Signal quality affects audiovisual speech integration in cochlear implant users and normal-hearing listeners

Hannah Shatzer, The Ohio State University

In everyday life, the primary mode of speech communication with others is face-to-face. The availability of visual speech cues from the movements of a talker's mouth, teeth, tongue, and jaw necessitates the inclusion of that visual information with the acoustic speech signal via the automatic process of audiovisual (AV) integration. Visual speech information can hold substantial influence over the perception of auditory speech via AV integration, as evidenced by a robust multisensory illusion known as the McGurk effect: When an auditory speech signal (e.g. /aba/) is paired with incongruent visual lip movements (e.g. /aga/), the resulting percept is a blending of features from both signals (/ada/; McGurk & MacDonald, 1976). Even when auditory speech is clearly intelligible but syntactically or semantically complex, visual speech can have a surprisingly strong influence on the overall percept (Reisberg et al., 1987; Kim & Davis, 2003). Visual speech has an even stronger influence on the resulting AV percept when the acoustic signal is degraded or otherwise less intelligible, as in a noisy environment. Visual speech can provide an intelligibility benefit of 11 dB or more when an auditory speech signal is masked by noise (Sumby & Pollack, 1954; Grant & Seitz, 2000; MacLeod & Summerfield, 1987).

Cochlear implant (CI) users are part of a hearing-impaired population that benefits even more from the inclusion of visual speech cues for comprehension. CI users are constantly exposed to degraded auditory speech via the electric signaling of their implants, and tend to show greater AV speech benefit from additional visual cues than individuals with normal hearing

I am enormously grateful to my collaborators on this project: Dr. Mark Pitt from the Department of Psychology at Ohio State, Drs. Aaron Moberly and Kara Vasil from the Department of Otolaryngology at Ohio State, and Drs. Antoine Shahin and Jess Kerlin from the Center for Mind and Brain at the University of California, Davis. I also thank the OSU research assistants who aided in data collection on this project: Lauren Boyce, Gabrielle Grose, Stephany Ngombu, Kristina Pickett, and Natalie Schoenfeld. This work was funded by NIH NIDCD grant #R01 DC013543 to Antoine Shahin.

(Desai et al., 2008). Postlingually deafened CI users have enhanced visual lipreading ability before implantation, when they must rely upon visual cues during pre-CI deafness to understand and verbally communicate with others (Bernstein, Auer, & Tucker, 2001). This visual speech ability is maintained post-implantation, and they continue to demonstrate increased reliance on visual cues in AV speech conditions despite having regained some auditory ability (Rouger et al., 2007). Several studies have indicated that CI users have stronger speechreading skills and AV benefit on average relative to normal-hearing (NH) individuals, though there is a high degree of performance variability among CI users (Bernstein et al., 2001; Desai et al., 2008; Kaiser et al., 2003; Champoux et al., 2009).

While the benefits of providing informative visual speech to improve auditory speech perception have been shown in several studies, the relative influence of less informative visual cues in both NH and CI subjects is less studied—for example, visual speech that has been blurred, inverted, or otherwise manipulated. A study by Huyse, Berhommier, and Leybaert (2013) demonstrated that both normal-hearing and cochlear-implanted children show an increased reliance on auditory speech when visual speech cues have been degraded, with better lipreaders in both groups still showing the greatest weighting of the unreliable visual cues. The McGurk effect is also impacted by varying the information quality available from the auditory and visual signals; when clear visual cues are available; instances of the McGurk effect are much higher than when the visual speech is blurred but the auditory speech is clear (Hazan et al., 2010). These studies point to the idea that auditory and visual cues have flexible, highly variable weighting during the integration process, and that perceivers are able to weight more reliable cues more highly in the resulting AV percept.

The neural regions underlying AV speech integration for both informative and degraded stimuli are known to include auditory and visual cortices, in addition to higher-level processing regions such as angular gyrus, supramarginal gyrus, planum temporale, inferior frontal gyrus, and ventral premotor cortex, with posterior superior temporal sulcus/gyrus believed to be the site of multisensory integration (Calvert & Lewis, 2004; Okada et al., 2010; Nath & Beauchamp, 2011). However, while structures involved in AV integration are relatively well-confirmed, the functional relationships between these areas are less certain. The primary auditory and visual cortices do appear to have direct functional connectivity that allows for information communication and early influence of one modality on processing of the other modality (Besle et al., 2008; Powers et al., 2012). For example, the inclusion of visual speech speeds neural processing of auditory speech relative to just the acoustic signal alone, suggesting an enhanced efficiency of acoustic processing when additional, informative visual speech cues are provided (van Wassenhove et al., 2005). Thus, it is reasonable to expect that the reliability and quality of one modality's signal may impact the processing of another signal—if auditory speech is degraded but clear visual cues are available, the perceiver will weight the visual modality more strongly in the AV integration process and show stronger influence of visual cues on the final AV percept.

The current study sought to further examine the dynamics of neural processing related to AV signal quality and how that processing may differ between NH perceivers and CI users, thus shedding light on potential neural AV integration differences between these two populations. Participants in an electroencephalography (EEG) experiment completed a McGurk task in which the quality of both signals was manipulated by blurring the visual mouth movements and varying the signal-to-noise ratio (SNR) of the auditory speech signal. Participants were predicted to

weight the most reliable signal most strongly in making perceptual judgments about what they heard, with CI users relying more heavily upon the visual speech signal than NH perceivers when those cues were clearly available. I also predicted that when the visual signal was clear, both groups would demonstrate suppression of early auditory cortex activity via amplitude reduction of the P1-N1-P2 auditory evoked potentials (AEPs), and CI users would show greater suppression relative to NH perceivers that reflected their increased reliance on the visual signal for perception. Clear visual cues would also have a stronger suppression effect on AEPs when the auditory speech was presented at a lower SNR and therefore less informative.

Method

Participants

Twenty adult CI users (9 females) and 14 age-matched NH controls (6 females) were recruited for the study (average age = 66 ± 9.6 years). All participants were right-handed, native speakers of American English, were screened for cognitive impairment, and self-reported normal or corrected-to-normal vision. All CI users had at least 2 years of experience using their implant, and scored at least 60% on the CID phoneme identification test in order to be included in the study. Table 1 describes the characteristics of the CI users in the current sample. NH controls were all tested for normal audiometric performance prior to participation. All participants were monetarily compensated for taking part in the study. Informed consent was obtained in accordance with procedures approved by the Ohio State University Institutional Review Board.

Subject	Sex	Age (yrs)	Duration of Deafness (yrs)	Age at First Implant (yrs)	Years of CI Experience	Side of Implant	CID Phonemes Correct (%)
1	F	67	32	62	5	Right	96.4
2	M	67	52	60	6	Left	98.5
3	M	70	55	62	8	Bilateral	88.6
4	M	73	20	71	3	Right	--
5	F	64	10	61	3	Bilateral	80.3
6	M	60	4	57	3	Bilateral	92.9
7	F	36	21	31	5	Left	85.1
8	M	80	76	74	6	Left	69.5
9	F	66	24	54	12	Bilateral	100
10	M	60	46	55	5	Left	78.7
11	M	60	7	55	5	Right	89.4
12	F	51	51	35	16	Bilateral	99.3
13	M	55	37	50	5	Bilateral	94.3
14	F	66	60	63	3	Right	67.3
15	F	70	58	56	14	Bilateral	87.9
16	M	54	51	36	18	Bilateral	75.9
17	M	82	14	80	2	Left	65.2
18	M	77	17	73	4	Left	69.5
19	F	85	55	45	40	Right	89.4
20	F	56	32	44	12	Left	76.6

Table 1. Demographic characteristics of CI participants.

Stimuli

Three vowel-consonant-vowel stimuli were selected: /aba/, /aga/, and /awa/. These particular consonants were chosen because various combinations of auditory and visual utterances may result in a third, fused percept through the McGurk effect. A female native English speaker from central Ohio was recorded producing multiple utterances of these stimuli through simultaneous use of a Panasonic GDVx100A video camera with a frame rate of 59.94 f/s and an Atlas Soundolier microphone with a sampling rate of 48,000 Hz. Productions of /afa/ were also recorded as a carrier stimulus.

One token of each stimulus was selected for use in the experiment. Video stimuli for /aba/, /aga/, and /awa/ were edited using Adobe Premiere Pro CS6 (Adobe Systems, San Jose,

CA). Each stimulus was edited to a 3-second video, starting and ending with the lips closed for approximately half a second. Videos were downsampled to 29.97 f/s and each frame was exported as a 1280x720 pixel .jpg image to improve the smoothness of playback during the experiment. The blurred version of the videos was created using a custom Matlab script to apply a 50x50 pixel (sd = 10 pixels) Gaussian filter to each frame image. Prior pretesting has shown this level of blurring to preserve overall gross motor movements of the mouth and jaw while obscuring finer articulatory detail. Auditory stimuli, including the carrier stimulus /afa/, were similarly edited to 3 seconds, with silence at the beginning and end of the video that corresponded to the duration of mouth closure before and after the utterance. Auditory stimuli were normalized to 70 dB SPL and exported as .wav files.

A custom Matlab program was used to create the noise manipulation for the auditory stimuli. The /b/, /g/, and /w/ consonants were spliced out of their original utterances and placed into the /afa/ template in place of /f/--this procedure was completed in order to ensure that the vowel articulations would be identical across the three stimuli. 250 ms of white noise was added during the consonant period in the spliced stimuli, including a 20 ms ramping period at the beginning and end of the noise. The signal-to-noise ratio (SNR) was manipulated directly during the auditory thresholding task prior to the AV task.

Procedure

Threshold task. Each participant completed an auditory-only task in Matlab to determine their individual SNR threshold for the auditory stimuli prior to the AV task. Unilateral CI users who also used a hearing aid removed that device and only used the CI for this study, though that ear was not plugged. Participants were presented with a randomized order of 60 trials, 20 trials per auditory stimulus. Each stimulus was initially presented with a SNR of -6 dB for the

consonant in noise, and participants were asked to identify the consonant by pressing one of three response keys on the keyboard labeled with /b/, /g/, and /w/. The SNR for each consonant was manipulated via a 4-down, 1-up staircase: For every 4 trials in which the given consonant was correctly identified, the SNR was decreased by 1 dB by manipulating the volume of the consonant. For every trial in which the consonant was incorrectly identified, the SNR was increased by 1. If the presented SNR exceeded 0 dB, the consonant volume was held at 0 dB and the noise volume was decreased from 0 dB in 2 dB increments for each step during the task. The final SNR threshold obtained by the end of the task was used to create the High SNR and Low SNR stimuli for each participant for use in the AV task: The high SNR stimuli were presented at 6 dB above the SNR threshold, and the low SNR stimuli were presented at 6 dB below the SNR threshold. This procedure equalized auditory performance across both CI and NH participants by ensuring that the High SNR and Low SNR conditions correspond to individual auditory perceptual ability.

AV task. After completion of the threshold task, participants engaged in a short, behavioral version of the AV task that was to be used in the EEG session. Participants sat approximately 1 meter away from an 18” Dell monitor, with the video stimuli presented at their original size on the screen and auditory stimuli presented from a speaker located behind and the monitor using Presentation software (Neurobehavioral Systems, Albany, CA). Stimuli were presented in a fully-crossed 2x2x3x3 (video quality x audio quality x video consonant x audio consonant) design in blocks of 72 trials, with each block containing two trials of each stimulus combination in a randomized order (Figure 1). During this behavioral pretest, participants completed a total of 3 blocks (216 trials).

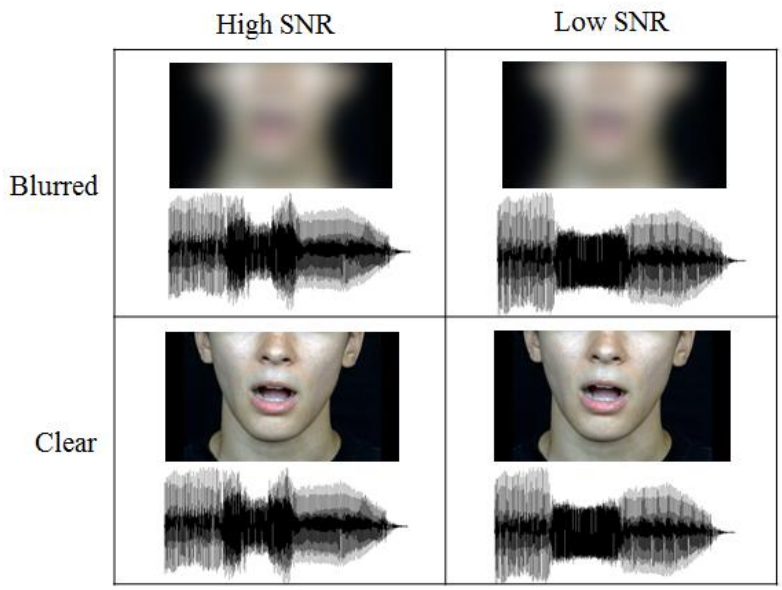

Figure 1. Possible combinations of stimuli by signal quality.

During the task, participants were presented with a clear or blurred video of /aba/, /aga/, or /awa/ simultaneously with the high SNR or low SNR audio of one of the three stimuli. As in the threshold task, they were asked to identify with a key press on the keyboard whether they heard /b/, /g/, or /w/, with two important changes: First, they were instructed to pay attention to the video, but to respond specifically with what they heard. Second, they also had the option of pressing a fourth response key, labeled 'non,' if the participant perceived a fusion, the McGurk effect, or a consonant other than the three available response options. If a participant reported hearing a consonant other than /b/, /g/, or /w/, they verbally reported what sound they perceived to an experimenter before pressing the 'non' key. Responses were not timed, and participants received a short break between each block. This shortened, behavior-only version of the AV task was completed primarily to collect the verbal information on 'non' percepts. Prior to the EEG session, all participants also completed the CUNY test, which involves identify words and sentences presented in A-only, V-only, and AV conditions to estimate auditory speech and visual speechreading abilities.

EEG task and equipment. Participants completed an extended version of the same audiovisual identification task described above during EEG recording. Participants again sat approximately a meter away from a high-resolution monitor, with audio presented from a speaker placed behind and slightly below the monitor. Participants were given the same instructions as in the behavioral version of the task, but with one exception: they did not verbally report when they perceived a consonant other than /b/, /g/, or /w/, instead just pressing the 'non' key with no specific report of what they heard before continuing to the next trial. Eight blocks of 72 trials were presented during the EEG session, with short breaks between each block for an opportunity to rest. Each trial lasted about 6 seconds, with stimulus markers at the onset of the video, onset of mouth movements, onset of sound, offset of sound, and offset of mouth movements. Participants were given a visual prompt to respond after the stimulus finished playing. They were instructed to respond using their left hand, and the order of response keys was varied for each participant via a Latin square to avoid response artifacts associated with the use of certain fingers to respond.

The EEG was recorded at a sampling rate of 1000 Hz using 64 electrodes from a 128-channel cap (BrainVision actiCHamp system, 10-20 Ag-AgCl electrode placement). As the location of the external magnet and processer of the cochlear implant on the scalp interfered with the placement of some electrodes for CI users, those electrodes were removed prior to the gelling process. The number of electrodes removed varied from one implant to the next, though typically anywhere from 1-5 electrodes had to be removed for each implant. NH participants were recorded from all 64 electrodes.

Data Analysis

Behavior. In the behavioral data from the EEG session, each participant's response per trial for incongruent AV stimuli (e.g. visual /aba/ and auditory /aga/) was coded for accuracy as either a visually-matched or auditory-matched response. If the response given matched the consonant presented in the visual stimulus, that response was coded as a 1 for visually-matched and a 0 for auditory-matched. Conversely, if the response matched the consonant presented in the auditory stimulus, it was coded as a 0 for visually-matched and a 1 for auditory-matched. If the response given on an incongruent trial did not match either of the presented consonants but was not a 'non' response, it was coded as a 0 for both an auditory-matched and visually-matched response. If a 'non' response was given on a trial, it was coded as an equally auditory- and visually-matched response (0.5 for each), under the assumption that the percept was a fusion or McGurk that combined features of the auditory and visually presented consonants. This assumption was supported by the verbally reported percepts in the behavioral AV task, which suggested that nearly all 'non' responses were of this fused nature. Congruent trials (e.g. visual and auditory /aba/) were coded as a 1 for an auditory-matched response, as participants were responding with what they heard. The coded responses were then averaged to determine the overall proportion of responses that were auditory-matched versus visually-matched for each of the four visual quality x auditory quality conditions.

EEG. EEG analyses were conducted with EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014), along with the use of in-house Matlab code. For each participant, continuous EEG files were decimated to 250 Hz, then epoched from -0.5 to 4 seconds around the start of each trial. Independent components analysis (ICA) included all channels that were used during recording for each individual subject (note that the number of channels varied depending on the number of electrodes removed for each CI participant),

resulting in the identical number of ICA components. ICA components that represented eyeblinks and CI artifact noise were removed, then channels that had been removed during the study for CI subjects were interpolated. Data files were filtered with a zero-phase Butterworth filter between 0.1 and 30 Hz, then average referenced. Finally, files were epoched from -100 to 600 ms around the onset of the auditory stimulus for each trial, re-baselined to the 100 ms pre-stimulus interval, and averaged across trials within each condition to produce ERP waveforms for the four conditions of interest: clear x high SNR, clear x low SNR, blurred x high SNR, and blurred x low SNR.

Results

Behavior. Figure 2 depicts the average proportion of auditory-matched and visually-matched responses per condition for NH participants, while Figure 3 depicts the same results for CI participants. As predicted, both groups appear to rely upon the most informative signal available within each condition to decide what they heard. The more intelligible, high SNR auditory speech yielded auditory-dominant responses regardless of the quality of the visual signal for NH participants (Figure 2). Uncertainty in responses was higher for the low SNR audio, but NH participants still gave auditory dominant responses when the visual mouth movements were blurred. The only case in which more visually dominant responses were present was in the clear x low SNR condition, when the visual signal was more informative than the acoustic signal. For CI participants (Figure 3), variability in performance was much higher than in NH users. On average, they responded with auditory-dominant responses when the video was blurred, and visually-dominant responses when the video was clear, suggesting that CI users are

making greater use of available visual cues to determine what they are hearing relative to NH participants.

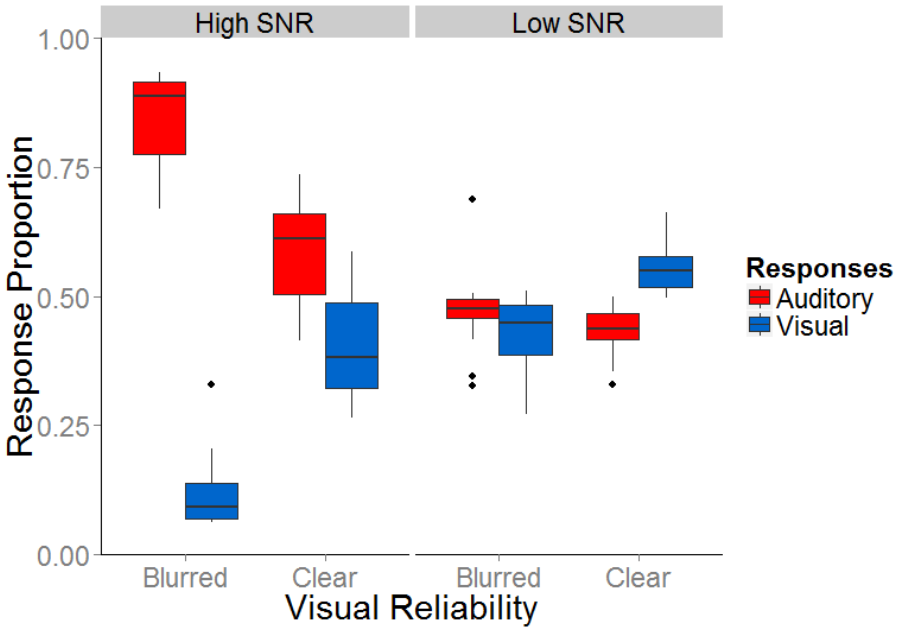

Figure 2. Boxplots depicting proportions of auditory-matched and visually-matched responses given by participants with normal hearing (N = 14). The left panel compares response proportions between blurred and clear videos with high SNR audio; the right panel compares responses with low SNR audio.

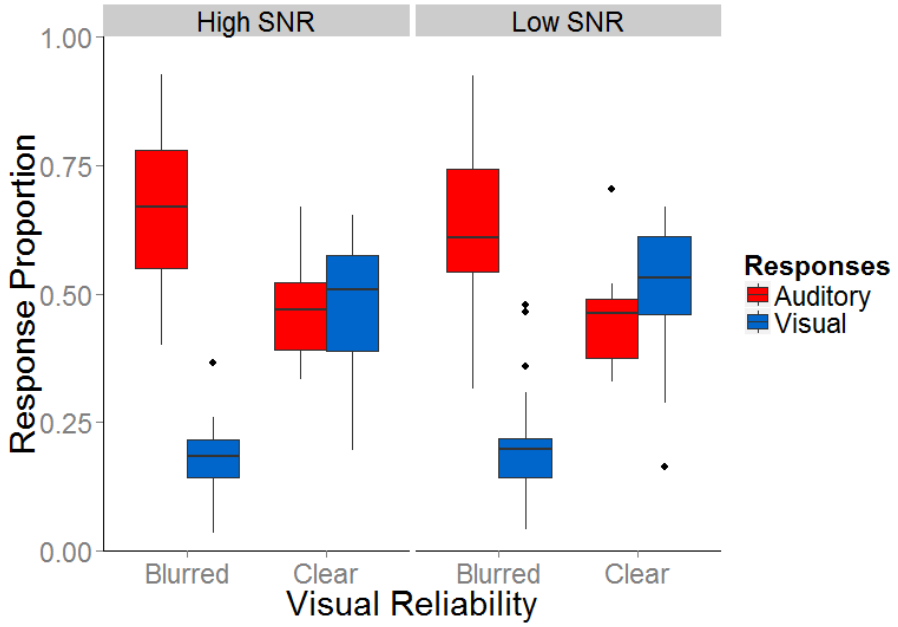

Figure 3. Boxplots depicting proportions of auditory-matched and visually-matched responses given by participants with CIs (N = 20).

Mixed effects regression was used to fit models to the data with group, auditory quality, and visual quality and their corresponding interactions as predictors of the proportion of auditory responses with a random effect for participant. The best-fitting model indicated a significant main effect of group ($t = 5.24$, $p < 0.0001$), auditory quality ($t = -3.619$, $p = 0.0003$), and visual quality ($t = -19.861$, $p < 0.0001$). Additionally, there was a significant interaction between group and auditory quality ($t = -20.979$, $p < 0.0001$) and between group and visual quality ($t = -3.27$, $p = 0.001$). The interaction between group and auditory quality reflects the lack of significant changes in response proportions between high SNR and low SNR auditory conditions for CI users, whereas NH participants gave significantly lower proportions of auditory-matched responses in the low SNR conditions when acoustic information was less reliable. The interaction between group and visual quality reflects the decreased proportion of auditory-matched responses from CI participants relative to NH participants when the video was clear. Generally, results are consistent with the initial predictions for behavioral responses from both NH and CI participants.

EEG. Auditory evoked potential (AEP) waveforms were generated by time-locking EEG activity to the onset of the acoustic stimulus, then averaging over trials for the auditory and visual quality conditions. The waveforms in Figures 4 and 5 are averaged over three fronto-centro-parietal electrode sites (Cz, FCz, and CPz), as these sites are known to robustly reflect auditory-evoked activity (Luck, 2014).

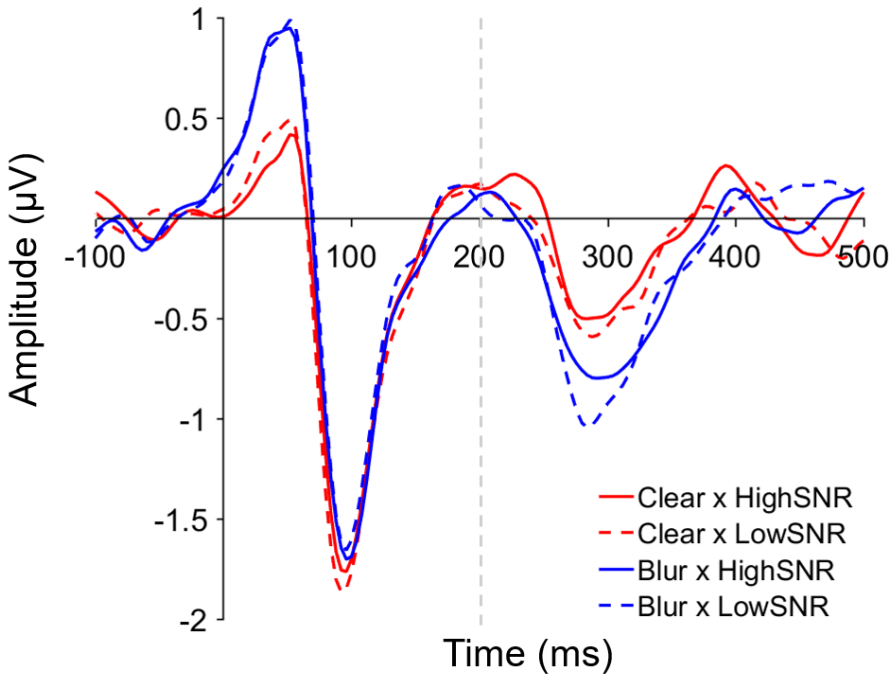

Figure 4. AEP elicited by the onset of the acoustic stimulus for NH participants (N = 14). Onset of noise occurs at ~ 200 ms post-acoustic onset (dashed line).

NH participants (Figure 4) indicate clear suppression of the P1 peak for clear stimuli relative to blurred stimuli. There do not appear to be differences between conditions at N1 or P2, but the subsequent negative peak does again show suppression for the clear visual conditions, with the largest amplitude for the blurred and low SNR condition. Note that this difference in amplitudes corresponding to the auditory SNR occurs after the onset of noise, at which point any differences in activity related to the SNR would be expected to arise. CI participants (Figure 5) also show a suppression of clear conditions relative to blurred conditions for P1, with a lack of amplitude differences between the conditions for N1 and P2. In contrast to NH participants, however, CI users do not show any indication of amplitude differences related to auditory quality after the onset of the noise ~200 ms post-acoustic onset. This lack of effects from the auditory SNR is consistent with their behavioral data, which showed no significant differences in the proportion of auditory-matched or visually-matched responses between the high SNR and low SNR conditions. One may also note that the amplitudes of the P1-N1-P2 complex appear to be

suppressed in CI users relative to NH participants; however, a direct statistical comparison between these groups is unwise due to potential confounds affecting the AEP amplitudes. For example, CI users were recorded from fewer electrodes than NH participants, and had more ICA components removed for artifacts, which may contribute to their smaller amplitudes.

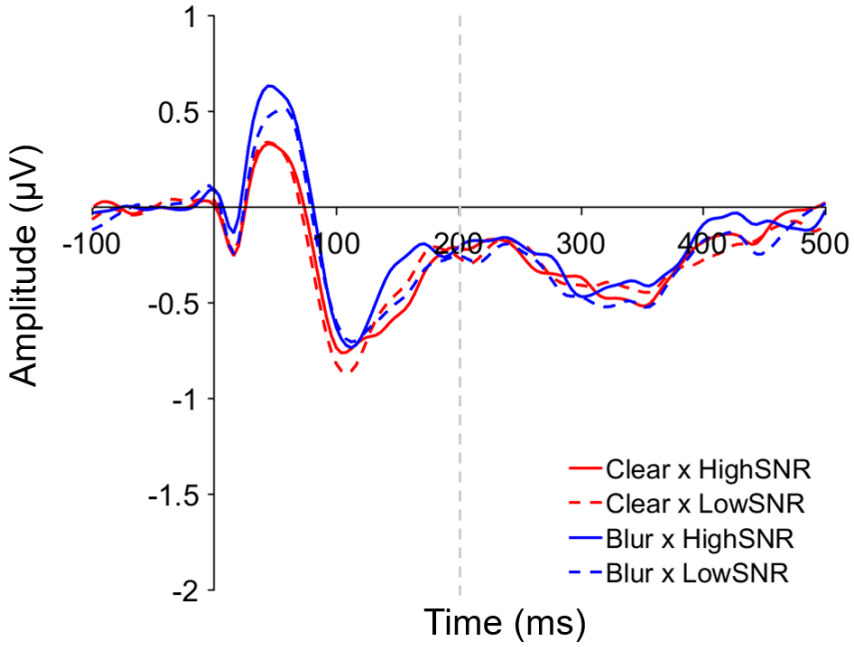

Figure 5. AEP elicited by the onset of the acoustic stimulus for CI participants (N = 20). Onset of noise occurs at ~ 200 ms post-acoustic onset (dashed line).

Statistical comparisons of P1-N1-P2 amplitudes, therefore, were made only within groups. Mixed effects regression models were fitted to the P1 amplitude measurements for the NH participants with the auditory and visual conditions as predictors and a random effect of participant. The best-fitting model revealed a significant main effect of visual quality ($t = 4.86$, $p < 0.0001$), confirming that P1 was significantly suppressed for the clear visual conditions relative to blurred. There were no statistically significant differences between conditions at N1 or P2. While the averaged AEP waveforms do trend towards a P1 suppression for clear conditions in CI users similar to that of NH participants, this difference was not statistically significant for CI users (they also did not show significant differences between conditions at N1 or P2). This lack

of significance is likely due to the high degree of variability observed in the AEP data from CI participants—in Figure 6, it is clear that the P1 amplitudes for CI users do follow the same pattern as NH participants, but there is a greater degree of overlap and variability in amplitudes for CIs.

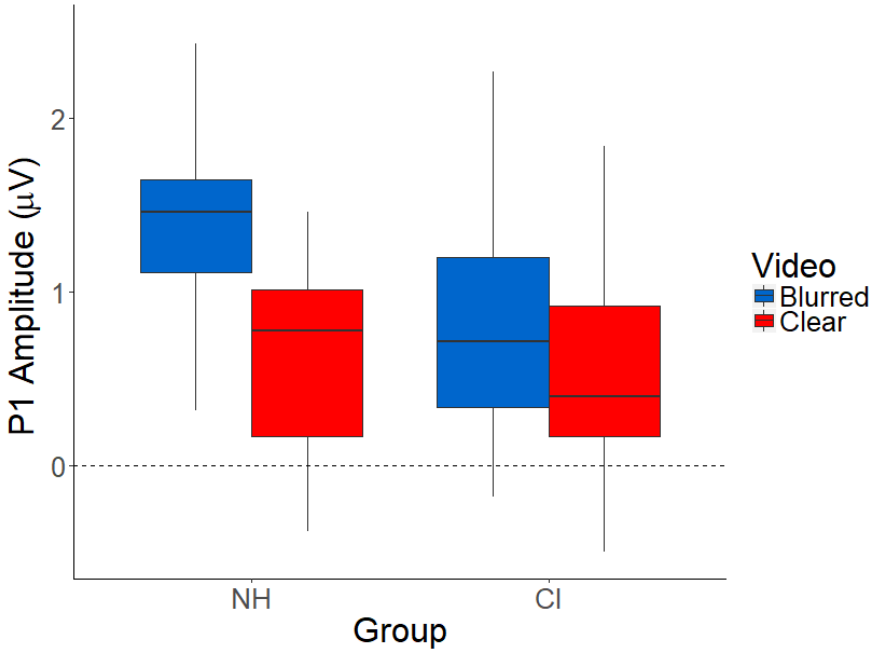

Figure 6. Boxplots of P1 amplitudes for participants with normal hearing (N = 14; see Figure 4) and with cochlear implants (N = 20; see Figure 5).

In summary, AEP results suggest that the inclusion of clear visual information does indeed reduce auditory-evoked activity in NH perceivers (with a trend toward significance in CI users), thus aligning with my initial hypothesis that reliable visual information suppresses early auditory cortex, leading to a greater visual influence on the overall speech percept than if visual information is blurred and uninformative. A later interaction between auditory and visual conditions after the onset of noise also suggests that the influence of clear visual information on early auditory cortex is stronger relative to highly uninformative speech (blurred video and low SNR audio), an effect that was not observed in CI users. Additionally, the smaller P1-N1-P2 amplitudes overall for CI users may suggest an overall greater influence of visual information on

early auditory cortex activity than in NH perceivers; however, this cannot be fully confirmed with a true statistical comparison between groups due to the differences in recording and artifact removal between the two participant groups.

Discussion

The behavioral and neural results from the present study suggest that the quality and reliability of auditory and visual speech signals impact the integrated percept of both modalities. In both CI and NH participants, clear visual speech decreased the proportion of auditory-matched responses and increased visually-matched responses, demonstrating a substantial influence of visual cues on the resulting AV percept. This effect was strongest when visual cues were clear and the auditory speech was presented at a lower SNR for NH participants, though effects of clear visual speech were equal in both high and low SNR conditions for CI participants. Both groups also indicated a suppressive effect of clear visual speech on the P1 amplitude evoked by auditory speech, though this effect was not significant in CI users—not only were the P1 amplitudes smaller to begin with in the CI sample, but there was also greater variability in amplitudes that may partly be due to the variability in the number of electrodes used for recording and number of ICA components removed for each of the CI participants. A later interaction between auditory quality and visual quality was observed after the noise onset in NH participants, but not in CI participants—the clear visual conditions were significantly suppressed relative to the least informative condition with blurred speech and low SNR audio, suggesting that for NH perceivers, the inclusion of clear visual information may also affect acoustic signal processing at a later stage beyond initial engagement of the primary auditory cortex.

These results generally support my predictions about the weighting of informative vs. uninformative auditory and visual speech signals in AV integration: Clear, informative visual speech has a substantial influence on the earliest stages of acoustic signal processing, as P1 is the AEP component associated with initial activation of early auditory cortex after the onset of an acoustic stimulus (Plourde, 2006). This amplitude reduction indicates reduced sensitivity to low-level features of the auditory stimulus, such as acoustic onsets, when clear visual speech is available—therefore contributing to the enhanced weighting of visual cues. Additionally, the neural results for NH participants in this study demonstrate an interaction of auditory and visual signal quality on auditory evoked activity: Peak suppression is greatest when the visual speech is clear and informative, relative to when both the auditory and visual signals are degraded and uninformative. This again suggests that the inclusion of clear visual speech reduces the acoustic processing load and allows for a stronger weighting of cues from the visual modality in the overall AV speech percept. However, effects of clear visual speech were concentrated at P1, and no significant differences were found at N1-P2. Previous research has shown reduction specifically of N1-P2 amplitudes via visual influence on auditory processing mechanisms (Besle et al., 2004; Stekelenburg & Vroomen, 2007), so the absence of those effects here is surprising. As the N1-P2 complex can index both low-level and high-level features of auditory speech, perhaps blurring the visual speech cues mainly affects low-level processing, not higher-level linguistic processing of the acoustic signal (Carpenter & Shahin, 2013).

I did not find a statistically significant suppression of peak amplitudes in the low SNR relative to high SNR audio when the video was clear, as I had originally predicted, so it appears that the benefit of clear speech cues is not immediately contingent on the quality of the acoustic signal from the neural results. However, I did find an increased proportion of visually-matched

responses relative to auditory-matched responses when comparing the clear x low SNR and clear x high SNR conditions in NH participants, so it is curious that this particular difference is not reflected in the AEP data. Since the effects of manipulating the auditory SNR occur later in the ERP recordings, it is possible that analysis of oscillatory activity may reveal these differences in NH perceivers—for example, since enhanced beta is an indication of stronger AV integration, we may see stronger beta activity for the clear x low SNR than in the clear x high SNR that reflects the stronger influence of visual information (Keil et al., 2011).

The current results also potentially implicate greater suppression of AEPs in CI users relative to NH participants with the inclusion of visual speech information. However, due to the necessary differences in EEG recording and ICA removal in NH and CI participants, a true comparison of P1-N1-P2 amplitudes across groups to confirm whether CI users show greater suppression of early auditory cortex with clear visual speech than NH perceivers is challenging. Future work that investigates the underlying neural mechanisms of integrating degraded AV signals could use functional near-infrared spectroscopy (fNIRS) as a neuroimaging method that would allow for easier statistical comparison between groups—as fNIRS uses near-infrared light to measure hemodynamic responses in the brain, it is both safe for CI users and would not require reduced measurement density to accommodate the implant (Chen et al., 2017).

Some interesting observations arise from the CI results that do not align with my original predictions: First, CI users reported higher proportions of auditory-matched responses in the blurred x low SNR condition relative to NH participants, suggesting that they were better able to perform the task and respond with what they heard than NH participants. This finding may be due to their enhanced experience with listening to degraded speech, which NH listeners do not encounter as often in daily life. Indeed, experienced CI users tend to perform better than

untrained NH perceivers on perceptual tasks that involve a degraded auditory speech signal (Başquent, 2012). NH participants in the EEG task also reported an unusually high number of 'non' responses (which were coded as equally visually-matched and auditory-matched responses) relative to CI users, which was not the case during their behavioral testing when they were required to verbally report what they heard during when they pressed the 'non' key. As they did not have this requirement during EEG recording, it is possible that NH participants, who are unaccustomed to listening to degraded speech regularly, were pressing the 'non' key as an indication of uncertainty or difficulty with the task rather than a true fusion or McGurk percept.

Additionally, there are no significant differences between the high SNR and low SNR auditory conditions for CI users as I had anticipated. It is possible that some of the CI users in this study were simply not sensitive to the 12 dB difference between high SNR and low SNR conditions—the perception of speech in noise is one of the biggest challenges for CI users due to the lack of fine spectral detail in the electric signal transmitted from the implant (Müller et al., 2002; Friesen et al., 2001). While the SNR threshold task was successful in obtaining overall accuracy of at least 60% in most of the CI users in this study, the SNR threshold was sometimes exceptionally high relative to the NH controls—the range of thresholds obtained for CI users was anywhere between -2 to upwards of 40 dB, at which point a 12 dB difference in the SNR is barely perceptible for even a NH participant. Potentially, a more dichotomous presentation of no noise vs. noise superimposed on the consonant could have elicited differences in CI users' behavioral and neural performance that reflected the auditory signal quality. This study also did not control for implant model, processing strategy, or unilateral/bilateral implantation and the presence of residual hearing, all factors that could potentially impact the acoustic processing of

speech and contribute to the lack of significant differences in performance based on auditory quality.

In conclusion, the current study implies a flexible weighting of auditory and visual speech cues to take advantage of the most informative signal, thus impacting the integrative process and resulting AV percept in significant ways. This weighting is evident for the quality of visual speech in both NH perceivers and CI users and suggests that the inclusion of clear visual cues can reduce auditory processing load in both populations, leading to strong visual influence during AV integration. The role of auditory signal quality for CI users is less clear, though results suggest that NH perceivers benefit most from clear visual speech relative to uninformative cues from both modalities. Future work will need to more closely examine these neural mechanisms of AV integration, particularly in CI users, to determine how they are impacted by changes in auditory signal quality.

References

Başkent, D. (2012). Effect of Speech Degradation on Top-Down Repair: Phonemic Restoration with Simulations of Cochlear Implants and Combined Electric-Acoustic Stimulation. *Journal of the Association for Research in Otolaryngology, 13(5),* 683-692.

Bernstein, L. E., Auer, E. T., Jr & Tucker, P. E. (2001). Enhanced speechreading in deaf adults: Can short-term training/practice close the gap for hearing adults? *Journal of Speech, Language, and Hearing Research, 44(1),* 5–18.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *Journal of Neuroscience, 28,* 14301-14310.

Besle, J., Fort, A., Delpuech, C., & Giard, M.H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20,* 2225-2234.

Calvert, G.A. & Lewis, J.W. (2004). Hemodynamic studies of audiovisual interactions. In G.A. Calvert, C. Spence, & B.E. Stein (Eds.), *The Handbook of Multisensory Processing,* 483-502.

Carpenter, A.L. & Shahin, A.J. (2013). Development of the N1-P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience Letters, 544,* 56-61.

Champoux, F., Lepore, F., Gagné, J., & Théoret, H. (2009). Visual stimuli can impair auditory processing in cochlear implant users. *Neuropsychologia, 47,* 17-22.

Chen, L., Stropahl, M., Schönwiesner, M., & Debener, S. (2017). Enhanced visual adaptation in cochlear implant users revealed by concurrent EEG-fNIRS. *NeuroImage, 146,* 600-608.

Desai, S., Stickney, G. & Zeng, F. G. (2008). Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. *The Journal of the Acoustical Society of America, 123(1),* 428–440.

Friesen, L.M., Shannon, R.V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110,* 1150.

Grant, K.W. & Seitz, P.F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 103,* 2677-2690.

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication, 52(11-12),* 996-1009.

Huyse, A., Berthommier, F., & Leybaert, J. (2013). Degradation of Labial Information Modifies Audiovisual Speech Perception in Cochlear-Implanted Children. *Ear and Hearing, 34(1),* 110-121.

Kaiser, A. R., Kirk, K. I., Lachs, L. & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language and Hearing Research, 46(2),* 390–404.

Keil, J., Muller, N., Ihssen, N., & Weisz, N. (2011). On the variability of the McGurk effect: Audiovisual integration depends on prestimulus brain states. *Cerebral Cortex, 22,* 221-231.

Kim, J. & Davis, C. (2003). Hearing foreign voices: Does knowing what is said affect visual-masked-speech detection? *Perception, 32(1),* 111–120.

Luck, S.J. (2014). *An Introduction to the Event-Related Potential Technique,* 2nd Edition. MIT Press.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21,* 131-141.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264(5588),* 746–748.

Müller, J., Schon, F., & Helms, J. (2002). Speech Understanding in Quiet and Noise in Bilateral Users of the MED-EL COMBI 40/40+ Cochlear Implant System. *Ear and Hearing, 23(3),* 198-206.

Nath, A.R. & Beauchamp, M.S. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *Journal of Neuroscience, 31,* 1704-1714.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T., & Hickok, G. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex, 20,* 2486-2495.

Plourde, G. (2006). Auditory evoked potentials. *Best Practice & Research: Clinical Anesthesiology, 20(1),* 129-139.

Powers, A.R., Hevey, A.R., & Wallace, M.T. (2012). Neural correlates of multisensory perceptual learning. *Journal of Neuroscience, 32,* 6263-6274.

Reisberg, D., McLean, J. & Goldfield, A. (1987). Easy to hear but hard to understand. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading*, 97–114. Hillsdale, NJ: Erlbaum.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the USA, 104,* 7295-7300.

Stekelenburg, J.J. & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19,* 1964-1973.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 12–15.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the USA, 102,* 1181-1186.