TITLE: Visual distraction costs: the cochlear-implant advantage?

INTRO:

Everyday face-to-face communication occurs in a variety of uncontrolled environments where speech from the target talker is sometimes difficult to understand. The difficulties in understanding the auditory speech signal can be attributed to factors inherent to the listener, such as having hearing loss or listening in a non-native language, the speaker, such as voice quality or accent, and/or the environment, such as background noise or reverberation. Visual information can benefit listeners when it complements the auditory information, but real world environments also contain additional, irrelevant, visual information, such as cars passing on the street or other people in the room, that may distract a listener from the desired speech. The listener must use cognitive resources to suppress auditory and visual distractors and focus attention on the relevant auditory and visual cues in order to best understand the target talker and facilitate successful communication.

As listeners age, they face additional difficulties with speech perception due to loss of hearing acuity (e.g., Schuknecht, 1955, Gates and Mills, 2005), temporal processing ability (e.g., Gordon-Salant and Fitzgibbons, 1999, Pichora-Fuller, 2003), speed of processing (e.g., Geary et al., 1993), and working memory (e.g., Hasher and Zacks, 1988). Those with severe-to-profound hearing loss are sometimes treated with the surgical insertion of a cochlear implant (CI). This auditory prosthesis partially restores access to sound by mostly maintaining the temporal envelope, but causes severe spectral degradation of the auditory signal. With practice, CI users are often able to regain the ability to understand speech, especially with the benefit of visual cues in addition to the auditory signal.

While listeners of all ages can benefit from visual speech cues (Sommers et al., 2005, Tye-Murray et al., 2007), some have argued for decreased ability to integrate audio-visual information with age. Further, Ben-David and colleagues (2014) have argued that older adults have more difficulty inhibiting distractors than with detection of targets or speed of processing. Visual distraction in CI listeners, on the other hand, is less well studied. Several studies have found an advantage for the congenitally deaf on visual tasks (e.g., Strelnikov at al., 2009), including an advantage in visual selective attention (Dye et al., 2009). Champoux and colleagues (2008), however, showed interference of visual distractors for “inexperienced” CI users and proposed that auditory areas had been taken over by visual processes due to long periods of deafness which was reversible with cortical plasticity in more experienced CI users.

One way to test the influence of visual distractions is through the use of a visual secondary task. Gagné and colleagues (year) have looked extensively at listening effort revealed by a secondary task paradigm and shown differences due to aging. In difficult listening conditions, performance on a speech recognition task can become dramatically affected when the listener is also asked to perform another task at the same time. Van Gerven and Guerriero (2017) found effects of age during crossmodal (audio and visual) dual-tasks. However, the ability of CI users to perform speech recognition in a cross-modal dual-task paradigm is yet unknown.

Many experienced CI users are extremely good at recognizing speech when it is accompanied by visual cues, because the visual cues complement the temporal acoustic cues available to CI listeners. One type of speech where temporal cues are modified is foreign-accented speech. . Accented speech has many of the stress patterns, pronunciations, and prosodic contours of the talker’s native language superimposed on the English speech. With the lack of spectral information due to the CIs’ processors and the distortion of temporal cues due to the accented talker, the speech recognition of even the highest-performing CI user can be tested.

Previous work (Gordon-Salant, unpublished) tested two groups of listeners with normal hearing divided by age on AV stimuli of two talkers (one accented, one unaccented) and found no difference due to age for visual distraction, but all listeners performed worse when listening to accented talkers. Older listeners were differentially and negatively affected when a secondary task was added to the speech recognition task, especially when the talker had an accent.

QUESTION:

The purpose of this study was to determine the effect of visual distraction on speech perception when listening to a spectrally-degraded audio signal, whether through a CI or vocoded speech. In order to further expose the effect of CI users’ experience relying on temporal cues, two additional factors were employed: the accent of the target talker (perturbation of temporal cues) and the addition of a secondary task (increase in listening effort).

The hypotheses tested were that:

1. Visual distraction adversely affects listeners who are limited to spectrally-degraded speech, such as through a CI,

2. This effect is greater when speech is presented with a foreign accent or a secondary task is added, and

3. CI listeners perform better than their NH peers, even with the age- and auditory-only-performance-matching. This difference may be due to their practice with the spectral distortion of the CI device and presumably with experience in switching attention between auditory and visual inputs.

METHOD:

*Participants*:

Two groups of 20 participants were recruited. One group was composed of CI users with at least one year of experience with their device(s). They were tested using their own processors with their personal, clinical settings. They listened with their CIs only - if they were bimodal or only used a CI unilaterally, the hearing aid was removed and the non-implanted ear was blocked with xx for the experiment. The second group was composed of age-matched (± 2-3 years) normal-hearing participants (thresholds <= 25 dB HL at .5, 1, 2, and 4 kHz) who were matched in performance to CI listeners’ auditory-only abilities by adaptively changing the number of vocoded channels in sentence stimuli (details to follow in Procedures). All participants were native English speakers with normal or corrected to normal vision.

*Stimuli:*

AV stimuli were created using Adobe Premiere Pro. The foreground videos were professionally recorded at the National Foreign Language Center and consist of a close-up of the head and shoulders of a talker speaking sentences from the IEEE (Harvard) corpus. In order to study the effects of accent, two talkers were chosen: a male native-speaker of American English and a male native-speaker of Spanish from Peru with a moderate accent, as determined by pilot testing. The background video selected for this study depicted a busy street scene with people walking past and no accompanying audio (courtesy of ExploringAlabama on Youtube.com).

Stimuli were created to test the following conditions: audio-visual (AV), AV and visual distraction (AV+D), AV and visual secondary task (AV+ST), and secondary task alone. Twenty IEEE sentences were presented in each condition for a total of 60 sentences from each talker (120 unique sentences). Each IEEE sentence has five key words and was scored for number of keywords correct - irrespective of order. Sentences were presented in a fixed order within each condition, but conditions were presented in random order.

The audio for the CI group was low-pass filtered at 4 kHz with a 3rd-order Butterworth filter using Adobe Audition and combined with six-talker babble at a signal-to-noise ratio (SNR) of +10 dB. The babble consisted of six male speakers, three with American English accents and three with Spanish accents. The level of the stimulus sentences was equalized on the basis of root-mean-square (RMS) intensity in Matlab. Segments of babble matched in length to each stimuli were then equalized to a RMS value equivalent to 10 dB less than that of the stimuli before being combined in a single audio track for presentation via loudspeaker.

For the NH control group, the IEEE sentences with babble added at a +10-dB SNR were vocoded using Matlab. For an n-channel vocoder, the auditory speech signal was bandpass filtered using 3rd-order forward-backward Butterworth filters into n logarithmically spaced bands (36 dB/octave) between 200 and 4000 Hz with no pre-emphasis added. The temporal speech envelope from each band was extracted with a Hilbert envelope cutoff of 400 Hz and then used to modulate n sine carriers placed at the center frequency of each channel. The various modulated sine carriers were then combined to create the final vocoded output. This method was used to create 31 vocoded versions of the stimuli representing 2-32 channels.

*Procedures*:

Preliminary Measures:

Participants first completed a vision screening using a Snellen Eye Chart, scoring 20/50 or better using both eyes and corrective lenses if needed. They also completed the MoCA (Montreal Cognitive Assessment), with a minimum score of 26 in order to rule out any cognitive impairment. Age-matched participants (within 3 years of age) with normal hearing were required to have thresholds at 0.5, 1, 2, and 4 kHz within the normal range (≤25 dB HL).

If the participant passed these screening measures, he or she participant was asked to complete three other cognitive measures: the NIH Toolbox Flanker Inhibitory Control and Attention Test (Zelazo et al., 2013), the NIH Toolbox Pattern Comparison Processing Speed Test (Carlozzi et al., 2013), and the NIH Toolbox Dimension Change Card Sort.

Method to equate performance of acoustic listeners to CI listeners

Listeners with CIs were presented with the audio of one list of 10 sentences from the IEEE corpus in quiet spoken by a separate native English talker and asked to repeat the sentences aloud. These sentences were not used in the later experiment. The number of keywords correct was used to compute the audio-only performance of each participant.

Listeners with NH were presented with two sentences spoken by a talker not used in the later experiment accompanied by their written transcripts in both high (32 channels) and low (4 channels) vocoded conditions in order to gain familiarity with the spectral degradation of the auditory signal. They were then presented with 40 test sentences randomly distributed among 4, 8, 16, and 32 channels and asked to repeat the sentences. The number of keywords was computed into a percent correct for each condition and a psychometric function was fit to the data points. This psychometric function allowed the computation of the number of channels a NH participant needed in order to match performance (within 10%) of their age-matched CI participant. When the required number of channels was determined, the listener heard the practice sentences again with the number of channels that was then used for the rest of the experiment and a confirmation list of 10 sentences was administered to assure that performance was indeed matched.

Experimental Procedures

Participants were seated in a double-walled sound-attenuating booth 1.5 meters from a 37-inch LCD television screen at zero degrees azimuth and at eye level when seated. A single loudspeaker, also at one meter distance and zero degrees azimuth, was used to present the audio tracks at 65 dBHL. dB HL.

*Conditions:*

In the auditory-visual condition (AV), the listeners watched a video with a solid-colored background and repeated the sentences spoken by the talker. In the visual distraction (AV+D) condition, the background was replaced by a scene of people walking past. The task remained to repeat the sentence spoken by the talker while ignoring the people walking past. In the audio-visual with secondary task (AV+ST) condition, participants were asked to repeat the sentence and report the number of people who entered the camera view and walked away from the camera’s point of view. Participants were scored by number of keywords correct in the sentences and the number of people walking away. In the secondary task alone condition, the participants were asked to only count the number of people walking away while ignoring the spoken sentences. A practice video with all conditions was presented and if mistakes were made, participants were allowed to repeat conditions until the task was clearly understood.

ANALYSIS:

Percent correct dataspeech understand scores were arc-sine transformed. The first analysis used the arc-sine transformed percent correct repetition of the target sentences as the dependent variable in a repeated measures analysis of variance (RM-ANOVA) with both between- and within-subject variables. The between-subject variable was listener group (2 levels: CI listeners, vocoded NH listeners), and the within-subject variables were task condition (3 levels: AV, AV+D, AV+ST) and talker accent (2 levels: unaccented, accented). This analysis should reveal the main effects of visual distraction, secondary task, and talker accent, and interactions between them.

The “cost” of the secondary task was calculated for each subject by determining the performance on the task of counting people walking away while also repeating sentences versus that in the ST-alone condition. A t-test confirmed no difference in performance between groups on the ST alone condition. These cost values were then used as the dependent variable in an ANOVA with CI group versus NH vocoded group as the between-subject factor and talker (unaccented and accented) as the within-subject factor.

Trial-by-trial data were collected and will be used for mixed effects models as well.

PRELIMINARY RESULTS:

Six CI users have been tested so far. ANOVA revealed significant effects of talker accent [F(1,5)=66.703, p<0.001] with recognition of the accented talker poorer than that of the native English talker. There was also a main effect of distractor condition [F(2,10)=21.951, p<0.01] with the speech recognition in the distractor + secondary task condition worse than the others, but no significant interaction. Currently, there is also no significant difference between AV and AV + distraction conditions.

Figure 1. Average Performance of 6 adults with CIs divided by Native Spanish and Native English talkers in three distraction conditions. Bars show standard error of the mean.

Figure 2. Performance on Secondary Task by 6 adult users of CIs with a significant drop in performance while also repeating sentences, but no difference in performance due to talker accent.

DISCUSSION:

The lack of significant differences between AV and AV+D conditions in the current CI population does not support the hypothesis that visual distraction adversely affects listeners, nor the hypothesis that this effect is greater when the talker has a foreign accent. The current results do support the hypothesis that adding a secondary task shows a greater effect on speech perception than AV alone or with a distracting video.

At this time, we cannot speak to group differences as no NH data have been collected, but it is possible that the lack of significant effects of visual distraction may be an expression of the well-developed attention-switching skills of those using CIs.

Future work will continue gathering data on participants with CIs as well as begin the age- and performance-matching process with NH listeners (as soon as all 31 sets of videos are complete). An additional line of inquiry would involve testing both participants with CIs and NH in quiet and looking at the group effects of listening in noise with visual distractions present.

REFERENCES:

Ben-David, B. M., Eidels, A., & Donkin, C. (2014). Effects of aging and distractors on detection of redundant visual targets and capacity: do older adults integrate visual targets differently than younger adults?. PloS one, 9(12), e113551.

Champoux, F., Lepore, F., Gagné, J. P., & Théoret, H. (2009). Visual stimuli can impair auditory processing in cochlear implant users. Neuropsychologia, 47(1), 17-22.

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. Ear and hearing, 23(5), 439-449.

Cohen, J. I., & Gordon-Salant, S. (2017). The effect of visual distraction on auditory-visual speech perception by younger and older listeners. The Journal of the Acoustical Society of America, 141(5), EL470-EL476.

Dye, M. W., Hauser, P. C., & Bavelier, D. (2009). Is visual selective attention in deaf individuals enhanced or deficient? The case of the useful field of view. PloS one, 4(5), e5640.

Forster, S., & Lavie, N. (2007). High perceptual load makes everybody equal. Psychological science, 18(5), 377-381.

Gagne, J. P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. Trends in hearing, 21, 2331216516687287.

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet*, *366*(9491), 1111-1120.

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and aging*, *8*(2), 242.

Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech, Language, and Hearing Research*, *42*(2), 300-311.

Gosselin, P. A., & Gagne, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. Journal of Speech, Language, and Hearing Research, 54(3), 944-958.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In *Psychology of learning and motivation* (Vol. 22, pp. 193-225). Academic Press.

Hay‐McCutcheon, M. J., Pisoni, D. B., & Kirk, K. I. (2005). Audiovisual speech perception in elderly cochlear implant recipients. The Laryngoscope, 115(10), 1887-1894.

Musacchia, G., Arum, L., Nicol, T., Garstecki, D., & Kraus, N. (2009). Audiovisual deficits in older adults with hearing loss: biological evidence. Ear and hearing, 30(5), 505-514.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology*, *42*(sup1), 59-67.

Schuknecht, H. F. (1955). Presbycusis. *The Laryngoscope*, *65*(6), 402-419.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. Ear and hearing, 26(3), 263-275.

Strelnikov, K., Rouger, J., Demonet, J. F., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2013). Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain*, 136(12), 3682-3695.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. Ear and Hearing, 28(5), 656-668.

Van Gerven, P. W., & Guerreiro, M. J. (2016). Selective attention and sensory modality in aging: curses and blessings. Frontiers in human neuroscience, 10, 147.

Walden, B.E., Busacco, D.A. &, Montgomery, A.A. (1993). Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. Journal of Speech and Hearing Research, 36(2), 431-6.