RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 27 (2005)
Indiana University

Audiovisual Asynchrony Detection and Speech Perception in Normal-Hearing Listeners and Hearing-Impaired Listeners with Cochlear Implants[1]

Marcia J. Hay McCutcheon,[2] David B. Pisoni[2] and Kristopher K. Hunt[2]

Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405

---

[1] This work was supported by NIH-NIDCD Training Grant T32DC00012 to Indiana University and the Psi Iota Xi Sorority. This is a draft of a paper to be submitted for publication.

[2] DeVault Otologic Research Laboratory, Department of Otolaryngology—Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Audiovisual Asynchrony Detection and Speech Perception in Normal-Hearing Listeners and Hearing-Impaired Listeners with Cochlear Implants

Abstract. This study examined the effects of hearing loss and aging on the detection of AV asynchrony in both normal hearing adults and hearing impaired listeners with cochlear implants. Additionally, the relationship between AV asynchrony detection skills and speech perception was assessed. Twenty-five cochlear implant users and 22 normal-hearing adults participated in this study. Both elderly and middle-aged sub-groups of individuals were included in the normal hearing and cochlear implant groups. Individuals were asked to make judgments about the synchrony of AV speech and to complete three speech perception tests, the CUNY, HINT and CNC. No significant differences were observed in the detection of AV asynchronous speech between the normal-hearing listeners and the cochlear implant users. The data revealed, however, that older adults had wider windows over which they identified AV asynchronous speech as being synchronous than younger adults. Additionally, for normal hearing listeners the temporal AV asynchrony processing window was found to be correlated with speech perception measures. Specifically, wider temporal windows were associated with poorer speech perception skills. This trend was not observed in the hearing impaired population. These findings suggest that there may be fundamental differences in how speech is perceived in individuals with cochlear implants.

Introduction

The assessment of speech perception skills in cochlear implant users commonly involves the use of cues provided through one modality, namely, the information acquired through listening-alone. However, in order to fully understand speech perception processes it also is appropriate to evaluate the impact that visual cues have on individual word and sentence recognition abilities. In the normal hearing population, a number of studies have shown that both visual and auditory information play an important role in speech perception. For example, Sumby and Pollack (1954) demonstrated the importance of visual information for speech understanding in the presence of background noise. This study demonstrated that for extremely poor listening environments (i.e., a -30 dB signal-to-noise ratio) a 40% to 80% increase in word recognition can be achieved when visual cues are added to the auditory stimuli. The benefits of visual cues for the identification of sentences in noise also was demonstrated by Middelweerd and Plomp (1987). Additionally, McGurk and McDonald (1976) demonstrated that when a visual production of one consonant is paired with an auditory production of another consonant, a third consonant, neither visually or auditorily presented, is perceived. Specifically, when a visual /g/ is paired with an auditory /b/, a perceived /d/ will often occur. These studies demonstrate that the use of auditory and visual cues is crucial for the understanding of speech.

The temporal alterations of audiovisual signals also has been examined to determine how speech perception is affected when listening to degraded asynchronous multimodal signals (Grant & Greenberg, 2001; McGrath & Summerfield, 1985; Pandey, Kunov, & Abel, 1986). In one study, Grant and Greenberg demonstrated that normal hearing individuals could successfully recognize Harvard/IEEE sentences when they were filtered and exposed to audio-visual delays. The sentences were filtered into 1/3-octave bands, and two of the bands, one low and one high frequency band, were used to construct the final stimuli. They found that when the auditory signal led the visual signal by up to approximately 40 ms or the visual signal led the auditory signal by up to 160 to 200 milliseconds, the stimuli could be successfully recognized. McGrath and Summerfield also demonstrated that when using an F0-modulated pulse train audio feed as part of an audiovisual signal, AV asynchronous speech was not affected when the audio and visual delays were less than 160 milliseconds. Similar findings were reported by Pandey, Kunov, and Abel who demonstrated that in the presence of background noise, AV asynchronous sentences can be successfully perceived for asynchrony levels up to approximately 120 milliseconds. More recently, Conrey and Pisoni (2006) demonstrated that young adults were able to identify isolated AV asynchronous words as being synchronous over a temporal window of approximately 150 milliseconds. The ability to recognize speech even though the audio and visual cues are not synchronous is clearly advantageous in noisy or reverberant environments where the audio and visual cues might not be aligned.

However, very little exploration of how individuals with hearing impairment integrate auditory and visual signals has been conducted. A hearing loss would imply that not all of the auditory frequency bandwidths from an audiovisual speech signal are adequately perceived, thereby potentially preventing the complete integration of auditory and visual stimuli. Grant and Seitz (1998) studied a group of individuals with mild sloping to severe hearing losses to determine the importance of synchronous AV speech stimuli for speech understanding. Their findings showed that speech recognition for audiovisual sentences was not affected until the audio delay exceeded 200 ms.

The effects of aging on AV asynchrony detection have also received little attention in the past. It has been reported that elderly listeners have more difficulty than younger adults with temporal processing of speech (Fitzgibbons & Gordon-Salant, 1996; Gordon-Salant & Fitzgibbons, 2001, 2004; Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994; Snell, 1997; Sommers, Tye-Murray, & Spehar, 2005; Spehar, Tye-Murray, & Sommers, 2004). For example, the detection of brief temporal gaps between pairs of tones or noisebursts is about twice the magnitude for elderly individuals as it is for younger listeners (Fitzgibbons & Gordon-Salant, 1996; Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994; Snell, 1997). Also, processing rapid speech has been shown to be more challenging for elderly individuals compared to younger adults (Gordon-Salant & Fitzgibbons, 2001, 2004). Finally, research has suggested that although older individuals often show a decline of speechreading abilities, their ability to comprehend time-altered visual speech is not compromised (Sommers, Tye-Murray, & Spehar, 2005; Spehar, Tye-Murray, & Sommers, 2004). It could be hypothesized, therefore, that due to the changes in auditory and visual processing, elderly individuals might integrate auditory and visual signals in a fundamentally different manner than younger individuals. Differences in auditory and visual integration could affect the perception of AV asynchronous speech.

Although AV asynchrony perception has not been examined in the elderly population, the integration of auditory and visual information has been assessed using several behavioral measures. For example, Cienkowski and Carney (2002) examined the McGurk effect in younger and older adults to assess the effects of aging on the ability to integrate auditory and visual information. They found that the percentage of fused responses to an auditory /bi/ and a visual /gi/ and an auditory /pi/ and visual /ki/ were not significantly different in younger and older age groups, suggesting that older adults are just as successful at integrating auditory and visual signals as younger adults. However, because individual measures of auditory and visual performance were not measured, it is not clear whether the younger and older study participants were integrating the auditory and visual cues in similar manners. That is, older adults could have relied more heavily either on auditory or visual cues to fuse stimuli whereas the younger adults could have used auditory and visual cues equally for fusion.

To address these concerns, Sommers, Tye-Murray, and Spehar (2005) recently examined the individual contributions that auditory and visual information provide for the integration of AV stimuli.

To minimize differences in unimodal performance, all study participants had normal hearing, as defined as hearing thresholds better than 25 dB HL from 250 Hz to 4000 Hz, and normal or normal-corrected vision. The study participants were asked to repeat consonants, isolated words and sentences under several different signal-to-noise ratios that would produce similar auditory performance between the two age groups. The speech materials also were presented in audiovisual and visual-only modalities in order to assess the effects of auditory or visual cues on speech understanding. The older normal hearing participants demonstrated significantly poorer speechreading skills than the younger adults suggesting that aging may be associated with declines in mechanisms that are responsible for the successful encoding of visual information, independently of hearing status. The findings also demonstrated that auditory and visual enhancement scores were comparable for the younger and older age groups. Therefore, the age differences that were obtained in audiovisual performance appear to reflect age related differences in speechreading abilities rather than the ability to integrate or combine auditory and visual stimuli.

In order to further examine the effects of hearing loss and aging on the integration of auditory and visual information in speech perception, we examined AV asynchrony perception in middle-aged and elderly normal hearing adults and cochlear implant users. Specifically, the present study measured the AV asynchrony detection skills in normal hearing adults and cochlear implant users to determine whether individuals who use cochlear implants detect AV asynchronous stimuli differently than normal hearing persons. This study also examined the effects that aging may have on the perception of AV asynchronous stimuli. Finally, because speech understanding deteriorates as the AV signal becomes increasingly asynchronous (Grant & Greenberg, 2001; Grant & Seitz, 1998), another goal of this study was to assess the association between AV asynchrony detection and speech perception abilities. Determining how normal hearing adults and cochlear implant users perceive AV asynchrony might provide some new insights into the sources of variability that underlie the wide range of speech perception skills that are typically observed within the cochlear implant population.

Method

Study Participants

Both normal hearing listeners and cochlear implant users participated in this study. Two different groups of English speaking cochlear implant users were recruited, 13 elderly adults ranging in age from 66 to 80 (mean 73 years old), and 12 middle-aged adults ranging in age from 41 to 54 years old (mean 47 years old). These individuals received either a Cochlear Corporation, an Advanced Bionics, or a Med El cochlear implant between the years of 1995 and 2004 at the Indiana University School of Medicine, Department of Otolaryngology—Head and Neck Surgery. One elderly participant had bilateral implants; the first device was implanted in 1996 while the second device was implanted in 2004. These adults were all native English speakers and none of them reported a history of stroke or head injury. The normal hearing participants consisted of 12 middle-aged adults ranging in age from 41 to 55 years old (mean age 48 years old), and 10 elderly adults ranging in age from 65 to 79 years old with a mean age of 70 years old. All of the normal hearing participants were recruited locally through posted advertisements and word of mouth. All of the normal hearing study participants reported that English was their first language, that they did not have prior speechreading training, and that they had no history of stroke or head injury.

Screening Tests

Pure-tone air-conductions thresholds were obtained for all normal hearing listeners from octaves 250 Hz to 4000 Hz using a Grason-Stadler GSI 61 Clinical Audiometer and EAR insert earphones. For purposes of this study, normal hearing was defined as behavioral thresholds of 25 dB HL or better at all test frequencies. Additionally, all individuals included in this study had symmetrical audiometric hearing configurations (i.e., less than 20 dB HL difference between ears at one test frequency). Additionally, the sound field audiometric behavioral thresholds also were obtained for the cochlear implant users.

Screening for vision was completed prior to testing to ensure that all participants were capable of perceiving and encoding visual speech information. Normal or corrected-to-normal visual acuity of 20/25 or better was indicated for all study participants. Additionally, the Mini Mental Status Exam was administered to all individuals to assess cognitive function (Folstein, Folstein, & McHugh, 1975). All individuals who participated in this study received a score of 27 or better out of a possible 30 points. The mean score for cognitively intact individuals in the Folstein, Folstein, and McHugh study was 27.6 with a range of 24 to 30.

Procedures and Stimuli

Three speech perception measures were administered to all study participants. The Consonant-Nucleus-Consonant (CNC) word recognition test (Peterson & Lehiste, 1962), the Hearing in Noise Test (HINT) sentence recognition test (Nilsson, Soli, & Sullivan, 1994), and the City University of New York (CUNY) sentence test (Boothroyd, Hnath-Chisolm, Hanin, & Kishon-Rabin, 1988) were presented to study participants in an IAC booth. The auditory stimuli were presented at 70 dB SPL for the cochlear implant users and at 63 dB SPL for the normal hearing study participants. Background speech noise also was used for the normal hearing participants and presented at 70 dB SPL, thereby leading to a -7 dB signal-to-noise ratio. The CNC word test was administered first, followed by the HINT and the CUNY sentence tests. Additionally, the CUNY sentence test was presented in three modalities in the following order: auditory-only (A), visual-only (V) and audio-visually (AV). All study participants were instructed to repeat the stimuli they heard or saw for these tasks. Guessing was encouraged. For all tests, a percentage correct score was obtained as the dependent measure.

The speech AV asynchrony task conducted in this experiment was the same one employed by Conrey and Pisoni (2006). A list of ten familiar English words was presented to the listeners using a single talker. The words were chosen from the Hoosier Audiovisual Multitalker Database which contains digitized movies of isolated monosyllabic words spoken by single talkers (Lachs & Hernandez, 1998). The most intelligible talker of this database, as determined by Lachs (1999), was chosen for stimulus presentation. To prepare synchronous and asynchronous AV stimuli, Final Cut Pro 3 (copyright 2003, Apple Computer, Inc.) was used to manipulate the audio and visual signals. The stimuli were prepared such that the only cues that could be used to make judgments about the synchrony of the signals were temporally based between the audio and visual leads. Specifically, the audio track did not play while the screen was blank and all of the speech sounds and active articulatory movements remained within the movie.

Previous research on AV synchrony perception has revealed that normal-hearing young adults have a fairly wide range over which they will judge AV signals as being synchronous or asynchronous. That is, AV stimuli are typically judged as being asynchronous with 100% accuracy when the audio signal leads the visual signal by 300 ms (i.e., A300V) or more and when the visual signal leads the audio signal by 500 ms (i.e., V500A) or more (Conrey & Pisoni, 2006). For this study, 25 asynchrony levels that covered a range of 800 ms from A300V to V500A were used. Each successive level of asynchrony, either audio-leading or visual-leading, differed by 33.33 ms increments. Nine stimuli had auditory leads, one was synchronous, and 15 had visual leads for each of the ten stimulus words that were used. As a result, a total of 250 trials were presented to the participants in a randomized order. The visual and audio stimuli were presented using an Apple G4 computer and Advent sound field speakers, respectively. The speakers were placed at ± 45° azimuth from the listeners who were seated approximately 19 inches from both the speakers and a Dell flat screen computer monitor.

Before the session began, the participants were given both written and oral instructions for performing the task and were presented with examples of asynchronous and synchronous AV stimuli. For each trial, the participants were asked to judge whether the AV stimulus was synchronous or asynchronous ("in sync" or "not in sync"). They were instructed to press one button on a button box if they thought the audio and visual stimuli were synchronous and a different button if they thought the stimuli were asynchronous. In order to alert the participants for an upcoming AV token, a fixation mark ("+") flashed on the computer screen for 200 ms which was then followed by a blank screen for 300 ms.

Results

The behavioral audiometric threshold data for cochlear implant users and the normal-hearing individuals are displayed in Tables 1 and 2, respectively. The cochlear implant user data presented in Table 1 reveal similar mean behavioral threshold responses for younger and older cochlear implant users at 250 Hz and 500 Hz. A one way ANOVA revealed no significant differences in thresholds between the two groups for these two warble tone behavioral thresholds. Significant differences between the younger and older cochlear implant users were obtained for the 1000 Hz (F (1,25) = 6.16, $p = 0.02$), 2000 Hz (F (1,25) = 7.14, $p = 0.01$) and 4000 Hz (F (1,25) = 4.56, $p = 0.04$) behavioral thresholds.

Young Subjects	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**
abf1	28	24	24	22	26
abi1	14	22	20	22	22
abk1	22	34	22	22	28
abn1	24	26	28	22	26
abs1	24	28	26	22	38
abv1	40	40	40	35	35
abw1	18	26	24	22	24
aby1	18	28	24	32	32
acb1	22	26	28	20	28
acr1	44	36	30	28	32
adh1	36	35	32	28	28
adi1	36	12	12	18	26
Mean	27.2	28.1	25.8	24.4	28.8
Elderly Subjects	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**
abg1	24	32	36	36	32
abh1	28	36	38	28	36
abj1	24	24	28	28	30
abm1	32	32	32	31	36
abo1	26	24	30	26	28
abp1	24	26	32	30	38
abr1	30	35	25	20	20
abq1	20	22	28	28	28
abz1	32	38	36	32	36
acc1	22	22	24	28	34
acd1	26	32	28	26	30
acf1	20	30	28	28	34
adc1-R	42	42	40	34	44
adc1-L	34	36	40	36	40
Mean	27.4	30.8	31.8	29.6	33.3

Table 1. Sound field behavioral audiometric thresholds for cochlear implant users. Thresholds are listed in dB HL for each test frequency.

One way ANOVAs performed using the normal hearing behavioral threshold data presented in Table 2 revealed significant differences in thresholds between younger and older adults for the right ear at 1000 Hz (*F* (1,21) = 8.42, *p* = 0.009) and 4000 Hz (*F* (1,21) = 8.79, *p* = 0.008). Left ear significant differences between the two aged groups also were noted at 1000 Hz (*F* (1,21) = 4.48, *p* = 0.04) and 4000 Hz (*F* (1,21) = 4.85, *p* = 0.04). A significant difference in the left ear pure tone average (PTA) was revealed (*F* (1,21) = 5.50, *p* = 0.03) but no significant difference between the two aged groups for the right PTA was indicated. Previous research has documented that individuals over the age of 60 experience significant hearing loss at frequencies above 4000 Hz (Lee, Matthews, Dubno, & Mills, 2005; Pearson et al., 1995). We cannot, therefore, rule out the possibility that the older adults who participated in this study did not have significant hearing loss at 8000 Hz. A hearing loss at 8000 Hz could have implications for the outcome measures (i.e., speech perception and asynchrony detection tasks) that were obtained during the course of the project.

Young Subjects	**Right Ear**						**Left Ear**					
	250 Hz	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**	**PTA-R**	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**	**PTA-L**
NH1	20	15	5	5	15	8.3	15	5	5	0	5	3.3
NH2	10	5	0	-5	10	0	15	10	0	5	5	5
NH5	5	20	10	15	15	15	10	10	5	10	10	8.3
NH9	15	15	20	10	15	15	10	10	10	10	15	10
NH10	10	10	10	10	0	10	10	10	5	10	5	8.3
NH11	15	20	10	10	15	13.3	10	10	5	10	15	8.3
NH14	5	5	0	5	5	3.3	5	5	5	5	15	5
NH15	10	10	10	15	5	11.6	10	5	10	15	15	10
NH17	0	0	10	5	5	5	10	5	5	5	5	5
NH18	10	10	10	5	5	8.3	10	15	10	15	10	13.3
NH19	15	15	10	10	5	11.7	15	10	15	5	15	10
NH20	10	5	10	10	20	8.3	15	5	5	15	10	8.3
Mean	10.4	10.8	8.8	7.9	9.6	9.2	11.3	8.3	6.7	8.8	10.4	7.9
Elderly Subjects												
NH25	20	25	20	10	25	18.3	20	25	25	10	20	20
NH28	5	10	15	10	25	11.7	10	5	5	20	25	10
NH32	10	5	20	0	15	8.3	10	10	5	5	15	6.7
NH33	15	15	15	15	15	15	10	10	10	10	10	10
NH35	20	5	10	5	10	6.7	15	10	10	15	5	11.7
NH36	10	5	15	20	20	13.3	0	15	5	10	15	10
NH42	15	10	20	10	15	13.3	20	20	20	5	15	15
NH47	10	10	15	10	15	11.7	5	10	10	15	15	11.7
NH49	5	0	5	5	10	3.3	0	0	10	5	15	5
NH48	20	10	15	25	20	16.7	10	10	15	20	15	15
Mean	13	9.5	15	11	17	11.8	10	11.5	11.5	11.5	15	11.5

Table 2. Behavioral audiometric thresholds for normal hearing listeners. Thresholds were obtained using insert earphones and are listed in dB HL.

An individual cochlear implant user example of an AV asynchrony function is displayed in Figure 1 Panel A. In this figure, the mean proportion of synchronous responses is presented as a function of the asynchrony level in milliseconds. On the abscissa, the negative asynchrony levels indicate that the auditory signal led the visual signal by a specified time (e.g., A300V), the zero point indicates that both the audio and visual signals were synchronous in time (i.e., 0), and the positive asynchrony levels indicate that the visual signal led the audio signal (e.g., V400A). The ordinate axis represents the proportion of synchronous responses that were reported at a specific asynchrony level. Recall that each AV asynchrony level was presented using 10 different words and the listener's task was to judge whether or not the stimulus was out of sync. For this particular example, the trials of A300V, A267V, V367A, V400A, V433A, V467A, and V500A were judged to be asynchronous with 100% accuracy. This individual reported that the audio and visual signals were completely synchronous for the asynchrony levels of A67V, A33V, 0, V33A, V67A, V100A, and V133A. For all other asynchrony levels, the study participant inconsistently reported that the AV stimuli were synchronous.

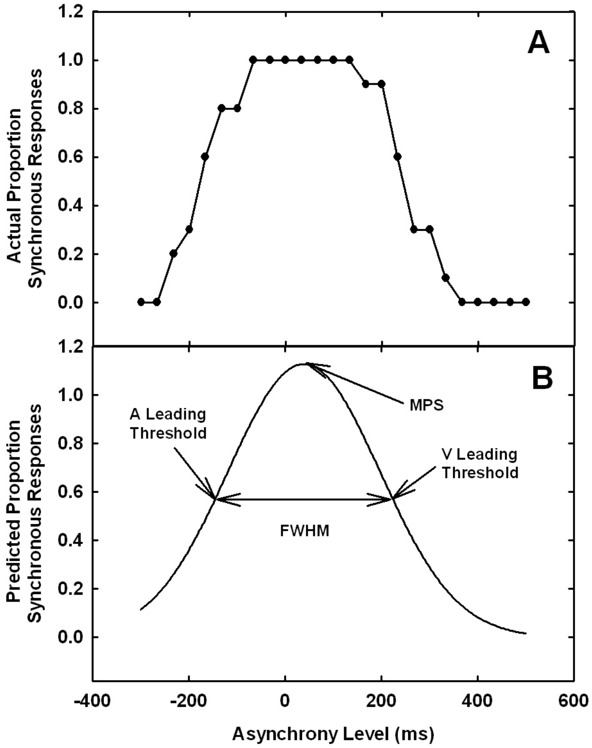

Figure 1. An individual AV asynchrony function. Panel A displays the observed function and Panel B shows the Gaussian curve fitted to the observed function. The proportion of synchronous responses is shown as a function of the asynchrony level. See text for details.

In order to quantify this AV asynchrony function, symmetrical Gaussian curves were fitted to individual asynchrony curves through the use of Sigma Plot 9.01 software and the following equation:

$$y = ae^{\left[-0.5\left(\frac{x-xo}{b}\right)^{2}\right]} \tag{1}$$

In this equation, *y* is the observed proportion of synchronous responses for each individual at each asynchrony level, *x*. The x-intercept, *xo*, represents the mean point of synchrony (MPS). Both *a* and *b* are generated parameters from the Sigma Plot software that aid with curve fitting. The Gaussian curve fitted to the individual asynchrony function shown in Panel A is displayed in Panel B of Figure 1. The four features that describe this AV asynchrony function are the MPS, the auditory (A) leading threshold, the visual (V) leading threshold and the full-width half maximum (FWHM). The A-leading threshold is the asynchrony level for the y value at 50% of the distance from the minimum to the maximum of the auditory leading portion of the curve (i.e., the left portion of the curve). Similarly, the V-leading threshold is the asynchrony level for the corresponding y value at 50% of the distance from the maximum to the minimum of the visual leading portion (i.e., the right portion) of the Gaussian function. The FWHM is the value of the asynchrony width at the half-maxima of the function. For this individual, the MPS was 39.15 ms, the A leading threshold was -145.23 ms, the V leading threshold was 226.65, and the FWHM was 371.88 ms.

The mean AV asynchrony data for all of the cochlear implant users and the normal hearing adults are shown in Figure 2. For both panels, the mean proportion of synchronous responses is displayed as a function of the asynchrony level. The top panel of the figure shows the data for the cochlear implant users and the bottom panel displays the data for the normal hearing adults. The overall results for the younger adults (averaged over cochlear implant users and normal hearing adults) are shown with the dotted line. The data for the older adults are shown using the solid line. The MPS, the A-leading threshold, the V-leading threshold and the FWHM values for the normal hearing and cochlear implant users are presented in Table 3. Two-way ANOVA analyses were performed using age and hearing status as independent variables and MPS, A-leading threshold, V-leading threshold, and FWHM as the dependent variables. A significant age-effect finding was revealed for the A-leading threshold (*F* (1,46) = 4.989, *p* = 0.03) and for the FWHM (*F* (1,46) = 4.921, *p* = 0.03). For these two variables, the younger adults (i.e., cochlear implant users and normal hearing adults) had A-leading thresholds that were closer to the point of AV synchrony (i.e., 0 on the abscissa in Figure 2) and had narrower FWHMs than the older adults. No other main effects or interactions were obtained for any of the other analyses.

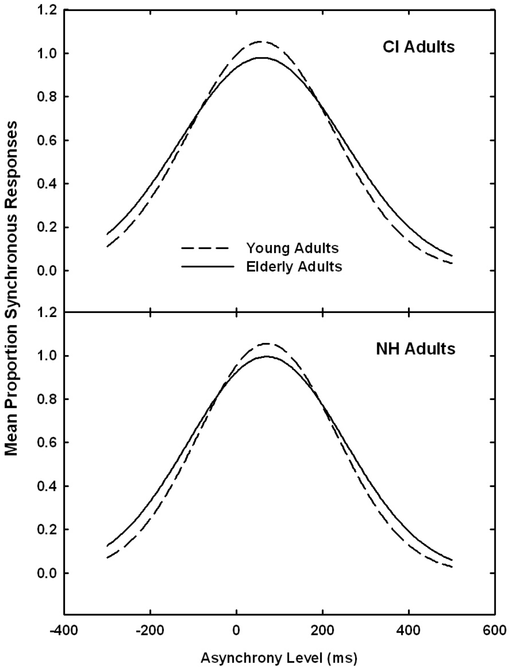

Figure 2. The mean AV asynchrony data for the cochlear implant users (top panel) and normal hearing adults (bottom panel). The mean proportion of the synchronous responses is displayed as a function of the asynchrony level.

	MPS	A leading Threshold	V leading Threshold	FWHM
Young - CI	58.4210	-135.9933	260.8817	396.8750
Elderly - CI	59.8846	-154.1708	295.8292	450.0000
Young - NH	72.3434	-112.3147	262.6853	375.0000
Elderly - NH	70.7044	-134.0327	294.0923	428.1250

Table 3. Mean Asynchrony Data. Values are in milliseconds; MPS: mean point of synchrony; FWHM: Full Width Half Maximum.

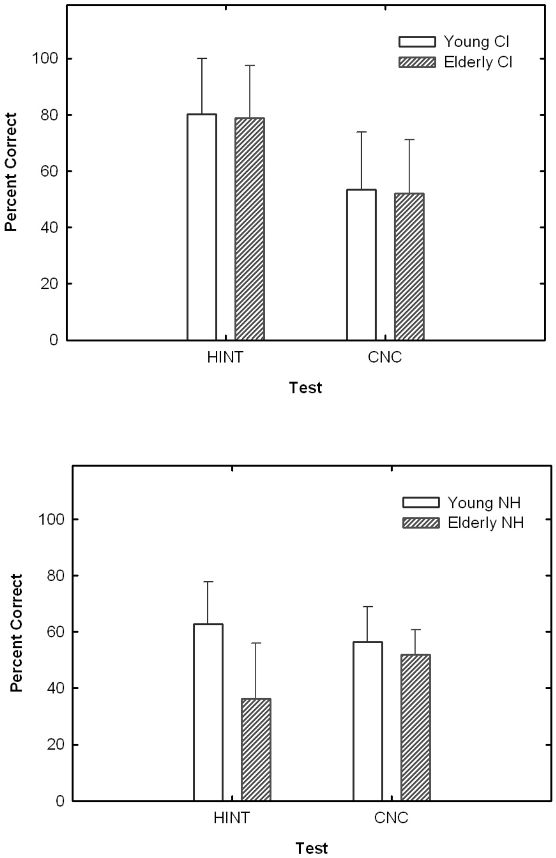

Figure 3. Mean percent correct scores and standard deviations for the HINT and CNC speech perception tests for the cochlear implant users (top panel) and normal hearing adults (bottom panel).

The results from the HINT, CNC, and CUNY speech perception tests are presented in Figures 3 and 4. Figure 3 displays the mean and standard deviations for the HINT and CNC sentence and word tests. The scores from the cochlear implant users are presented in the top panel; the scores from the normal hearing study participants are shown in the bottom panel. For both panels, the white bars show the data for the young individuals and the hatched bars show the data for the elderly participants. A two way ANOVA analysis of the HINT scores revealed a main effect for hearing status (F (1,43) = 31.34, $p <$ 0.001), and age (F (1,43) = 6.77, p = 0.013) and an interaction (F (1,43) = 5.37, p = 0.025). No significant differences were observed for the CNC data. Additionally, no significant correlations were observed between the FWHM data and the HINT and CNC data. There was, however, a trend for poorer HINT scores to be association with wider FWHMs for the young and elderly normal hearing adults, but this pattern did not reach significance (r = -0.371, p = 0.08).

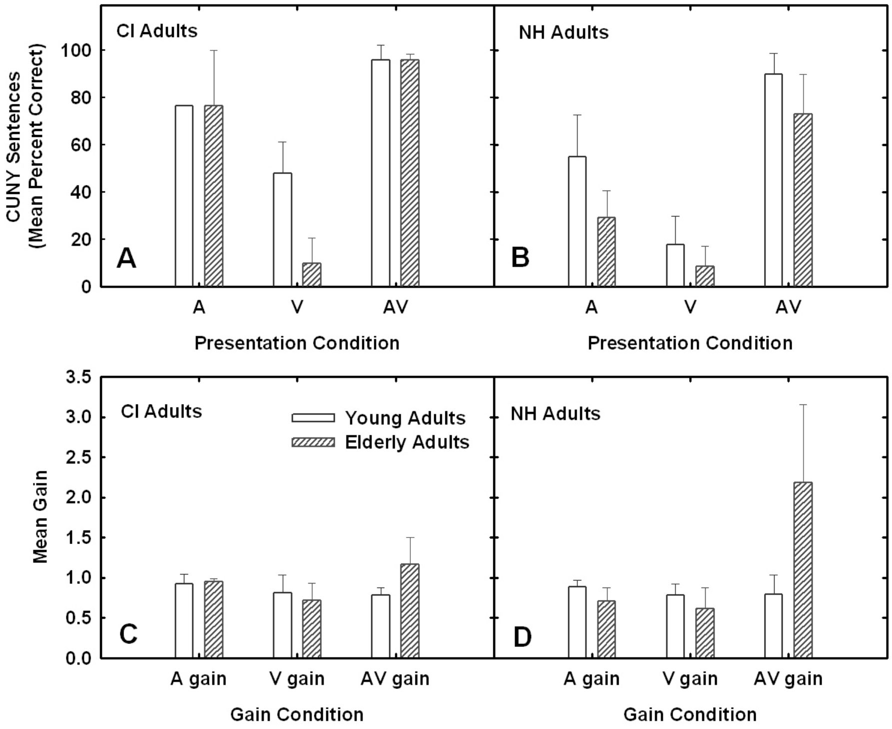

Figure 4. Mean percent correct (panels A and B) and gain scores (panels C and D) for the CUNY sentence test for the cochlear implant and normal hearing participants.

Figure 4 shows the results from the CUNY sentence test, displayed for the presentation modalities, auditory-alone (A), visual-alone (V) and audiovisually (AV). Panels A and C show the data for the cochlear implant users, and Panels B and D show the data for the normal hearing participants. The mean percent correct CUNY scores for each presentation condition (i.e., A, V and AV) are shown in Panels A and B for the cochlear implant and normal hearing participants, respectively. Additionally, Panels C and D show the A, V, and AV-gain or enhancement scores obtained from the CUNY sentence test. These gain measures assess how both auditory and visual cues contribute to overall speech understanding and are described in further detail below. The white bars in all panels show the data for the elderly adults and the hatched bars show the data for the younger adults. The error bars represent one standard deviation around the mean.

In order to assess the effects of aging and hearing status on the perception of the CUNY sentences, a series of two-way ANOVA analyses were conducted using the scores from the A, V and AV presentations. For the CUNY A presentation, significant main effects were found for age (F (1,43) = 5.69, $p = 0.02$) and hearing status (F (1,43) = 41.27, $p < 0.001$), along with an interaction (F (1,43) = 5.65, $p = 0.02$). As shown in Figure 4 Panels A and B, the mean CUNY A results for the normal hearing participants were lower than the scores for the cochlear implant users. Recall that the cochlear implant users listened to the sentences in quiet, and the normal hearing participants listened to the CUNY sentences in the presence of background noise in order to simulate the effects of a hearing loss and reduce performance from ceiling effects.

The main effect of hearing status was most likely the result of the added background noise that was used for the normal hearing adults and not for the cochlear implant users. In addition, the interaction between age and hearing status may be due to differences in the listening conditions between the two groups. Specifically, the added background noise probably reduced the performance of the older normal hearing adults to a greater extent than it affected the performance of the younger normal hearing adults. Degraded speech understanding in noise for elderly normal hearing individuals has been previously documented (Stuart & Phillips, 1996). These differences in performance for the older and younger groups were not observed in the cochlear implant population most likely due to the absence of background noise in the listening environment.

Main effects for the V-only presentation were observed for both age (F (1,43) = 50.34, $p < 0.001$) and hearing status (F (1,43) = 22.55, $p < 0.001$). An interaction also was observed (F (1,43) = 18.95, $p < 0.001$). The data for the CUNY V-only scores in Figure 4 Panels A and B reveal that the younger cochlear implant users and normal hearing adults were better speechreaders than the older adults, a finding that has been observed previously (Hay-McCutcheon, Pisoni, & Kirk, 2005; Sommers, Tye-Murray, & Spehar, 2005). However, the younger cochlear implant users identified V speech better than the younger normal hearing individuals. In addition, the data indicate that both the older normal hearing participants and the older cochlear implant users performed similarly in this speechreading task.

Finally, a two-way ANOVA of the CUNY AV scores revealed main effects for age (F (1,43) = 8.90, $p = 0.005$), hearing status (F (1,43) = 26.92, $p < 0.001$), and an interaction. The data shown in Figure 3 revealed three findings: first, both the older and younger cochlear implant users performed similarly on this task; second, the cochlear implant users performed better than the normal hearing participants, and third, the younger normal hearing adults achieved higher scores than the older normal hearing adults. The results shown here, however, should be interpreted with some caution because ceiling effects were observed for the AV results obtained for the cochlear implant users.

In order to assess the separate contributions that auditory and visual cues provide for speech understanding, auditory and visual gain scores were calculated. The A, V and AV scores obtained from the CUNY speech perception test were used to calculate the benefit that audition-alone (A-gain) and the vision-alone (V-gain) cues provide for speech perception (Lachs, Pisoni, & Kirk, 2001; Sommers, Tye-Murray, & Spehar, 2005). Specifically, the “A-gain” score represents the improvement in speech perception due to the addition of visual information to the auditory signal (i.e., AV-V/100-V), and conversely, the “V-gain” score represents the improvement in speech perception due to the addition of auditory cues to the visual signal (i.e., AV-A/100-A). For the A-gain score, the contributions that the visual cues add to the audiovisual results are subtracted and this value is subsequently divided by the difference between the possible visual-alone score and the obtained visual-alone score. Similarly, for the V-gain score, the contributions that the auditory cues add to the audiovisual results are determined and then divided by the difference between the possible auditory-alone score and the obtained auditory-alone score.

We also assessed the overall integration of auditory and visual information, and determined the superadditive nature of the use of combined modalities for speech perception (i.e., AV-gain = AV/A+V). For individual cases, if the combined AV performance is the same as the addition of the A-alone and V-alone scores, then AV-gain would be equal to one, and little AV enhancement would be indicated. Alternatively, if the combined AV scores are greater than the simple sum of the A and V scores alone, then the integration of auditory and visual information is beneficial (i.e., superadditive) for speech understanding.

The gain scores are presented in Panels C and D of Figure 4 for the cochlear implant users and the normal hearing adults, respectively. A series of two-way ANOVA analyses were conducted using the A-gain, V-gain and AV-gain scores as the dependent measures and hearing status and age as the independent variables. Because the AV results for the cochlear implant users were at ceiling, and these data were used to determine the gain measures, the following results need to be cautiously considered.

For the A-gain scores, significant main effects were obtained for age (F (1,43) = 4.823, p = 0.034) and hearing status (F (1,43) = 19.64, $p < 0.001$). In addition, an interaction was also observed (F (1,43) = 10.88, p = 0.002). For the V-gain scores, only a main effect for age was observed (F (1,43) = 4.46, p = 0.04). Main effects for age (F (1,43) = 37.67, $p < 0.001$), hearing status (F (1,43) = 12.81, p = 0.001), and an interaction effect (F (1,43) = 12.02, p = 0.001) were found for the AV-gain scores. The AV-gain scores displayed in Figure 4 reveal that the elderly normal hearing adults and cochlear implant users had scores above one, suggesting that the combined use of the auditory and visual cues provided greater benefit to the older adults than the younger adults.

To assess the relations between the CUNY speech perception scores and the AV asynchrony detection, Pearson correlations were conducted. The results are summarized in Figure 5. In this figure, the results for the normal hearing listeners are presented in the three left graphs; the results for the cochlear implant users are presented in the three right graphs. The white circles and triangles represent individual data from the young normal hearing adults and cochlear implant users, respectively. The gray diamonds and squares represent data from the elderly normal hearing adults and cochlear implant users.

The Pearson correlations, presented in each panel, reveal that for normal hearing individuals the wider the width of the FWHM the poorer the performance on CUNY A, V and AV tasks. This trend, however, was not observed in the scores obtained for the cochlear implant users. For the V results, there was a tendency for the cochlear implant users to have wider FWHMs with poorer perception, but this pattern did not reach significance (r = -0.368, p = 0.071). A significant correlation (r = -0.438, p = 0.029) was obtained for the AV data suggesting that wider FWHMs resulted in poorer speech perception scores. Because of ceiling effects, this finding needs to be viewed with some caution.

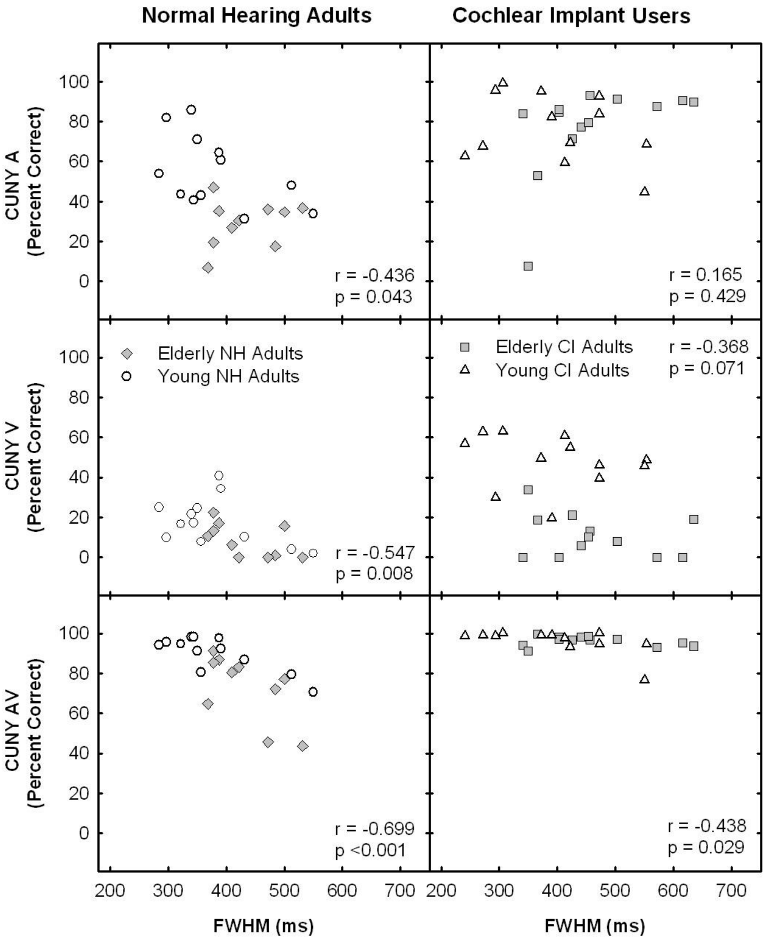

Figure 5. Correlation results for the CUNY and FWHM data for the normal hearing adults (left three panels) and cochlear implant users (right three panels).

The Pearson correlation coefficients for the gain scores and the FWHMs are presented in Figure 6. The normal hearing data are presented in the left three graphs and the data from the cochlear implant users are presented in the right three graphs. The white circles and the gray diamonds represent the data for the normal hearing young and older adults, respectively. The white triangles and the gray squares represent the data for the young and older cochlear implant users, respectively.

For the normal hearing individuals, a significant trend was observed for the A-gain ($r = -0.689$, $p < 0.001$) and V-gain ($r = -0.659$, $p < 0.001$) scores to decrease with increasing FWHM width. This trend was not observed for the AV-gain results ($r = 0.114$, $p = 0.613$). Although there was a tendency for the A-gain ($r = -0.346$, $p = 0.090$) scores to decrease with increasing FWHM width, this observation was not significant for the cochlear implant users. A significant correlation was observed for the V-gain ($r = -0.610$, $p = 0.001$) scores for the cochlear implant users. Specifically, lower V-gain scores were correlated with wider FWHMs. For both the normal hearing and the cochlear implant users, the AV-gain scores were not significantly correlated with the FWHM width.

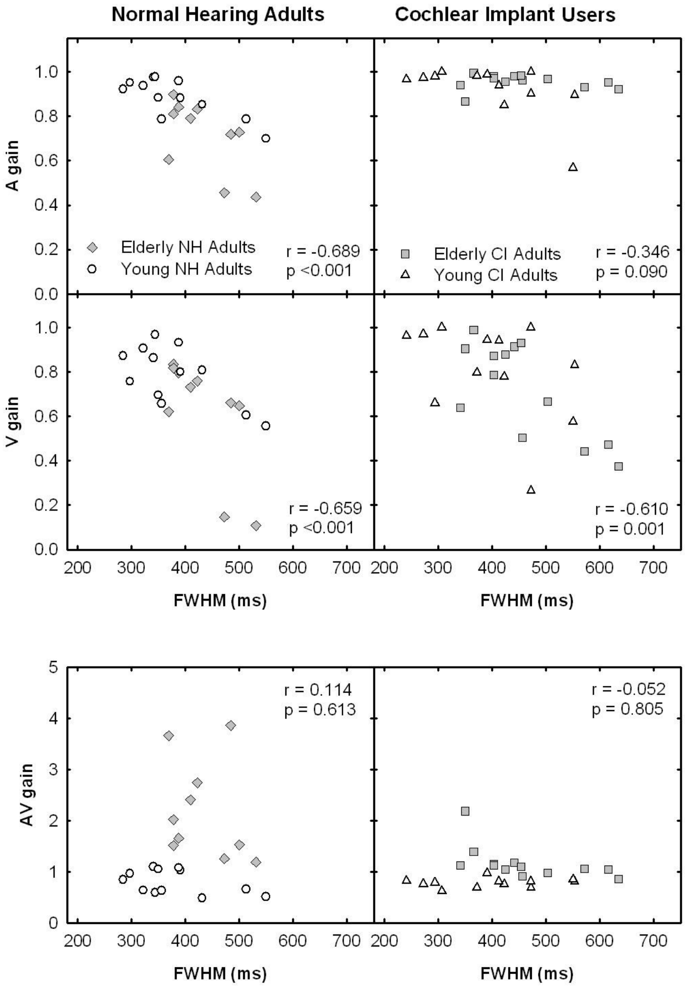

Figure 6. Correlation results for the gain scores and the FWHM (ms) data for the normal hearing adults (left three panels) and the cochlear implant users (right three panels).

Discussion

The goal of this study was to examine how normal hearing listeners and cochlear implant users perceive AV asynchrony in speech and assess the association between AV asynchrony detection and speech understanding abilities. The results of this study revealed no significant differences in performance between the normal hearing adults and the cochlear implant users in detecting and perceiving AV asynchronous single words. There was, however, a significant difference between the number of spoken words that were identified as being asynchronous for the elderly and middle-aged study participants. Specifically, the elderly normal hearing and cochlear implant individuals identified asynchronous words as being synchronous over a wider time window than did the younger adults. That is, the average FWHMs for the elderly normal hearing and cochlear implant population was approximately 440 ms compared to an average of 386 ms for the younger normal hearing individuals and cochlear implant users. Moreover, we found that the width of the asynchrony function was significantly correlated with the CUNY A, V and AV results for the normal hearing study participants. Correlations indicated that wider FWHMs were associated with poorer speech perception skills, a finding that replicated the earlier results of Conrey and Pisoni (2006). This pattern also was observed with the HINT scores in the current study, but the findings were not significant. Conversely, for individuals with cochlear implants, the A and V CUNY scores were not significantly correlated with the FWHM data. Overall, the results suggest that older individuals have more difficulty identifying AV asynchronous stimuli and that AV skills are correlated with speech perception abilities.

AV Asynchrony Detection

The AV asynchrony findings reported here are similar to those previously reported. The findings of Conrey & Pisoni (2006) suggested that the FWHM was on average 372 ms for young adults aged 18 to 22 years old, which was very similar to the FWHMs reported in the present study for the younger adults (i.e., 386 ms). Grant and Greenberg (2001), McGrath and Summerfield (1985), and Pandy, Kunov, and Abel (1986) also reported findings on speech understanding using AV asynchronous material and all three papers noted that sentences can be successfully identified when the auditory and visual components are approximately 200-250 milliseconds out of sync. Although the findings from the current study cannot be directly compared to the results of these earlier studies because of procedural differences, it is clear that AV asynchronous speech is perceived as synchronous over a window of approximately several hundred milliseconds.

AV Asynchrony Detection and Aging

The findings from this study also suggest that age rather than hearing impairment is more closely tied with the detection of AV asynchronous speech. The data displayed in Figure 2 suggest that compared to younger adults, older individuals have a significantly wider temporal integration window over which they identify AV asynchronous speech as being synchronous. The present findings suggest that individuals with a severe-to-profound hearing loss who use a cochlear implant do not have more difficulty detecting AV asynchronous speech than individuals with normal hearing. To more fully understand the effects of cochlear implantation on the perception of AV asynchronous speech, further work should focus on the identification, rather than just the detection of AV asynchronous speech in individuals who use a cochlear implant. Additionally, future work should address the effects that the degree of hearing loss has upon both the detection and understanding of AV asynchronous speech. Through these types of studies it may be possible to more clearly describe the effect of hearing loss on the perception of AV asynchronous speech.

Age-related effects also have been reported in several previous studies that evaluated auditory perception abilities in younger and older listeners (Fitzgibbons & Gordon-Salant, 1996; Gordon-Salant & Fitzgibbons, 2004; Pichora-Fuller & Souza, 2003; Stuart & Phillips, 1996). Specifically, several studies have found that older normal hearing listeners have more difficulty than young normal hearing adults with temporal processing tasks such as gap detection, sound duration discrimination, and identifying time compressed speech (Fitzgibbons & Gordon-Salant, 1996; Gordon-Salant & Fitzgibbons, 2004). Other studies have reported that younger adults correctly answer more comprehension questions after listening to a passage presented in a -15 dB signal-to-noise ratio condition than do older adults (Schneider, Daneman, Murphy, & Kwong See, 2000). Stuart and Phillips (1996) also demonstrated that older normal hearing listeners identified fewer monosyllabic words when presented in background noise than did younger adult listeners. Evidence suggests that the declines in speech perception performance can be partially attributed to changes that have occurred within the peripheral auditory system (Humes, 1996; Schneider, Daneman, Murphy, & Kwong See, 2000; Souza & Turner, 1994). However, as noted above, the observed differences between younger and older adults with more complex tasks such as gap detection and sound duration discrimination cannot exclusively be attributed to peripheral sensory deterioration (Fitzgibbons & Gordon-Salant, 1996; Gordon-Salant & Fitzgibbons, 2004; Pichora-Fuller & Souza, 2003). Most likely, neurophysiological changes that occur within the central auditory system also contribute to the declines in performance observed in complex listening tasks with elderly individuals.

In terms of the auditory periphery, both younger and older adults included in this study had hearing within normal limits. However, within the range of normal hearing, the behavioral audiometric threshold data suggested that the older normal hearing individuals had significantly higher thresholds than the younger normal hearing individuals at 1000 Hz and 4000 Hz. Additionally, the sound field thresholds for the cochlear implant users revealed that the older study participants had significantly lower thresholds at 1000 Hz, 2000 Hz and 4000 Hz. It is possible, therefore, that the differences in AV asynchrony detection could have been a direct consequence of the physiological differences in the peripheral auditory system between younger and older individuals.

Differences in the central auditory systems between younger and older adults also should be considered when explaining the observed differences in AV asynchrony detection. Specifically, several studies have shown that older adults experience difficulty with tasks that require them to divide their attention. Madden, Pierce, and Allen (1996) demonstrated that the reaction time to identify specific tokens from a group of distracting tokens was significantly longer in an elderly group of individuals aged 63 to 70 than in a young group of individuals aged 18 to 29 years old. Mayr (2001) reported that in a task requiring study participants to switch between different types of decisions between trials, older adults (mean age 71 years old, SD=3.3 years) had a significant longer reaction time for this task than did younger adults (mean age 33 years old, SD=1.4 years). It is possible, therefore, that the attentional demands that were required in the current study (i.e., attending to both the auditory and visual streams and making a conscious and explicit decision about their synchrony) placed greater processing demands on the older participants than the younger participants, and this could have contributed to the observed differences in performance between the two aged groups.

AV Asynchrony Detection and Speech Perception in Cochlear Implant Users

Contrary to the findings of the data for the normal hearing individuals, the relation between the AV asynchrony detection task and the speech perception scores were not highly correlated as shown in Figure 4. A correlation was observed between the AV CUNY results and the FWHM data, but this finding needs to be interpreted with some caution due to the ceiling effects noted with the AV CUNY data. Because the mechanism responsible for the correlation of the CUNY speech perception data and the FWHM in normal hearing individuals is not well understood, it is difficult to determine the reason for the lack of correlation found between the speech perception findings and the FWHM data in the cochlear implant users.

It is possible that for normal hearing individuals both the auditory and visual domains were effectively utilized to detect the presence of AV asynchrony, and that the processing of the AV input signal along the peripheral and central nervous system pathways was successfully and effectively completed. Additionally, it is possible that the processing of AV asynchronous stimuli and speech occurs using similar mechanisms in normal hearing individuals. Conversely, the cochlear implant users have had inadequate or altered peripheral auditory pathway processing prior to and following implantation. This change in processing strategies has been documented in neural plasticity data obtained from animal models (Shepherd, Baxi, & Hardie, 1999; Shepherd & Hardie, 2001). Thus altered peripheral processing could have an impact on the central processing of the signal, which would ultimately affect the perception of the AV asynchronous stimuli and the speech tokens associated with the speech perception tasks. The integration of auditory and visual information may occur in a fundamentally different manner for cochlear implant users compared to the normal hearing individuals and this processing might have an impact on both the perception of AV synchronous and asynchronous speech stimuli.

Conclusions

In summary, the findings from this experiment suggest that aging has a greater effect on the detection of AV asynchronous speech than a severe-to-profound hearing loss that has been partially corrected through the use of a cochlear implant. For normal hearing adults, the width of the temporal processing window over which AV asynchronous speech was identified as being synchronous was correlated with speech perception skills. We found that the perception of wider temporal windows was associated with poorer speech understanding. Conversely, for cochlear implant users, the temporal width of the AV asynchrony function was not correlated with speech perception skills. The findings suggest that the perception of speech may occur in a fundamentally different manner for hearing-impaired individuals who use cochlear implants users than it does for normal hearing adults.

References

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., & Kishon-Rabin, L. (1988). Voice fundamental frequency as an auditory supplement to the speechreading of sentences. *Ear & Hearing, 9*, 306-312.

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

Conrey, B., & Pisoni, D.B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *Journal of the Acoustical Society of America, 119*, 4065-4073.

Fitzgibbons, P.J., & Gordon-Salant, S. (1996). Auditory temporal processing in elderly listeners. *Journal of the American Academy of Audiology, 7*, 183-189.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189-198.

Gordon-Salant, S., & Fitzgibbons, P.J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research, 44*, 709-719.

Gordon-Salant, S., & Fitzgibbons, P.J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *Journal of the Acoustical Society of America, 115*(4), 1808-1817.

Grant, K.W., & Greenberg, S. (2001). *Speech intelligibility derived from asynchronous processing of auditory-visual information.* Paper presented at the International Conference on Auditory Visual Speech Processing, Scheelsminde, Denmark.

Grant, K.W., & Seitz, P.F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America, 104*, 2438-2450.

Hay-McCutcheon, M.J., Pisoni, D.B., & Kirk, K.I. (2005). *Speech recognition skills in the elderly cochlear implant population: A preliminary examination.* Paper presented at the MidWinter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, Feb 19-24, 2005.

Humes, L.E. (1996). Speech understanding in the elderly. *Journal of the American Academy of Audiology, 7*, 161-167.

Lachs, L. (1999). Use of partial stimulus information in spoken word recognition without auditory stimulation. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 81-118). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., & Hernandez, L.R. (1998). Update: The Hoosier audiovisual multitalker database. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 377-388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lachs, L., Pisoni, D.B., & Kirk, K.I. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear & Hearing, 22*, 236-251.

Lee, F.S., Matthews, L.J., Dubno, J.R., & Mills, J.H. (2005). Longitudinal study of pure-tone thresholds in older persons. *Ear & Hearing, 26*, 1-11.

Madden, D.J., Pierce, T.W., & Allen, P.A. (1996). Adult age differences in the use of distractor homogeneity during visual search. *Psychology and Aging, 11*, 454-474.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging, 16*, 96-109.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, 678-685.

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

Middelweerd, M.J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America, 82*, 2145-2147.

Nilsson, M., Soli, S.D., & Sullivan, J.A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*, 1085-1099.

Pandey, P.C., Kunov, H., & Abel, S.M. (1986). Disruptive effects of auditory signal delay on speech perception with lipreading. *The Journal of Auditory Research, 26*, 27-41.

Pearson, J.D., Morrell, D.H., Gordon-Salant, S., Brant, L.J., Metter, E.J., Klein, L.L., et al. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *Journal of the Acoustical Society of America, 97*, 1196-1205.

Peterson, G.E., & Lehiste, I. (1962). Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders, 27*, 62-65.

Pichora-Fuller, M.K., & Souza, P.E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42*, 12S11-12S16.

Schneider, B.A., Daneman, M., Murphy, D.R., & Kwong See, S. (2000). Listening to discourse in distracting settings: The effects of aging. *Psychology and Aging, 15*, 110-125.

Schneider, B.A., Pichora-Fuller, M.K., Kowalchuk, D., & Lamb, M. (1994). Gap detection and the precedence effect in young and old adults. *Journal of the Acoustical Society of America, 95*, 980-991.

Shepherd, R.K., Baxi, J.H., & Hardie, N.A. (1999). Response of inferior colliculus neurons to electrical stimulation of the auditory nerve in neonatally deafened cats. *Journal of Neurophysiology, 82*, 1363-1380.

Shepherd, R.K., & Hardie, N.A. (2001). Deafness-induced changes in the auditory pathway: Implications for cochlear implants. *Audiology and Neuro-Otology, 6*, 305-318.

Snell, K.B. (1997). Age-related changes in temporal gap detection. *Journal of the Acoustical Society of America, 101*, 2214-2220.

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

Souza, P.E., & Turner, C.W. (1994). Masking of speech in young and elderly listeners with hearing loss. *Journal of Speech and Hearing Research, 37*, 655-661.

Spehar, B., Tye-Murray, N., & Sommers, M. (2004). Time-compressed visual speech and age: A first report. *Ear & Hearing, 25*, 565-572.

Stuart, A., & Phillips, D.P. (1996). Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners. *Ear & Hearing, 17*, 478-489.

Sumby, W.H., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.