineural hearing loss (pure-tone averages between 41 and 70 dB) were assigned to the moderate hearing loss group (MOD). Attempts were made to include participants in the experimental groups who had bilaterally symmetric hearing loss with flat configurations (no more than 10 dB slope per octave) but three participants in the MOD category had asymmetric hearing loss. Hence the thresholds in their better ears were used to meet the selection criteria. These 10 participants (8 males, 2 females) ranged in age from 20 to 74 years (M = 49.6). Participants with severe-to-profound sensorineural hearing loss (thresholds greater than 71 dB HL) were assigned to the other experimental group (SEV) which consisted of 6 males and 4 females ranging in age from 22 to 68 years (M = 48.7). Figure 1 displays the mean air conduction thresholds for the three groups of participants. All of the participants in the experimental groups (MOD and SEV) used their own amplification devices (hearing aids and cochlear implants) set to their prescribed settings. We neither optimized their device parameters nor included their audiologists in the study. Demographic details and amplification devices are summarized in Table 1.

All participants had normal middle ear function bilaterally as evidenced by normal tympanometric results (i.e., normal tympanometric peak pressure, ear canal volume, static admittance, and tympanometric width) using screening normative data from Roup et al (1998). Participants were also native speakers of American English and had no major health issues other than hearing loss. Because the study addressed the issue of audio-visual integration, it was important for the participants to have good visual capabilities. A Snellen chart was used to verify that participants had good visual acuity and/or used corrective lenses during the study. All subjects were able to correctly repeat the letters on a Snellen chart that represented 20/20 vision.

Stimuli and Instrumentation

For the purposes of this study, stimuli from the Connected Speech Test (CST; Cox, Alexander, and Gilmore, 1987; Cox et al, 1988) were used. The original CST uses everyday sentences pooled into 48 passages (24 list pairs) of connected speech. Each passage consists of 2 list pairs where each list contains 10 sentences, with 25 key words each, centered on a familiar topic (e.g., Envelope/Grasshopper, etc.). Four passages were administered for each experimental condition to enhance the reliability of the obtained results.

The CST stimuli were re-recorded to reflect speech produced in the various surgical mask conditions. Given that monitored-live-voice presentation of stimuli is less reliable, this re-recording was necessary in order to maintain consistency across the subjects and experimental conditions. The stimuli were re-recorded using a digital audio recorder (Marantz Model PMD660 portable solid state digital recorder) by an adult American male speaker who had a General American Dialect. The talker’s speech was clearly intelligible without any abnormality in vocal characteristics as judged qualitatively by the researchers. The talker was instructed to speak as naturally as possible without deviating away from the microphone. Because the CST was re-recorded with a new speaker, we expected some departure from the original standardization and validation of the CST. Therefore, any results produced from the new recordings cannot be directly compared to the original CST. All of the stimuli were recorded in a double-walled sound treated booth meeting ANSI Standard S3.1-1999 (ANSI, R2008) maximum permissible ambient noise levels for audiometric test rooms. The stimuli were recorded using a Shure SM93 microphone positioned approximately 10 inches from the speaker who was seated in the sound booth. The microphone was connected to the digital recorder positioned outside of the booth.

A digital video recorder (Canon Vixia HG21A) was placed one meter away from the speaker's face at zero degrees azimuth for the audio-visual conditions with and without masks. His full facial image and upper chest were recorded. A sampling rate of 48 kHz and 32-bit analog-to-digital converter was used to record the stimuli. While the speech stimuli were being recorded, the talker's speech characteristics (e.g., rate of speech, voice quality, intensity, etc.) were monitored constantly by the experimenters who were outside the sound booth and provided feedback to the talker regarding the acceptability of each production. The talker was instructed to maintain a normal conversational rate without any exaggerated articulatory movement. All CST stimuli were scaled and edited using Adobe Audition (Version 3.0) to maintain uniformity in loudness across the lists and experimental conditions. A 1000 Hz calibration tone of 20 seconds duration was created using Adobe Audition for calibration of the stimuli.

Multi-talker (4 talker) babble from the Bamford-Kowal-Bench Speech-in-Noise (BKB-SIN) Test (Bench, Kowal, and Bamford, 1979; Niquette et al, 2003; Etymotic Research, 2005) was used with the CST stimuli. It consisted of male and female talkers speaking random sentences simultaneously, making it difficult for a listener to understand what one particular talker was saying. The multi-talker babble was reproduced from the BKB-SIN using Adobe Audition, which was edited and looped to create a 50-minute sample burned to an audio CD to be presented along with the experimental stimuli. The babble was chosen to represent the background noise experienced by many professionals who work in an operating theatre who have to communicate using a surgical mask.

Procedure

Data were collected at two sites: (1) the Speech Perception Assessment Laboratory at The University of Memphis (UofM) and (2) the Auditory Electrophysiology and (Re)habilitation Laboratory at the University of Arkansas at Little Rock (UALR). Prior to data collection, all participants signed an informed consent approved by the UofM and UALR Institutional Review Boards for participation in this study, and basic ethical considerations were taken for the protection of the research participants. The same detailed study protocol was followed consistently across the two data collection sites. All participants were informed that the aim of the study was to look at the effect of different types of surgical masks on speech perception, and all were compensated for their participation. The initial goal was to have each site recruit half of the participants from each group, which we were able to do for the NH and MOD groups. For the SEV group, UALR had better access and ended up recruting 8 of the 10 participants.

Pure-tone thresholds were measured for all participants at the octave frequencies from 250 to 8000 Hz using a diagnostic two-channel audiometer (GSI-61) with supra-aural earphones (TDH-50) meeting ANSI S3.6-2004 specifications for audiometers (ANSI, 2004) and standard audiometric procedures. Prior to presentation of the stimuli, the 1000 Hz calibration tone was played to adjust the VU meter deflection of the audiometer to ‘0.’ The subjects were then instructed as follows: “*You will hear several lists of topic-related sentences. Some of the sentences are presented so that you can see the talker on the video monitor. Some of the sentences are presented with the video screen blank. After you hear each sentence, please repeat* *it as clearly as you can. If you are unsure, please guess. Be sure to face forward and try to keep your head still.*"

The listeners were seated in the sound-treated room, and the experimental stimuli were played from a laptop (Dell Precision M4700), routed to a monitor inside the sound room, and presented via a loudspeaker (Boston loudspeaker). The multi-talker babble was played from a CD player (Sony compact disc recorder-RCD-W500C/W100) and presented via the same speaker. Both the stimuli and the noise were routed to the audiometer and presented via the loud speaker at +10 dB signal-to-noise ratio (SNR, noise at 55 dB HL and stimuli at 65 dB HL). The participants were seated 1 meter away from the loudspeaker at 0 degrees azimuth and the computer monitor was located slightly to the side. The video on the monitor was clearly visible and placed at about one-half meter distance from the subject.

All groups of research participants listened to four CST passages in each of the following five conditions in a randomized manner:

- Condition 1 - NMA (No Mask Auditory only; Lists 19, 20, 37, 38)
- Condition 2 - NMAV (No Mask Audiovisual; Lists 13,14, 47, 48)
- Condition 3 - TMA (Transparent Mask Auditory only; Lists 9, 10, 11, 12)
- Condition 4 - TMAV (Transparent Mask Audiovisual; Lists 7, 8, 29, 30)
- Condition 5 - PMA (Paper Mask Auditory only; Lists 21, 22, 1, 2)

In the NMA condition, the stimuli were recorded without any mask and presented in the auditory modality only. In NMAV condition, listeners not only heard the stimuli through the speaker but also visualized the talker's face without a mask in the monitor. The talker produced the stimuli while wearing a paper mask in the PMA condition and while wearing a transparent mask in the TMA condition. In both of these mask conditions, the stimuli were presented in the auditory modality only. In the TMAV condition, the stimuli were recorded using the same mask as in TMA condition but presented in both auditory and visual modalities. During the auditory-only conditions (e.g., NMA, TMA and PMA), the listeners only heard the experimental stimuli through the speaker (whilst the screen remained blank). Because the paper mask obscured any visual input, the PMA condition was also an auditory-only condition. In contrast, during the audio-visual conditions (e.g., NMAV, TMAV), the listeners heard the stimuli through the speaker as well as visualized the talker's face in the monitor as he wore the different masks (see Figure 2).

A lapel microphone was attached to the collar of each subject's shirt at an approximate distance of 10 centimeters and the microphone was connected to a digital recorder which was placed outside the sound-treated booth for the experimenter to hear and score the responses. Presentation of the test stimuli was paused to allow the participant time to repeat each item and the experimenters to score their responses. Listeners' responses to the stimuli were scored as correct only if all key words were repeated correctly. Inter-judge scoring reliability of the listeners' responses was calculated on 50% of the data from each group (NH, MOD, SEV) to ensure accuracy in scoring the talk-back responses from the participants. The following formula was used: (agreements / [agreements + disagreements]) X 100%. Inter-judge scoring reliability was found to be 99%.

Results

Spectral Analysis of Stimuli

To perform the spectral analysis comparisons, recorded CST stimuli within each condition were edited, and the silent gaps between the sentences were deleted using Adobe Audition (Adobe Systems Incorporated, version 3.0) (previously described in Mendel et al, 2008). Next, the total root mean square (RMS) values were determined for the experimental conditions. The FFT size was set to 65536 (maximum) and Blackman-Harris filter was used relative to 0 dBFS (full-scale). Table 2 shows the total RMS values for the experimental conditions. A one-way analysis of variance (ANOVA) across conditions showed there was a significant main effect for mask condition when analyzing the RMS values [$F(4, 15) = 6.935$, $p < .0001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean RMS score for the NMA condition ($M = -21.91$) was significantly higher ($p<0.5$) than the PMA ($M = -22.02$), TMA ($M = -23.64$) and TMAV conditions ($M = -24.06$), respectively. However, the NMAV condition ($M = -21.90$) did not differ significantly from the NMA condition ($M = -21.91$). In addition, the RMS scores for TMA and TMAV conditions were not significantly different from each other ($p>0.05$) but were significantly lower ($p<0.05$) than the PMA condition. Taken together, these results indicate that the presence of a mask affected the transfer of speech information by significantly reducing the RMS values of the stimuli. Specifically, when there was no mask present (either in the audio or audio-visual conditions), the RMS values were significantly higher than the conditions with a mask. Further, the transparent mask conveyed the least speech information (lowest RMS values) compared to the other experimental conditions, despite the fact that it provided visual cues.

Speech Perception Results

The main aim of this study was to evaluate the effect of hearing status (NH, MOD, SEV) and mask type (no mask, paper mask, and transparent mask) on speech recognition performance. All percent correct scores on the CST were converted to rationalized arcsine units (RAU; Studebaker, 1985) for statistical analysis to stabilize the error of variance and avoid ceiling and/or floor effects. RAUs were converted back to percent correct for display of the data. All participant CST results are illustrated in Figure 3.

A two-way repeated measures ANOVA revealed a significant main effect for hearing status ($F(2,27) = 19.862$, $p<0.001$, partial $\eta 2 = 0.595$, power = 1.00), a significant main effect for type of mask ($F(4,108) = 22.410$, $p<0.001$, partial $\eta 2 = 0.454$, power = 1.00), and a significant interaction between hearing and type of mask ($F(8,149) = 6.732$, $p<0.001$, partial $\eta 2 = 0.595$, power = 1.00). Post hoc Tukey pairwise comparisons revealed statistically significant differences ($p<0.001$) between the participants with normal hearing and those with severe-to-profound hearing loss for all types of mask, with the subjects who had normal hearing performing significantly better than those with severe-to-profound hearing loss. There was no significant difference between the subjects with normal hearing and those with moderate hearing loss in their performance in the NMA, NMAV, and TMAV conditions suggesting that the addition of visual cues for those with better hearing did not have a great impact on speech perception performance.

Tukey all pairwise comparisons also revealed that those with normal hearing showed no statistically significant differences across the mask conditions. Those with moderate hearing loss showed statistically better performance in the NMAV condition compared to the paper (PMA) and transparent masks audio only (TMA) conditions ($p<0.001$). In addition, those with moderate loss showed better performance for the TMAV condition compared to the PMA or TMA conditions indicating the presence of visual cues through the transparent mask was better than the transparent mask with only auditory cues or the paper mask alone. Lastly, those with severe-to-profound hearing loss performed significantly better in all conditions that provided visual cues

(NMAV and TMAV) confirming their reliance on visual information to compensate for their hearing loss.

Discussion

The purpose of this study was to compare a conventional paper surgical face mask with a transparent ("see-through") prototype surgical face mask on speech perception performance for listeners with normal hearing, moderate, and severe-to-profound hearing loss. Each participant was presented with sentences in an audio-only or audiovisual format in the presence of background noise with an SNR of +10 dB. The purpose of the background noise was to simulate real world listening and achieve at least 90% speech understanding performance in normal hearing listeners.

Listeners with normal hearing and moderate hearing loss performed extremely well compared to the listeners with severe-to-profound hearing loss. Overall, for these two groups, there was little to no impact of the paper or transparent face masks on their speech recognition performance, with or without visual cues, and the presence of noise at +10 dB SNR resulted in little to no decrement of performance. Thus, despite the fact that the paper and transparent masks reduced the overall output level of the stimuli, this had no significant effect on speech recognition performance for the listeners with normal hearing or those with moderate hearing loss. Further, the lack of a significant difference in performance between those with normal hearing and those with moderate hearing loss for the NMA, NMAV, and TMAV conditions suggests that the addition of visual cues for those with better hearing did not have a great impact on speech perception performance. Even though the presence of visual cues did not have a statistically significant effect on scores for these participants, some stated at the end of the test that having a visual component always allowed them to answer more confidently; while sentences presented without visual cues made them more hesitant.

Listeners with severe-to-profound hearing loss had greater difficulty in speech perception, especially in the absence of visual cues. These listeners' overall scores were significantly below those with normal hearing and moderate hearing loss. However, those with severe-to-profound hearing loss showed statistically better performance in the NMAV and TMAV conditions compared to the PMA and TMA conditions suggesting the presence of visual cues through the transparent mask resulted in a significant improvement in scores.

All participants with severe-to-profound hearing loss were able to be tested; some completed the test with ease while others struggled through the whole process. A couple of the severe-profound listeners were experienced cochlear implant users with self-reported good ability to read lips, and mainly communicated orally. Some of these participants were also very comfortable using their cochlear implants during the auditory-only conditions, having enough experience relying on their cochlear implants in the past to receive auditory information without the need for visual cues. The presence of background noise had a negative impact on these participants' overall performance but the noise served mostly as a nuisance for them and did not hinder them from giving a complete answer. The rest of the severe-to-profound participants did not wear any amplification, primarily used sign language to communicate, and had great difficulty reading lips. Contrary to the cochlear implant users, these participants could not hear the background noise and therefore were not affected by it. The mean data show a clear trend of audiovisual cues resulting in better performance compared to conditions without visual cues. This is true when comparing the NMA and NMAV conditions (overall 26% improvement), as well as with the PMA or TMA and TMAV conditions (overall 27-28% improvement).

It should be mentioned that there were some differences among the participants in both hearing loss groups. Some participants in the moderate group wore hearing aids and others did not. As stated earlier, participants in the severe-to-profound group differed in both amplification use, communication mode, and speechreading abilities.

There were several limitations to this study. One limitation that was unexpected involved the condition in which the transparent mask was used. A glare appeared on the "see-through" portion of the mask from a light source. One of our listeners with severe-to-profound hearing loss had trouble speechreading because the glare became a visual distraction. Although this participant is not an experienced speechreader, it should still be taken into consideration that the glare could possibly serve as a hindrance to one's ability to speechread if they are interacting with a healthcare professional wearing a transparent mask. Although efforts to reduce the glare could have been taken, we chose to leave the glare to simulate the real world as much as possible. Another limitation was that background noise had to be continuously played in between the passages in order for the start of the passage to play at the same time as the noise. The speaker started talking immediately as soon as the particular passage started. This caught some of the participants off guard, which made some of them miss part of the first sentence.

During the CST presentation, only the audiovisual conditions allowed the participant to actually see the speaker on the screen. During the auditory-only conditions, the participants faced a blank screen. There is a possibility that individuals with severe-to-profound hearing loss who wear amplification or implants could benefit from visual cues from other areas of the face besides the mouth. This also could help further demonstrate that a paper mask also visually disrupts a person's ability to comprehend speech communication. Another adjustment that could have been made to this study is to have a pause implemented at the start of each passage so the start of each passage did not begin in such an abrupt manner that it caught the participant off guard.

Future research should examine the effect of different talkers to present the stimuli in all conditions. For example, the AzBio test is an audio-based sentence-in-noise test using two female and two male talkers (Spahr, Dorman, Litvak, Van Wie, Gifford, Loizou, Loiselle, Oakes, and Cook, 2012). The study could have more depth if different fundamental frequencies from different kinds of voices could be evaluated to see if they have a similar or different effect on a person's understanding of someone wearing a transparent surgical mask. In the present study, the participants all had to listen to the same voice, a man's voice, which has a lower fundamental frequency than a female. Future studies could also examine the effect of other visual cues (e.g., facial hair, lipstick, etc.).

Another possibility for further research would be to adjust the presentation level of the talker of the CST during each condition. Individuals engaged in everyday conversation do not speak at the same intensity during the whole conversation. Background noise also does not remain constant during conversation. Most individuals adjust the volume of their voice according to intensity fluctuations of background noise. A future study could further explore different types of noise with both speech and noise stimuli presented at different intensities and SNRs.

Summary and Conclusions

The results of this study demonstrated that in conditions in which background noise was present during the CST, participants with severe-to-profound hearing loss benefited from the presence of visual cues as evidenced by better performance in audiovisual conditions than any other condition. This finding is not entirely unexpected; however, it has never before been demonstrated empirically using a surgical face mask with a “see-through” window. For individuals with normal hearing or moderate hearing loss, the results showed consistently high scores regardless of mask condition. These results suggest that a transparent face mask, such as the one used here, does not decrease the acoustic integrity of the speech signal and offers speechreading advantages for listeners with severe-to-profound hearing losses over auditory only conditions. It is anticipated that listeners with normal hearing may show benefit with the transparent mask in more challenging listening environments where more noise exists, or with speakers who do not have a clear General American dialect.

References

American National Standards Institute (ANSI). (2004). *Specifications for Audiometers* (ANSI S3.6-2004). New York: American National Standards Institute.

American National Standards Institute (ANSI). (2008) *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms*. (ANSI S3.1-1999). New York: American National Standards Institute.

Amos N, Humes L. (2007). Contribution of high frequencies to speech recognition in quiet and noise in listeners with varying degrees of high-frequency sensorineural hearing loss. *J Speech Lang Hear Res 50*(4):819-834.

Bench J, Kowal A, Bamford J. (1979). The BKB (Bamford-Kowal-Bench) Sentence Lists for partially-hearing children. *Br J Audiol* 13*:*108-112.

Campbell R, Zihi J, Massaro D, Munhall K, Cohen M. (1997). Lipreading in the akinetopsic patient, L.M. *Brain* 120:1793-1803.

Champion, JJ, Holt RR. (2000). Dental care for children and young people who have a hearing impairment. *Br Dental J* 189(3):155-159.

Cox RM, Alexander GC, Gilmore C. (1987). Development of the connected speech test (CST). *Ear Hear 8 (Suppl)*:119S-126S.

Cox RM, Alexander GC, Gilmore C, Pusakulich KM. (1988). Use of the connected speech test (CST) with hearing-impaired listeners. *Ear Hear 9:*198-207.

Erber NP. (1974). Effects of angle, distance, and illumination on visual reception of speech by profoundly deaf children. *J Speech Hear Res 17*:99-112.

Etymotic Research. (2005). *Bamford-Kowal-Bench Speech-in-Noise Test* (Version 1.03) [Audio CD]. Elk Grove Village, IL: Author.

Feldman-Stewart DD, Brundage MD, Tishelman CC. (2005). A conceptual framework for patient–professional communication: an application to the cancer context. *Psycho-Oncol 14*(10):801-809.

Grant K, Walden B, Seitz P. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *J Acoust Soc Am 103*(5):2677-2690.

Jeffers J, Barley M. (1971). *Lipreading*. Springfield, IL: Charles C. Thomas.

Kochkin S. (2005). MarkeTrak VII: hearing loss population tops 31 million. *Hear Rev* 12:16–29.

McGurk H, MacDonald J. (1976). Hearing lips and seeing voices. *Nature 264*:746-748.

Mendel LL, Gardino JA, Atcherson SR. (2008). Speech understanding using surgical masks: a problem in health care? *J Am Acad Audiol 19(9)*:686-695.

Niquette P, Arcaroli J, Revit L, Parkinson A, Staller S, Skinner M, Killion M. (2003, March). *Development of the BKB-SIN Test*. Paper presented at the annual meeting of the American Auditory Society, Scottsdale, AZ.

Preminger J, Lin H, Payen M, Levitt H. (1998). Selective visual masking in speechreading. *J Speech Lang Hear Res 41(3*):564-575.

Roup CM, Wiley TW, Sfady SH, Stoppenbach DT. (1998) Tympanometric screening norms for adults. *Am J Audiol 7*:55–60.

Spahr AJ, Dorman MF, Litvak LM, Van Wie S, Gifford RH, Loizou PC, Loiselle LM, Oakes T, Cook S. (2012). Development and validation of the AzBio sentence lists. *Ear Hear 33*(1):112-117.

Thomas, S.M. and Jordan, T.R. (2004). Contributions of oral and extra-oral facial motion to visual and audiovisual speech perception. *J Exp Psychol Hum Percept Perform* 30:873-888.

Way J, Long A, Weihing J, Ritchie R, Jones R, Bush M, and Shinn J. (2013). Effect of noise on auditory processing in the operating room. *J Am College Surgeons* 216:933-938.

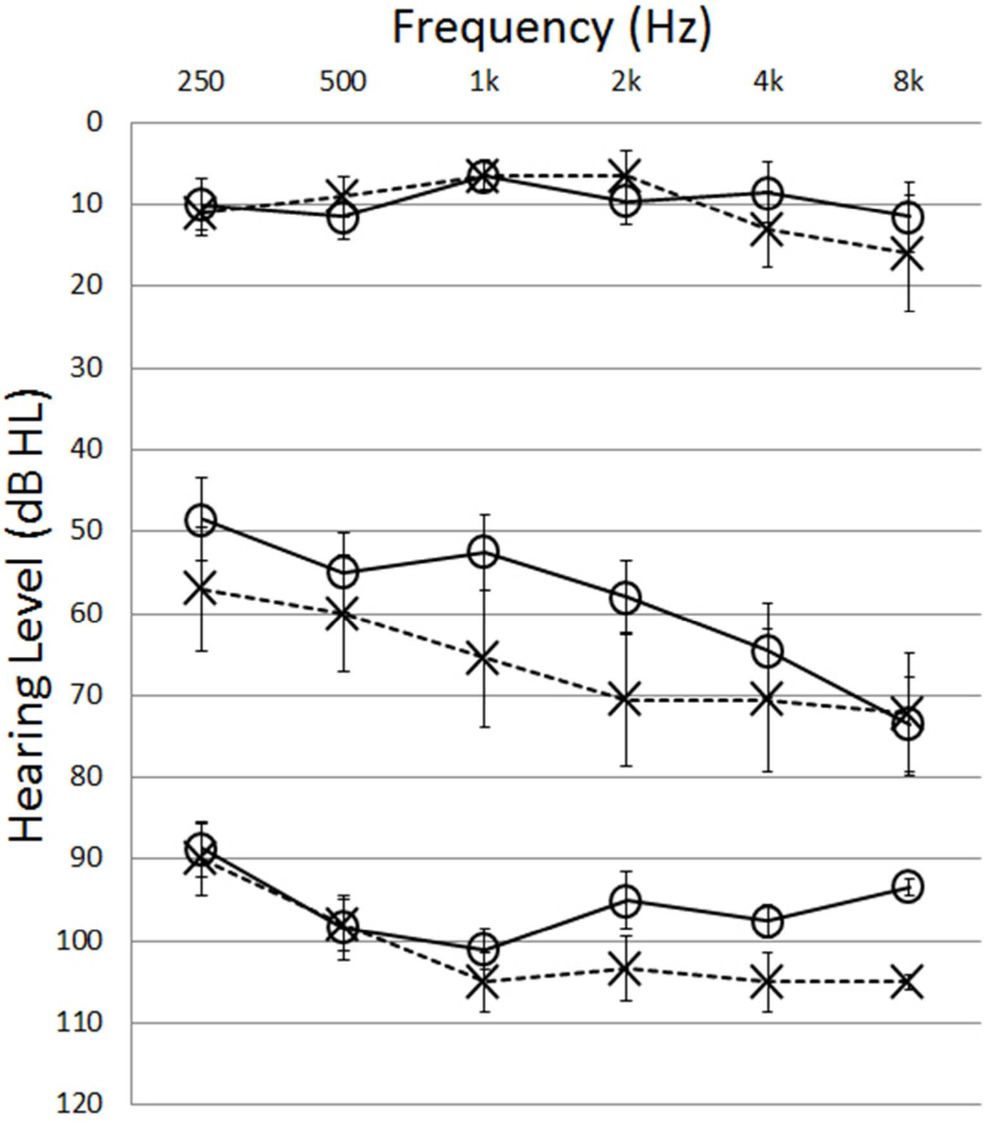
Wieczorek SS. (2013). *Lipreading Difficulty during Audiovisual Integration*. Independent Studies and Capstones. Paper 657. Program in Audiology and Communication Sciences, Washington University School of Medicine. Retrieved from http://digitalcommons.wustl.edu/pacs_capstones/657

Figure Legends

Figure 1. Mean air-conduction thresholds for the normal-hearing and hearing-impaired groups with error bars reflecting +/- 1 standard deviation and sample size. O = Right ear; X = Left ear.

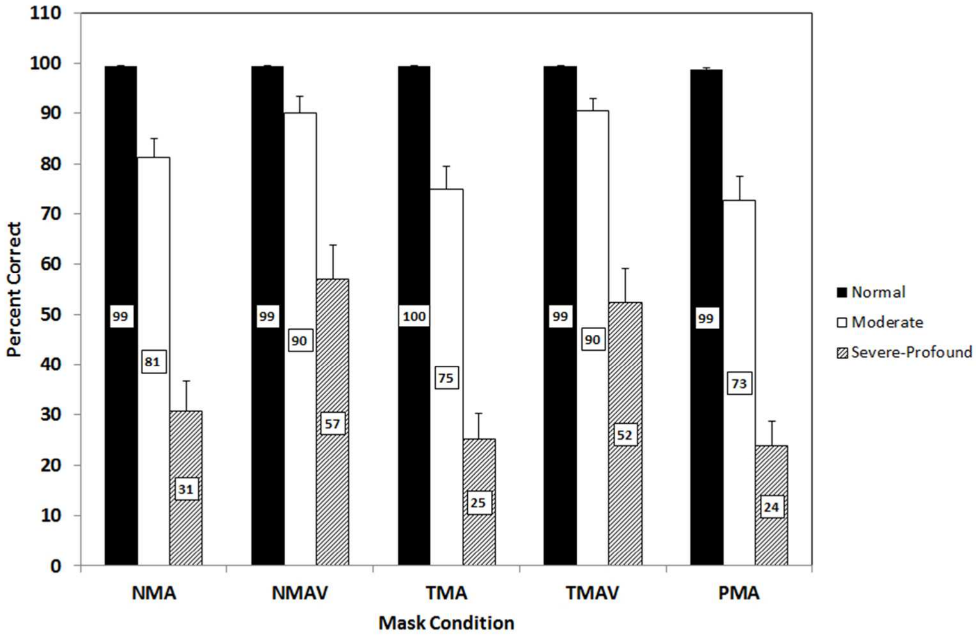
Figure 2. Examples of various experimental conditions on the monitor: No Mask Auditory Visual (NMA; left); Transparent Mask Auditory (TMAV; middle); No Mask Auditory, Transparent Mask Auditory, and Paper Mask Auditory (NMA, TMA, PMA; right). (Note: Actual video was in color and did not have black box covering eyes.)

Figure 3. Mean percent correct performance for all three groups of participants with error bars reflecting +/- 1 standard deviation and sample size (NH = Normal Hearing; MOD = Moderate; SEV = Severe-to-Profound) on the CST stimuli in the five experimental conditions (NMA = No Mask Auditory; NMAV – No Mask Auditory Visual; TMA= Transparent Mask Auditory; TMAV = Transparent Mask Auditory Visual; PMA = Paper Mask Auditory).



Mean air-conduction thresholds for the normal-hearing and hearing-impaired groups with error bars reflecting +/- 1 standard deviation and sample size. O = Right ear; X = Left ear.
45x51mm (300 x 300 DPI)

Examples of various experimental conditions on the monitor: No Mask Auditory Visual (NMA; left); Transparent Mask Auditory (TMAV; middle); No Mask Auditory, Transparent Mask Auditory, and Paper Mask Auditory (NMA, TMA, PMA; right). (Note: Actual video was in color and did not have black box covering eyes.)
98x39mm (300 x 300 DPI)



Mean percent correct performance for all three groups of participants with error bars reflecting +/- 1 standard deviation and sample size (NH = Normal Hearing; MOD = Moderate; SEV = Severe-to-Profound) on the CST stimuli in the five experimental conditions (NMA = No Mask Auditory; NMAV – No Mask Auditory Visual; TMA= Transparent Mask Auditory; TMAV = Transparent Mask Auditory Visual; PMA = Paper Mask Auditory).

76x49mm (300 x 300 DPI)

TABLE 1. Participant Characteristics.

Subject	Category	Age	Gender	Type of Amplification	Monaural/ Binaural/Bimodal (M/B/BM)
1	NH	25	F	NA	NA
2	NH	47	F	NA	NA
3	NH	24	M	NA	NA
4	NH	26	M	NA	NA
5	NH	44	M	NA	NA
6	NH	21	M	NA	NA
7	NH	19	F	NA	NA
8	NH	21	F	NA	NA
9	NH	58	M	NA	NA
10	NH	64	F	NA	NA
11	MOD	54	F	HA	B
12	MOD	74	M	HA	B
13	MOD	72	M	HA	B
14	MOD	64	M	HA	B
15	MOD	74	M	HA	B
16	MOD	25	M	HA	M
17	MOD	20	F	HA	M
18	MOD	26	M	NA	NA
19	MOD	34	M	NA	NA
20	MOD	53	M	HA	M
21	SEV	62	F	HA	B
22	SEV	65	M	HA+CI	BM
23	SEV	68	M	CI	M
24	SEV	38	F	HA	B
25	SEV	47	F	HA+CI	BM
26	SEV	22	M	ABI	M
27	SEV	54	M	CI	M
28	SEV	68	M	NA	NA
29	SEV	26	F	NA	NA
30	SEV	37	M	HA + CI	BM

NH: Normal Hearing; MOD: Moderate to Moderately severe; SEV: Severe-to-Profound; ABI: Auditory Brainstem Implant; HA: Hearing Aid; CI: Cochlear Implant; HA+CI: Hearing aid and Cochlear Implant; NA: Not applicable.

Table 2. Total RMS values in dB for each experimental condition. (NMA = No Mask Auditory; NMAV – No Mask Auditory Visual; TMA = Transparent Mask Auditory; TMAV = Transparent Mask Auditory Visual; PMA = Paper Mask Auditory)

Experimental Conditions	**RMS values (in dB)**
NMA	-21.91
NMAV	-21.91
TMA	-23.64
TMAV	-24.06
PMA	-22.02