Psychology and Aging
2011, Vol. 26, No. 2, 427–438

0882-7974/11/$12.00 DOI: 10.1037/a0021683

Does Audiovisual Speech Offer a Fountain of Youth for Old Ears? An Event-Related Brain Potential Study of Age Differences in Audiovisual Speech Perception

Axel H. Winneke and Natalie A. Phillips
Concordia University

The current study addressed the question whether audiovisual (AV) speech can improve speech perception in older and younger adults in a noisy environment. Event-related potentials (ERPs) were recorded to investigate age-related differences in the processes underlying AV speech perception. Participants performed an object categorization task in three conditions, namely auditory-only (A), visual-only (V), and AV_{speech}. Both age groups revealed an equivalent behavioral AV_{speech} benefit over unisensory trials. ERP analyses revealed an amplitude reduction of the auditory P1 and N1 on AV_{speech} trials relative to the summed unisensory (A + V) response in both age groups. These amplitude reductions are interpreted as an indication of multisensory efficiency as fewer neural resources were recruited to achieve better performance. Of interest, the observed P1 amplitude reduction was larger in older adults. Younger and older adults also showed an earlier auditory N1 in AV_{speech} relative to A and A + V trials, an effect that was again greater in the older adults. The degree of multisensory latency shift was predicted by basic auditory functioning (i.e., higher hearing thresholds were associated with larger latency shifts) in both age groups. Together, the results show that AV speech processing is not only intact in older adults, but that the facilitation of neural responses occurs earlier in and to a greater extent than in younger adults. Thus, older adults appear to benefit more from additional visual speech cues than younger adults, possibly to compensate for more impoverished unisensory inputs because of sensory aging.

Keywords: ERPs, audiovisual speech, multisensory processing, auditory evoked potentials, aging

Thanks to medical and technical advancements, better nutrition, and healthier lifestyles, the life expectancy and hence the proportion of senior citizens is increasing (Government of Canada, 2002; Statistics Canada, 2005). Healthy aging can lead to changes in both sensory-perceptual abilities and higher-order cognitive functions (Schneider & Pichora-Fuller, 2000). Despite growing interest in age-related changes in sensory and cognitive functioning, many aspects remain poorly understood. This study focused on the relationship between aging and changes in audiovisual (AV) speech perception. The ability to integrate both sources of sensory information is especially important when information in one or both of the sensory channels is unclear or ambiguous (e.g., when having a conversation with a lot of background noise). There is clear evidence that the availability of visual speech input in a noisy acoustic environment is perceptually equivalent to boosting the volume of the auditory speech by 10 to 15 dB (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954). This highlights the potential of AV speech to improve communication even in individuals who do not have a hearing impairment and are not trained in lipreading.

There is an inverse relation between increasing age in adulthood and the functioning of our sensory systems. With respect to auditory function, many older adults (OAs) experience an age-related hearing loss (presbycusis), which affects the perception of high-frequency sounds and can lead to difficulties in speech comprehension (Erber, 2002; Schneider & Pichora-Fuller, 2000; Wingfield, Tun, & McCoy, 2005). Even OAs with age-appropriate normal hearing can have speech perception deficits in quiet listening conditions, a deficit that is exacerbated in noisy environments (CHABA [Committee on Hearing and Bioacoustics], 1988; Kim, Frisina, Mapes, Hickman, & Frisina, 2006).

Similarly, visual abilities such as acuity and contrast sensitivity decline with age (Erber, 2002; Schneider & Pichora-Fuller, 2000). Despite ample evidence of sensory decline in OAs in each separate modality, less is known about their interactions in OAs. With both unisensory information channels compromised, one could expect that AV perception including AV speech would also decline with

---

Axel H. Winneke, and Natalie A. Phillips Department of Psychology/Centre for Research in Human Development, Concordia University, Montréal, Québec, Canada.

These studies were supported by a grant awarded to Natalie A. Phillips from the Canadian Institutes of Health Research (Grant MOP-97808). The authors also acknowledge membership in the Centre for Research on Language, Mind, and Brain, McGill University, Montréal, Québec, Canada. A portion of the data was reported at the 8th International Conference on Auditory-Visual Speech Processing (AVSP), September 2009, Norwich, United Kingdom.

Correspondence concerning this article should be addressed to Natalie A. Phillips, Department of Psychology, Concordia University, 7141 Sherbrooke Street West, Montréal, Québec Canada. E-mail: natalie.phillips@concordia.ca

age. As an alternative, age-related decline of sensory functioning is relevant to the inverse effectiveness hypothesis of multisensory interactions.[1] This hypothesis states that the gain derived from a multisensory stimulus is larger when the unisensory channels are less effective on their own (Stein & Meredith, 1993). Consequently, because of the decline in unisensory abilities, OAs could benefit *more* from the combination of audiovisual stimuli than younger adults (YAs) whose sensory channels are intact.

The Effect of Age on AV Processing

Previous research on age-related changes in AV speech perception has led to a variety of findings, possibly because of differences in stimulus materials (e.g., syllables, words, or sentences) and screening measures for participation (e.g., visual acuity, hearing level, cognitive functioning). For example, Cienkowski and Carney (2002) investigated AV speech perception in healthy younger and older participants who listened to syllables in a quiet environment. A third group consisted of young controls who listened to syllables in a noisy background to match hearing thresholds to that of the OAs. All participants demonstrated normal visual acuity, visual contrast sensitivity, and age-appropriate auditory hearing levels (with OAs showing mild hearing loss for higher frequencies). The task was to name the syllable they perceived; syllables were presented auditory-only (A-only) and audiovisually (AV) to measure the extent to which participants showed the McGurk effect (McGurk & MacDonald, 1976) in the AV condition. In a classic McGurk paradigm, an auditory syllable is dubbed onto a video of a speaker saying a different syllable (e.g., an auditory /ba/ combined with a visual /ga/, leading to the perception of /da/). The McGurk effect refers to a perceptual phenomenon in which participants perceive a syllable that was presented neither auditorily nor visually, suggesting that auditory and visual speech cues were integrated. In Cienkowski and Carney (2002), all groups integrated syllables equally well. However, when integration failed, the OAs and the hearing-matched young controls tended to choose visual rather than auditory alternatives more often than YAs with intact hearing (i.e., no noise). That is, OAs relied more on visual speech cues when auditory information was ambiguous.

Sommers, Tye-Murray, and Spehar (2005) showed poorer speechreading abilities for OAs compared with YAs. Participants with intact sensory functioning had to identify syllables, words, and sentences presented in V-only, A-only, and AV format. To measure the extent to which additional visual speech cues enhanced performance relative to A-alone trials (i.e., visual enhancement), error rate in the A-alone condition was equated in each group to 50% by titrating the intensity of background babble noise. The same signal/noise (S/N) ratio was used for the AV condition. OAs performed more poorly than YAs in the V-only and AV conditions. However, after factoring out V-only performance, both groups showed the same degree of behavioral improvement indicating that YAs and OAs were equally successful in combining auditory and visual speech cues.

Although previous studies have shown that the AV performance of OAs was generally poorer than YAs, which may be explained by poorer speechreading abilities in the OAs, the ability to integrate auditory and visual speech cues remained intact (Cienkowski & Carney, 2002; Sommers et al., 2005; Tye-Murray et al., 2008). This conclusion has also been made in a bimodal target detection task (Bucur, Allen, Sanders, Ruthruff, & Murphy, 2005) where OAs and YAs responded faster to AV targets than unimodal targets. This facilitation was because of the interaction of the two sensory channels allowing for the integration of multisensory information. It is interesting to note that OAs appeared to use the visual speech cues more than YAs, possibly to compensate for sensory decline (Cienkowski & Carney, 2002; Thompson, 1995) or limited attentional resources (Thompson & Malloy, 2004).

One might argue that OAs are “permanently” in suboptimal perceptual conditions because of sensory declines, and, according to the principle of inverse effectiveness, should benefit more from multisensory information. Laurienti, Burdette, Maldjian, and Wallace (2006) investigated this idea in a target discrimination task with YAs and OAs screened for normal sensory and cognitive functions. The stimuli consisted of colored disks (red and blue) presented on a computer screen (V-only), a female voice uttering the color words (A-only), or both disks and voice combined (AV). OAs responded significantly slower in all conditions but their relative benefit from the added visual stimulus to the auditory cue was significantly larger than for YAs. Using a similar target discrimination task, Hugenschmidt, Mozolic, and Laurienti (2009) demonstrated enhanced multisensory integration in OAs relative to YAs under both divided and modality specific-attention, namely a proportionally larger decrease in response times to multisensory relative to unisensory trials. The authors concluded that integrational mechanisms remained intact in OAs and that attentional demands (i.e., selective vs. divided) influenced multisensory integration equally in both age groups.

Despite these just described behavioral findings, there is relatively little information about the neural mechanisms underlying AV speech perception in OAs. There have been a few studies investigating neural processes of AV speech perception but these have been restricted to YAs and stimuli usually comprised syllables rather than words or sentences.

Previous AV speech studies investigating the electrophysiological processes mainly looked at early auditory event-related brain potentials (ERPs). ERPs provide an on-line measurement of sensory, perceptual, and cognitive processing. They reflect voltage variations in electrical brain activity in response to a stimulus and are extracted from the electroencephalogram (EEG) via signal averaging. The various ERP components are typically identified by their polarity (positive or negative), latency (occurrence after stimulus onset, measured in milliseconds), amplitude (in μV), and topographical distribution across the scalp (Rugg & Coles, 1995). Amplitude is generally taken to reflect the strength of underlying neural responses, while latency generally reflects the time point of the peak relative to the eliciting stimulus. Thus, earlier peaks are presumed to reflect the output of a process that is completed earlier

---

[1] In this article, we use the term “integration” to refer to the underlying neural and/or perceptual mechanisms or processes used to combine information from two sensory sources (e.g., Grant, 2002). In contrast, we use “enhancement” to refer to an improvement in behavioral performance in a multisensory condition relative to a unisensory condition; thus, enhancement is neutral with regard to the underlying mechanism and does not necessarily imply a multisensory interaction. However, we do use the term integration when describing behavioral data which can only be because of the interaction between two unisensory inputs (such as in the case of the McGurk effect or demonstrated violations of the race model, see below).

in time than a later peak. Early auditory ERPs consist of a sequence of obligatory brain responses referred to as the P1-N1-P2 complex. This complex is elicited by auditory stimuli and consists of a series of positive and negative voltage deflections which peak around 50 to 80 ms (P1), 100 to 150 ms (N1), and 200 to 250 ms (P2) after stimulus onset. The functional role of the N1 is related to stimulus detection and the encoding of auditory stimulus properties; its neural source has been localized to the auditory cortex (Eggermont & Ponton, 2002; Näätänen & Picton, 1987). The earlier P1 is thought to have subcortical contributions, including the reticular activating system (e.g., Erwin & Buchwald, 1987). The functional significance of the P2 is less well articulated, and it is usually described as the positivity after the N1, as part of a complex. As discussed later, the amplitude of auditory ERPs can be modulated by other components sensitive to contextual factors (e.g., attention) occurring in the same time period.

AV speech studies recording ERPs elicited by syllables have shown that (1) the amplitude of the auditory N1 during an AV speech condition was reduced relative to the summed ERP responses of the A and V conditions (Besle, Fort, Depuelch, & Giard, 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2005) and (2) that the auditory brain processes were faster (i.e., a latency decrease) relative to auditory-alone and summed A + V trials (Stekelenburg & Vroomen, 2007; van Wassenhove, et al., 2005). Because of visual speech cues preceding auditory speech cue by up to 300 ms, van Wassenhove and colleagues (2005) proposed that auditory processing of AV speech benefits from the visual cue predicting what the auditory system can expect. It is interesting to note that the authors showed that the N1 became increasingly earlier with increasing predictability of a spoken syllable. Stekelenburg and Vroomen (2007) observed similar reductions in N1 amplitude and latency. They demonstrated that this effect seemed to be related to the visual cue temporally preceding the auditory cue, because they observed similar electrophysiological responses in trials involving nonspeech stimuli.

The current study will examine the extent to which performance and electrophysiological patterns differ between healthy YAs and OAs during AV speech perception in noisy environments. Given that the individual sensory modalities function less optimally in OAs than YAs, we predict that OAs should benefit more from AV speech than YAs, in line with the inverse effectiveness hypothesis. At the behavioral level, OAs should show relatively faster and more accurate responses than YAs during multisensory AV trials than during unimodal conditions (i.e., only listening to or watching speech). At the neural level we predict effects for the early auditory P1, N1, and P2 sensory components. Specifically, we expect: (1) an amplitude reduction during AV trials relative to the unisensory trials (A and V) and their summed response (A + V), and (2) a N1 latency shift, such that it will peak earlier during AV trials compared with unimodal trials and summed responses (A + V). We hypothesize that this multisensory amplitude reduction and latency shift will be relatively larger in OAs than in YAs. Whether or not differences between the groups are also observed on the earlier (P1) and later (P2) components will provide important information on the timing of AV processing and interaction in OAs.

Previous studies that concluded that the ability to integrate auditory and visual speech cues remains intact in OAs were based solely on behavioral findings. However, behavioral or performance measures such as response time represent the endpoint of a number of processes including those that are perceptual, evaluative, and motoric in nature. Recording ERPs can identify when multisensory interaction takes place in the processing stream of AV speech. Thus, we exploit the high temporal resolution of ERPs in addition to behavioral performance. That is, by identifying which of the various ERP components are modulated by AV speech, the current study will be able to pinpoint the point in the processing stream multisensory effects are evident and whether or not the stage of this effect is similar or different in YAs and OAs.

Method

Participants

Twenty YAs and 19 OAs were tested; however, three YAs and two OAs were excluded because of poor behavioral performance that differed from the group mean by more than two standard deviations) or because of noisy electrophysiological recordings. The final sample consisted of 34 individuals ($N = 17$ in each age group) who were in reported good health. Participants were screened for intact sensory abilities. We assessed visual contrast sensitivity using the MARS Letter Contrast Sensitivity test (Haymes et al., 2006), auditory acuity by measuring pure tone averages (PTA; average hearing threshold for frequencies of 500, 1,000, and 2,000 Hz), cognitive functioning using the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005); these data plus important demographic information are summarized in Table 1.

Table 1
Demographics for Younger and Older Adults

	Younger adults	Older adults	*t*-test
N	17 (12 women)	17 (12 women)	
Age	24.5 (3.43)	68.5 (5.03)	
Years of education	17.0 (1.8)	15.1 (2.9)	$t(32) = 2.4$; $p = .025$
MoCA	28.4 (1.6)	27.3 (1.8)	$t(32) = 1.8$; $p = .07$
PTA	6.2 (4.3)	12.7 (4.4)	$t(32) = 4.3$; $p < .001$
MARS contrast sensitivity	1.7 (.04)	1.6 (0.1)	$t(32) = 4.4$; $p < .001$
S/N in dB	55/68	55/66	$t(32) = 4.9$; $p < .001$

Note. Values are *M* (*SD*) if not indicated. MoCA = Montreal Cognitive Assessment; PTA = pure tone average (left and right ear); S/N = signal-to-noise ratio.

Although the OAs had lower sensory functioning than the YAs, both groups had age-appropriate and clinically normal contrast sensitivity scores (Haymes et al., 2006) and PTAs (ANSI, 1989). Only participants with a PTA below 20 dB and PTA differences between the left and right ear of 10dB or less were included. Participants gave informed consent and the study was approved by the Concordia University research ethics board.

Stimuli

The stimuli consisted of 80 spoken object names, 40 of which were natural objects (e.g., tree, pear, etc. . .) and 40 were artificial or manmade objects (e.g., bike, clock, etc. . .). The items in the two categories did not differ on various psycholinguistic factors such as number of syllables, artificial: $M = 1.21$ ($SD = 0.41$); natural: $M = 1.25$ ($SD = 0.44$); Thorndike-Lorge written word frequency, artificial: $M = 645.1$ ($SD = 802.4$); natural: $M = 454.0$ ($SD = 739.4$); Brown verbal word frequency, artificial: $M = 11.6$ ($SD = 14.8$); natural: $M = 7.3$ ($SD = 11.3$); and familiarity, artificial: $M = 558.3$ ($SD = 49.4$); natural: $M = 536.8$ ($SD = 52.7$; MRC Psycholinguistic Database: Machine Usable Dictionary. Version 2.00).

To present the stimuli, we videotaped a female speaker uttering the object names and subsequently edited the videos using Adobe Premiere to reveal only the head, face, and neck of the speaker. We added 13 still frames on average ($SD = 2$) as lead-in before the onset and 16 still frames ($SD = 2$) as lead out after the offset of the lip movements. The video images subtended a visual angle of 8.3° × 8.3° and were presented on a 16.1” Cathode Ray Tube (CRT) monitor. During recording, the sound files were digitized at 48 kHz and were equalized off-line on sound intensity using Adobe Audition and PRAAT (Boersma & Weenink, 2006). The average duration of each spoken word was 617 ms (range = 417 to 860 ms). Auditory stimuli were presented binaurally at 55dB SPL using EARLINK tube ear inserts (Neuroscan, El Paso, TX).

Here we describe three different presentation conditions, namely auditory-only (A-only), visual-only (V-only), and AV_{speech}.[2] In the AV_{speech} condition, participants watched the videoclip of the woman speaking a stimulus word and heard the woman at the same time. Stimuli for the other two conditions were derived from these AV_{speech} stimuli. That is, the V-only condition consisted of the same stimuli as the AV_{speech} trials, but with the audio track removed. Likewise, the A-only trials were the same stimuli as AV_{speech} trials, but with the video removed.

In all presentation conditions, participants were exposed to background noise that was played at the same time the stimuli were presented. This noise consisted of a multitalker babble mask adapted from the Speech Perception in Noise test, Revised (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). We modified the original eight-speaker babble track by overlaying this track three times slightly shifted in time to create a background babble mask that was less variable in its intensity fluctuations. It is important to note the intensity of the background babble was individually adjusted relative to the word signals for each participant to ensure an equivalent auditory perceptual load across the two age groups. To achieve the S/N adjustment, we played a list of object names not included in the experiment in an auditory-only format and asked participants to repeat the word. We then adjusted the intensity of the babble noise until the participant identified about 55% to 60% of the words correctly. The S/N ratio was slightly more favorable for OAs (Table 1) in order to achieve the same level of performance as the YAs. This step was crucial to equate the perceptual difficulty of the unisensory auditory condition for the YAs and OAs so that any AV benefit could be measured from the same baseline.

To measure visual and auditory ERPs elicited by each stimulus, we inserted triggers or markers at the onset of the lip movement and the onset of the sound, respectively, in all AV_{speech} stimuli. These triggers consisted of short transistor-transistor logic (TTL) pulses, which were not perceivable by participants but served to signal the recording amplifiers as to the onset of a stimulus. Because the AV_{speech} stimuli served as the basis for all other conditions, both trigger points were present in all three conditions. That is, the V-only condition included a trigger to mark the onset of the sound even though the sound was not audible to the participant. This was necessary to assess multisensory interaction effects at the same time point (see below).

Procedure

Participants were seated in a comfortable chair in a dimly lit room. Prior to the experimental task, we obtained sensory and cognitive performance scores and established the customized S/N ratio. The experiment consisted of a single block of 160 trials from each stimulus condition. Each word was presented twice in each condition, and the trial sequence was randomized throughout the block. Thus, from trial to trial, participants were presented with words from the A-only, V-only, and AV conditions, controlling for any blockwise attentional effects. Stimulus presentation was controlled by Inquisit 2.0 (2006) software. At the beginning of each trial, a fixation dot was presented in the centre of the monitor for 200 to 300 ms. For A-only trials the dot was replaced by a blank screen, and for trials involving visual information (i.e., V-only, AV_{speech}) the fixation dot was replaced with a sequence of still frames of the speaker’s face as lead-in ($M = 460$ ms, $SD = 55$ ms), after which the speaker’s lips started to move in the V-only and AV_{speech} conditions. In the AV_{speech} condition, the lip movement preceded the first auditory speech cue on average by about 432 ms and varied from 36 to 600 ms ($SD = 92$ ms), depending on the word. In the V-only trials, no auditory speech was presented. After the video had faded out, there was a 450-ms interstimulus interval to give participants a sufficiently long response time window. Participants were instructed to respond as quickly as possible as to whether the stimulus word named a natural or manmade object by pressing one of two keys on a standard keyboard (i.e., “*S*” and “*L*” keys) with the side of response assignment counterbalanced across participants. The stimulus onset asynchrony between the onset of the first video frame of consecutive stimuli was 4.5 s.

---

[2] The experiment also included a fourth condition labeled AV_{photo} during which participants heard the speaker while looking at a photograph of her. Its purpose was to determine whether the AV speech benefits can be achieved merely by the presence of a visual signal (i.e., a still face) that precedes auditory speech information or whether it is necessary to have dynamic and congruent lip movements. However, the ERP responses to this condition did not differ substantially from the unisensory conditions; thus, we do not report on this condition further.

EEG Data Acquisition

A continuous EEG was recorded using an elastic nylon cap containing 32 tin electrodes (Electro-Cap International, Inc., Eaton, OH) and arranged according to the International 10/20 system using a cephalic (forehead) location as ground and referenced on-line to the left ear. Six electrodes were aligned along the midline of the scalp running from anterior to posterior regions (Fz, FCz, Cz, CPz, Pz, Oz). Electrodes over the left/right hemispheres included the respective electrode sites FP1/2, F3/4, F7/8, FT7/8 (frontal), FC3/4, C3/4 (Fronto-central) and CP3/4, T7/8, P3/4, O1/2 (parieto-occipital).

All EEG data were re-referenced offline to linked ear lobes. The EEG signal was amplified using NeuroScan Synamps (Neuroscan, El Paso, TX) and was recorded at a sampling rate of 500 Hz in a DC to 100 Hz bandwidth with electrical impedances kept below 5 kΩ. The continuous EEG was divided into 700-ms epochs defined by the onset of each stimulus trigger and included a 100-ms prestimulus baseline interval. EEG was filtered offline for frequencies between 1 and 30 Hz. Horizontal and vertical electro-oculograms (HEOG and VEOG) were used to monitor eye movements and trials with HEOG activity exceeding +/− 50 μV were rejected. To assure a sufficient number of retained trials, excessive VEOG artefacts (i.e., eye blinks) were corrected using a spatial filter correction technique (Method 2, NeuroScan Edit 4.3 manual, 2003). Trials with EEG activity and other motion artefacts exceeding +/− 100μV were rejected. Furthermore, only trials with correct responses were included in our analyses. For a participant to be included in the analysis, a minimum of 70 accepted trials per presentation condition had to be retained.

As mentioned above, each stimulus contained two triggers, one to mark the onset of the lip movement and the other to mark the onset of the sound. This was the case even for A-only trials where no lip-movement was apparent and for V-only trials where no spoken word was audible. This was important to assess multisensory interactions. To do so, we compared the ERPs to the AV_{speech} trials triggered by the onset of the sound (i.e., when signals from both modalities were available) to the sum of the ERPs on unisensory trials (i.e., A + V). For this comparison to be valid, each of the triggers had to be aligned to the same point in time, namely the onset of the sound which was real in the case of AV_{speech} and A-only trials but virtual in the case of V-only (i.e., the trigger was at the point where the auditory signal would have begun had it been audible). This careful alignment of time points allowed us to accurately assess any non-linear interaction effects present in the AV_{speech} trials (van Wassenhove, et al., 2005). Having triggers placed at the onset of lip movement and at the onset of the sound, we were able to measure visual and auditory evoked potentials (AEPs), respectively. Onset of lip movement elicited clear visual evoked potentials over lateral occipito-temporal electrodes but because this study focused on auditory responses, visual evoked potentials are not discussed further.

As mentioned earlier, the electrophysiological response to an auditory stimulus typically consists of a series of early, sensory-driven and automatic ERPs referred to as P1-N1-P2 complex (Eggermont & Ponton, 2002; Näätänen & Picton, 1987). Component amplitudes and latencies were recorded at the components' peaks. P1 was scored as the maximum positive amplitude in the time windows 40 to 130 ms after stimulus onset, N1 was scored as the most negative amplitude in the time window 80 to 150 ms, and P2 as the most positive peak in the 160- to 260-ms time window. The accuracy of the detection algorithm was checked and when necessary (i.e., a peak just outside the defined time window) manually adjusted. The analyses of the AEPs were limited to central electrode sites around the vertex where auditory ERPs are most pronounced.

Results

All repeated-measures ANOVAs were adjusted with the Greenhouse-Geisser non-sphericity correction (Greenhouse & Geisser, 1959) for effects with more than one degree of freedom (*df*) in the numerator. According to convention, uncorrected degrees of freedom, the Greenhouse-Geisser epsilon (ε), mean square error (*MSE*), and adjusted p values are reported. Significant main effects and interactions were followed by analyses of simple effects and, unless stated otherwise, the differences reported are significant at $\alpha = .05$ or below.

Behavioral Results

Accuracy. Table 2 presents the accuracy results for YAs and OAs. In order to investigate an effect of age on accuracy, a 2 (Age Group; YAs & OAs) × 3 (Condition; A-only, V-only, AV_{speech}) repeated-measures ANOVA was conducted. The analysis revealed a main effect of condition, $F(2, 64) = 276.1$, $MSE = 52.6$, $\varepsilon = .67$, $p < .001$, in that responses to AV_{speech} trials were more accurate than responses to A-only which in turn were more accurate than responses to V-only trials. There was also an Age Group × Condition interaction, $F(2, 64) = 9.6$, $MSE = 52.6$, $p = .002$; pairwise comparisons showed that the V-only condition was driving this interaction. For that condition, OAs performed less well than YAs, indicating poorer lipreading ability. No group differences were found for A-only because of the fact that we successfully equated the groups on listening performance.[3] It is interesting to note that accuracy scores for AV_{speech} did not differ between groups, reflecting equivalent performance under multisensory conditions. No main effect of age group, $F(1, 32) = 1.98$, $MSE = 69.4$, $p = .17$, was evident. Finally, it is important to note that performance in both unisensory conditions was above the 50% chance level for YAs and OAs (i.e., there were no floor effects) and that AV_{speech} accuracy scores were significantly below the 100% ceiling value.

Response time. Reaction time data are presented in Table 2. To investigate an effect of age on reaction time (RT), a 2 (Age Group) × 3 (Condition; A-only, V-only, AV_{speech}) repeated-measures ANOVA was conducted, which revealed a main effect of condition, $F(2, 64) = 901.3$, $MSE = 10731.5$, $\varepsilon = .61$, $p < .001$. AV_{speech} trials resulted in the fastest responses, whereas RTs in V-only trials were slowest relative to the other conditions (Table

---

[3] The discrepancy between performance during the pre-experimental S/N ratio adjustment and the performance during experimental A-only trials is likely because of differences in task requirements. During the experiment participants were asked to perform a two-choice categorization task, whereas during the S/N ratio adjustment participants were asked to repeat the word presented.

Table 2
*Mean (*SD*) for Accuracy (% Correct) and Reaction Time (in ms) for A-Only, V-Only, and AV_{Speech} Conditions in Younger and Older Adults*

	Younger adults	Older adults
Accuracy		
A	81.3 (5.0)	84.1 (7.1)
V	65.9 (7.1)	56.6 (11.6)
AV_{speech}	94.7 (2.3)	94.3 (3.7)
RT		
A	1127.1 (112.4)	1207.1 (130.2)
V	1776.8 (153.4)	1903.9 (193.7)
AV_{speech}	1034.1 (113.4)	1118.4 (103.5)

Note. RT = response time; A = auditory only; V = visual only; AV_{speech} = audiovisual speech.

2). A main effect of age group, $F(1, 32) = 5.5$, $MSE = 43{,}943.6$, $p = .026$, indicated that OAs responded more slowly than YAs.

Race model analysis. One approach to assessing multisensory interaction is to evaluate whether response time distributions fit predictions of the race model which states that information streams are independent and that only the fastest channel yields a response; that is, the response to multisensory trials is not faster than the fastest of the unisensory responses (Miller, 1982).

The race model is said to be violated when the probability (p) of a particular response time is higher in the multisensory condition than the joint probability of the unisensory responses, i.e., $p(AV) > p(A + V) - p(A \times V)$, for that given response time. A violation suggests that that RT facilitation is because of the interaction of the two sensory channels (Miller, 1982). To test whether race model predictions are met or violated, the RT data were plotted as cumulative distribution functions (CDFs). We divided the RT interval from 0.4-s to 2.5-s into 10-ms bins and calculated the likelihood that a response occurred at a given response time or faster. The CDFs of OAs and YAs are plotted in Figure 1. These data were analyzed by conducting paired t-tests at each time bin to determine whether the observed AV_{speech} response time probabilities, $p(AV)$, were higher than the joint probability of the unisensory responses, $p(A + V) - p(A \times V$ (i.e., test of race model violation).

In the YAs, the CDF values for RTs to AV_{speech} trials were significantly larger ($p < .05$) than the CDF values of the joint probability of the unisensory responses for each time bin from 590 ms to 1240 ms. These data were remarkably similar to those of the OAs ($p < .05$; 600–1260 ms). Responses to AV_{speech} trials were faster than unisensory responses (A-only and V-only) and faster than the race model predictions which is shown in Figure 1 by the CDF of AV_{speech} response times shifted to the left relative to the other curves (Figure 1).

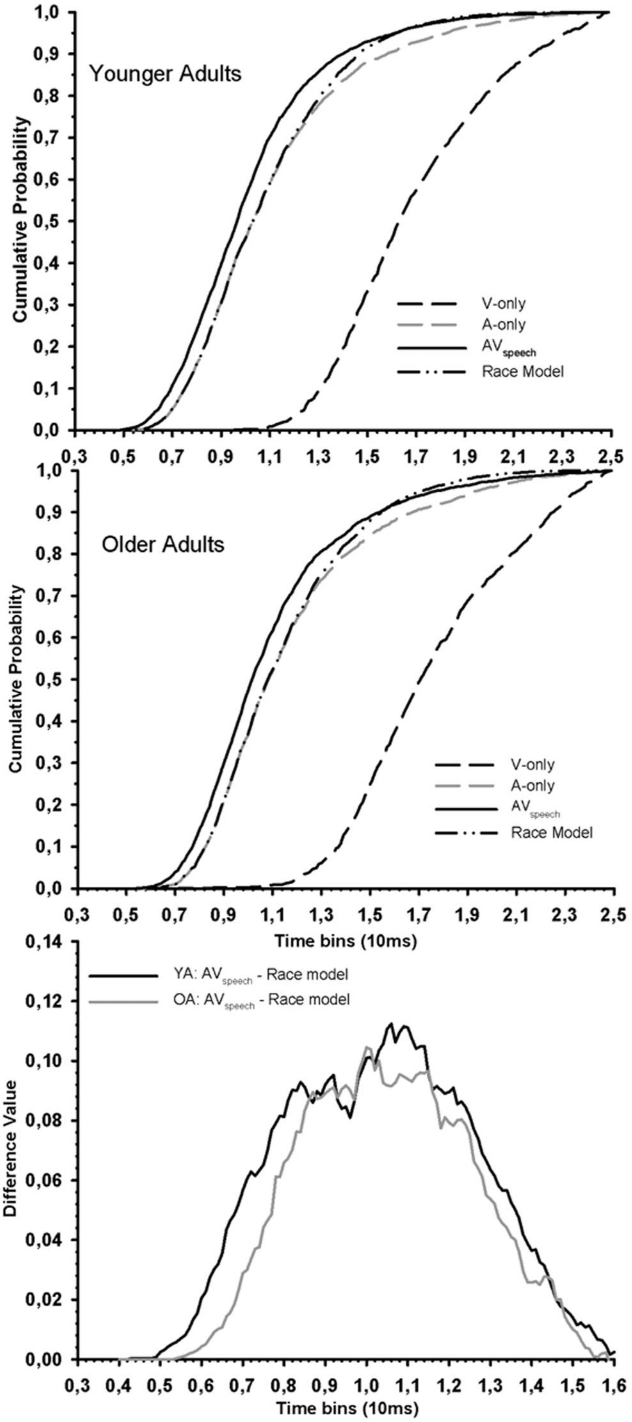

Figure 1. Cumulative distribution functions (CDFs) of reaction times obtained for younger adults (YAs; top) and older adults (OAs; middle) in the three presentation conditions: A = auditory only (gray, dashed), V = visual only (black, dashed), AV_{speech} (black, solid). The predicted CDF from the race model, $p(A + V) - p(A \times V)$, is presented in the black, dashed-dotted line. The bottom panel presents the difference values between the AV_{speech} condition and race model predictions for younger (black) and older adults (gray).

Auditory and visual enhancement. To examine the benefit derived from combining information from two modalities, we calculated the visual enhancement (VE), which reflects the amount of benefit gained from the additional visual speech cues, and auditory enhancement (AE), which is the amount of benefit gained from the additional auditory. These values were computed for accuracy data using a formula that equates for large variances across participants, $VE = (AV - A)/(100 - A)$; $AE = (AV - V)/(100 - V)$ (Sommers et al., 2005; Tye-Murray et al., 2008; Tye-Murray, Sommers, & Spehar, 2007). We were also interested in AE and VE for RT data but, in contrast to the accuracy data where there is a clearly defined maximum performance (i.e., 100% correct), we adjusted the formula slightly for RT, VE = (AV − A)/(A); AE = (AV − V)/(V). Separate one-way ANOVAs were conducted to investigate age differences for AE and VE. The results did not reveal any effects of age on enhancement for the accuracy data, AE: $F(1, 32) = .89$; $p = .35$; VE: $F(1, 32) = 1.5$; $p = .23$, or the RT data, AE: $F(1, 32) = .19$; $p = .66$; VE: $F(1, 32) = .45$; $p = .51$.

Electrophysiology of AEPs

Multisensory interaction for neural responses can be assessed by comparing the multisensory response to the arithmetic sum of the individual unisensory responses (Calvert, Hansen, Iversen, & Brammer, 2001). Significant deviations from this sum (i.e., either response enhancement or reduction) signify nonlinear interaction effects. We compared the ERP responses to the multisensory AV_{speech} condition to the sum of the responses elicited at the same time point in the two unisensory conditions A-only and V-only (i.e., A + V). As expected, V-only trials (i.e., visual speech without any auditory speech) did not elicit an AEP and are therefore not depicted in the figures.[4] Furthermore, because we focus on early auditory sensory processes reflected by the P1, N1, and P2, and AEPs are recorded maximally at the central midline electrode locations (Näätänen & Picton, 1987), we restricted our analyses to these electrode sites (FCz, Cz, CPz). Note that all figures depict group average waveforms at the central midline site Cz only, where the response is maximal.

To assess possible hemispheric differences, several repeated-measures ANOVAs were conducted with factors age (OAs and YAs), condition (A-only, A + V, AV_{speech}), hemisphere (left and right), and anteriority (3 sites from fronto-central to centro-parietal sites) on the P1, N1, and P2 latency and amplitude values. No main effects of hemisphere or interaction effects involving hemisphere were found.[5] Given that results from lateral sites did not yield additional information, subsequent ANOVAs included factors condition and site (3 midline sites: FCz, Cz, and CPz) and only the results from midline sites are reported here.

For each of the three auditory ERP components (i.e., P1, N1, P2) a separate ANOVA was conducted for the dependent variable of latency and amplitude, measured at the peak of the component of interest, with the factors age, condition, and site.[6]

P1 latency. Analysis of P1 latency ($M = 64.2$ ms, $SD = 16.4$) did not reveal a main effect of condition, $F(2, 64) = 2.5$, $MSE = 651.6$, $\varepsilon = .78$, $p = .11$; or age, $F(1, 32) = 1.7$, $MSE = 1147.6$, $p = .2$, nor an Age × Condition interaction, $F(2, 64) = .16$, $MSE = 651.6$, $p = .8$.

N1 latency. Analyses of N1 latency did not show a main effect of age, $F(1, 32 = 2.7$, $MSE = 1{,}696.7$, $p = .11$. However, there was a main effect of condition, $F(2, 64) = 40.6$, $MSE = 638.7$, $\varepsilon = .75$, $p < .001$, which showed that the N1 peaked significantly earlier during AV trials relative to A-only and to the summed A + V condition which did not differ from each other (Figures 3 and 4). An Age × Condition interaction, $F(2, 64) = 3.7$, $MSE = 638.7$, $p = .04$, indicated that, although the N1 AV latency shift from A-only and A + V trials to AV_{speech} trials was significant in both groups, it was larger in OAs (Figure 2). Planned

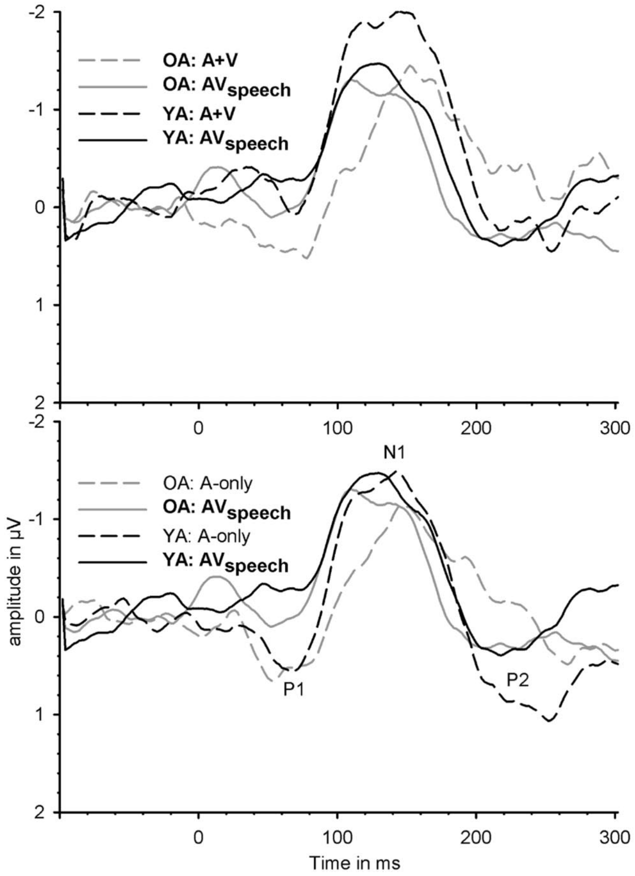

Figure 2. Group average waveforms of younger adults (YAs; black) and older adults (OAs; gray) at the Cz electrode site. Top panel: comparison of the unisensory trials (the sum of auditory and visual trials; A + V; dashed lines) and multisensory AV_{speech} trials (solid lines). Bottom panel: comparison of the unisensory auditory-only trials (A-only; dashed lines) and multisensory AV_{speech} trials. Note for the N1, the reduction in amplitude (i.e., it is less negative) and the earlier peak in the multisensory condition for both groups. Note also that the shift in N1 latency in the AV condition relative to the A-only condition was significantly larger in the OAs ($M = 33$ ms) compared with the YAs ($M = 19$ ms). The P1 was also significantly smaller in the OAs in the multisensory condition relative to A + V.

---

[4] As noted, the waveform for the V-only condition was computed by averaging the EEG traces that were time-locked to the temporal point of the onset of the auditory signal (which was not audible in the V condition). This allowed us to compare brain activity when information from both modalities was present (AV_{speech}) to the brain activity at the same point in time when information from only one modality was present (A or V). To be clear, we do not to say that there was no neural activity recorded at this time point in the V-only condition, otherwise the A + V waveform would be identical to the A-only waveform and, as is clear from the figures, this is not the case. Rather, there was no *auditory*-specific activity elicited in the V-only waveform.

[5] This is consistent with past research that finds no hemispheric differences in response to speech sounds (e.g., Grabow, Aronson, Rose, & Greene, 1980).

[6] None of the analyses reported below revealed a main or interaction effect involving the site factor.

comparisons showed that the N1 in response to A-only and A + V trials for OAs peaked significantly later than for YAs at central sites (A-only: 150 ms vs. 135 ms; A + V: 148 ms vs. 137 ms). It is interesting to note that the latency of the auditory N1 on AV_{speech} trials did not differ between age groups (120 ms for both), indicating that the younger and older groups showed equally early auditory responses in the multisensory condition.

P2 latency. The P2 latency analysis revealed a main effect of condition, $F(2, 64) = 3.6$, $MSE = 1242.3$, $\varepsilon = .88$, $p = .039$, with the P2 peaking earlier during the AV_{speech} condition relative to the A-only condition. No main effect of Age for P2 latency was evident, $F(1, 32) = .25$, $MSE = 3280.6$, $p = .62$.

P1 amplitude. Analysis of the P1 amplitude revealed a main effect of condition, $F(2, 64) = 17.7$, $MSE = .9$, $\varepsilon = .99$, $p < .001$, with responses to AV_{speech} trials being smaller than responses to the summed A + V and the A-only trials. Moreover, a main effect of age, $F(1, 32) = 6.4$, $MSE = 2.3$, $p = .016$, was observed with P1 amplitudes being larger for OAs than for YAs. An Age × Condition interaction, $F(2, 64) = 3.7$, $MSE = .9$, $p = .03$, indicated that the responses to AV_{speech} trials were significantly smaller than responses to the summed A + V trials (OAs, $p < .001$; YAs, $p = .054$, just missing the conventional level of significance). The interaction was attributable to the fact that the amplitude reduction was larger in the OAs (mean difference = .92 μV) than in the YAs (mean difference = .37 μV; see Figures 3 and 4 for ERP responses from YAs and OAs, respectively).

N1 amplitude. Analysis of the N1 amplitude revealed a main effect of condition, $F(2, 64) = 19.7$, $MSE = 1.7$, $\varepsilon = .7$, $p = .001$, with the summed response of A + V being significantly larger than responses to AV_{speech} and A-only trials. No main effect of age, $F(1, 32) = 3.5$, $MSE = 4.2$, $p = .7$, nor an Age × Condition interaction, $F(2, 64) = 1.5$, $MSE = 1.7$, $p = .23$, were obtained, indicating that the N1 reduction to AV_{speech} trials held for both groups.

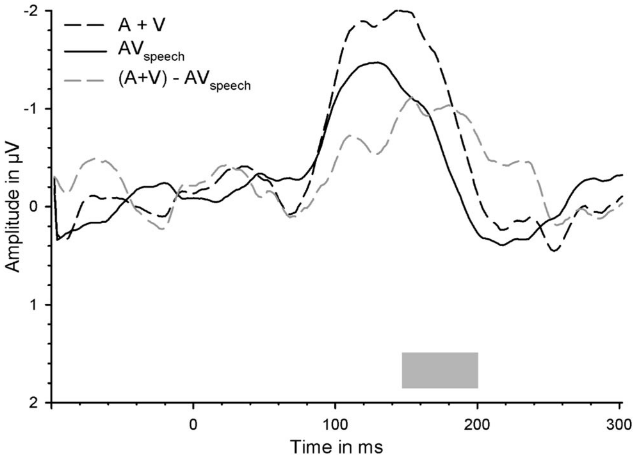

Figure 3. Group average waveforms of younger adults at the Cz electrode site depicting the reduction in P1 amplitude (i.e., it is less positive), N1 amplitude (i.e., it is less negative) and earlier N1 peak in the multisensory AV_{speech} condition (black, solid line) compared with the sum of the unisensory conditions (A + V; black, dashed line). The dashed, grey line displays the difference waveform $(A + V) - AV_{speech}$. The gray shading indicates the time interval for which the AV_{speech} waveform differed significantly from the summed A + V waveform ($p < .05$).

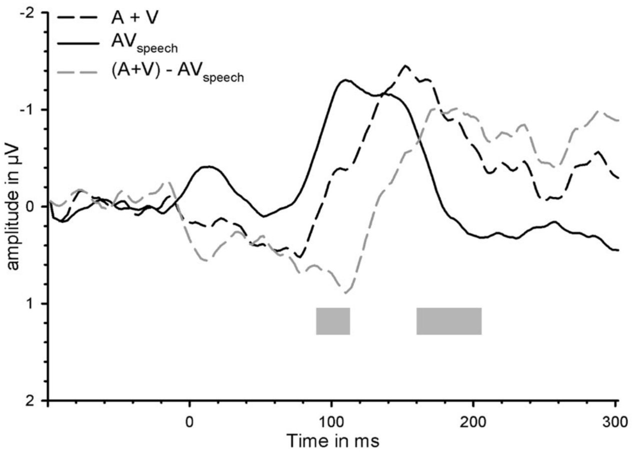

Figure 4. Group average waveforms of older adults at the Cz electrode site depicting the reduction in P1 amplitude (i.e., it is less positive), N1 amplitude (i.e., it is less negative) and earlier N1 peak in the multisensory AV_{speech} condition (black, solid line) compared with the sum of the unisensory conditions (A + V; black, dashed line). The dashed, grey line displays the difference waveform $(A + V) - AV_{speech}$. The gray shading indicates the time interval for which the AV_{speech} waveform differed significantly from the summed A + V waveform ($p < .05$).

P2 amplitude. Analysis of the P2 amplitude did not reveal a main effect of condition, $F(2, 64) = 1.02$, $MSE = 1.7$, $\varepsilon = .74$, $p = .37$, or main effect of age, $F(1, 32) = 1.6$, $MSE = 7.3$, $p = .2$, nor an Age × Condition interaction, $F(2, 64) = 1.3$, $MSE = 1.7$, $p = .28$.

Time point of multisensory interaction. Our analyses indicated that AV speech led to multisensory interaction at the level of early sensory processes as indexed by the auditory P1-N1-P2 complex. To assess the time point of multisensory interaction more closely, we computed ERP difference waveforms by subtracting the responses to AV speech trials from the summed response of A + V trials. We then conducted a *t*-test at each time point from 0 to 300 ms after stimulus onset (i.e., 150 time points) and applied the most conservative criterion for significance proposed by Guthrie and Buchwald (1991), namely a minimum of 12 consecutive *t* values larger than the critical value of 2.14 at the Cz electrode site, which is where AEPs are most prominent. OAs revealed significant differences from 88 to 114 ms after stimulus onset which is around the time period of the P1-N1 complex and from 160 to 208 ms corresponding to the N1-P2 ERP complex (Figure 4). For the YAs, significant differences between AV_{speech} and A + V emerged only for the later time window, namely for the N1-P2, 142 to 198 ms after stimulus onset (Figure 3).

The Role of Sensory Functioning

Predicated on the inverse effectiveness hypothesis and its predictions related to sensory effectiveness, we examined the relationship between basic sensory functioning (i.e., visual contrast sensitivity and auditory PTA thresholds) and our dependent variables. Contrast sensitivity correlated only with accuracy performance on AV_{speech} trials, $r(32) = .36$, $p = .037$. This relationship suggests that higher contrast sensitivity led to better AV_{speech} perception but interestingly not to better lipreading performance

(V-only) per se. However, a standard multiple regression with AV_{speech} accuracy as the dependent variable and age, cognitive functioning, hearing level, and contrast sensitivity as independent variables did not reach significance. A second multiple regression analysis was conducted between the N1 latency shift from A-only to AV trials as dependent variable and age, contrast sensitivity, cognitive functioning, and PTAs as independent variables. Table 2 shows the results of the analysis, including the bivariate correlations between the independent variables and the dependent variable, the unstandardized regression coefficients (*B*), the standardized regression coefficients (β), the squared semipartial correlations (sr^2), the intercept, *R* and R^2 (Tabachnik & Fidell, 2001). This regression revealed that hearing level was the only significant predictor of the size of the auditory N1 latency reduction, predicting almost 20% of the variance in N1 latency shift (Table 3). Figure 5 shows that higher hearing thresholds (i.e., poorer auditory functioning) led to a greater reduction in N1 latency on AV trials compared with A-only trials.

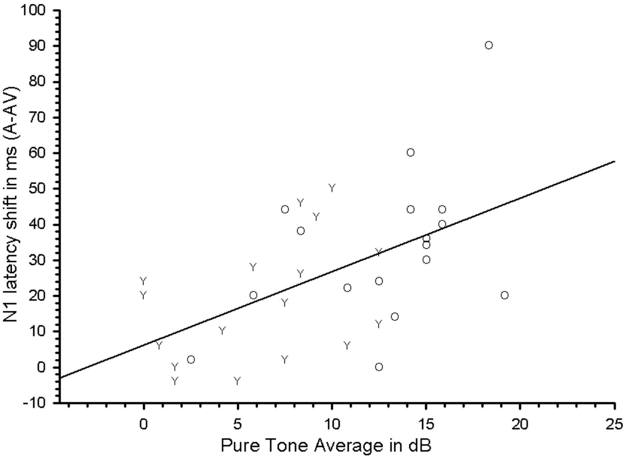

Figure 5. Regression of auditory functioning as measured by listening thresholds (PTA = pure tone average) on the shift in the auditory N1 latency from A-only to AV_{speech} trials (A − AV). O = older adults; Y = younger adults. Regression equation: N1 latency shift = .54 * PTA + 57.4 ms.

Discussion

This study is the first to investigate the behavioral benefit of and the neural processes associated with AV speech perception of spoken words in an ecologically realistic, noisy listening environment. More importantly, this study examined age differences in the ability to integrate auditory and visual speech cues and the underlying neural processes. Before addressing differences between OAs and YAs with regards to audiovisual speech processing, it is important to discuss any age differences in unisensory performance. Note that although all participants had clinically normal sensory function, the OAs performed more poorly on our measures of unisensory processing than did the YAs. Recall that in order to equate each individual participant on auditory perceptual load, the signal-to-noise ratio was titrated to achieve, on average, 80% response accuracy for A-only in both groups. This was important in order to estimate the amount of benefit derived from the additional visual speech cues in the AV_{speech} condition compared with an equivalent A-only condition. A more moderate S/N ratio was required to achieve this performance in OAs than YAs, suggesting that auditory functioning was decreased in this group. Significant age effects were also observed in the V-only condition (i.e., speechreading) during which OAs performed significantly poorer than YAs. Last, OAs responded more slowly on the categorization task, a finding consistent with commonly observed age-related slowing.

Table 3
Regression on N1 Latency Shift From A-Only to AV_{Speech} Trials

Variable	*R* with N1 latency shift	*B*	β	sr^2 (unique)
Age	.34	−.07	−.08	.002
CS	−.22	−1.28	−.01	.00002
MoCA	−.18	−1.69	−.14	.02
PTA	.54	2.19	.57*	.18
	Intercept = 57.4			
	R^2 = .31			
	Adjusted R^2 = .22			
	R = .56			

Note. CS = contrast sensitivity; MoCA = Montreal Cognitive Assessment; PTA = pure tone average.
* $p < .01$.

For audiovisual processing, the behavioral data clearly showed that the availability of AV speech cues led to superior performance (i.e., higher accuracy and faster response times) in both age groups compared with unisensory speech perception (i.e., only listening or only lipreading). This is in keeping with the benefit that has been shown repeatedly in studies presenting simple syllables (Besle et al., 2004; Cienkowski & Carney, 2002; Sumby & Pollack, 1954), as well as words or even sentences (Sommers et al., 2005).

Analysis of the reaction time data revealed violations of the race model indicating that the faster responses during AV_{speech} trials were likely due to an interaction of the two unisensory information channels and not simply the result of two redundant signals (Miller, 1982). The response time interval during which the race model was violated did not differ between YAs and OAs. Taken together, the behavioral findings showed that the ability to integrate auditory and visual speech cues remained intact in OAs, supporting previous findings (Bucur et al., 2005; Cienkowski & Carney, 2002; Hugenschmidt et al., 2009; Laurienti et al., 2006; Sommers et al., 2005; Thompson, 1995; Thompson & Malloy, 2004).

Our OAs exhibited poorer sensory functioning than the YAs and could be considered to be in a ‘permanently’ suboptimal environment. Considering results from previous studies we expected OAs to benefit relatively more from AV speech (Hugenschmidt et al., 2009; Laurienti et al., 2006). The underlying idea was that the availability of congruent AV speech cues would provide a way to compensate for decline in unisensory functioning. The auditory and visual enhancement scores did not support this hypothesis, because OAs and YAs showed equal multisensory benefits. It is interesting to note that the RTs for OAs on AV trials were as fast as the RTs for YAs on A-only trials. Moreover, the cumulative distribution functions of the RTs of OAs during the AV_{speech} condition overlapped with those of the A-only condition for YAs. Therefore, the behavioral AV_{speech} performance of the OAs approached the level of the unisensory performance in the YAs, a finding that has also been shown by Laurienti and colleagues (2006). In other words, the additional visual speech cues made the responses of OAs to speech tokens "younger."

However, striking evidence for the Inverse Effectiveness Hypothesis was found in the ERP data, with OAs showing more pronounced facilitations of neural responses on AV speech trials relative to YAs. We focused our analyses on early sensory ERP responses of the auditory system, namely the P1, N1, and P2. Both age groups showed an amplitude reduction of the auditory N1 in response to AV_{speech} trials relative to the summed response of the two unisensory conditions, $A + V$. This finding corresponds to previous studies on AV speech processing in YAs (Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove, et al., 2005) and, importantly, extends it to OAs. Näätänen and Picton (1987) have shown that N1 amplitude becomes smaller if the auditory stimulus is predictable. In the context of AV speech, van Wassenhove and colleagues (2005) proposed that the N1 amplitude reduction is because of the increased predictability of the auditory speech sound given that the visual speech cues precede the auditory signal.

The N1 amplitude reduction in the present study reflects multisensory interaction in the form of a response reduction in the AV condition compared with the sum of the unisensory responses (i.e., $AV < A + V$) and, based on previous research, suggests that visual information interacted with auditory cues at the level of the auditory cortex (Besle et al., 2004; Campbell, 2008; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). It is interesting to note that the size of the amplitude reduction from $A + V$ to AV trials did not differ statistically between YAs and OAs, suggesting that the neural processes underlying AV speech processing were intact in OAs.[7] This finding is in line with our behavioral data.

Similarly, our auditory P1 amplitude results indicate that both groups show a P1 amplitude reduction in AV trials relative to the $A + V$ response. It is interesting to note that this reduction was larger in OAs than in YAs and indicates that P1 multisensory interaction effects are greater in OAs. Less is known about the precise perceptual processes reflected by the auditory P1 (also referred to as the P50), but it is driven by physical characteristics of an auditory stimulus and is related to aspects of sensory gating (e.g., Grunwald et al., 2003). Some findings locate the source of the P1 in thalamic nuclei (e.g., Erwin & Buchwald, 1987) or peri-Sylvian auditory cortex (Grunwald et al., 2003), suggesting that it is elicited very early in the auditory perceptual processing stream. Further analyses of our data confirmed that multisensory interaction is more pronounced at an earlier time point in OAs than in YAs. The YAs showed multisensory interaction effects in the time range of the P1 and also from 142 to 198 ms after stimulus onset (i.e., in the latency range of the N1-P2 complex). However, the multisensory interaction in the OAs in the early P1 time window (88 to 114 ms) were even more pronounced (i.e., the amplitude reduction in the OAs was bigger than that of the YAs). Together, the amplitude and latency data suggest that the neural processes underlying AV speech perception in OAs are facilitated earlier and to a greater extent than in YAs.

In addition to the amplitude reduction, both groups exhibited a significant latency shift, with the multisensory N1 response peaking earlier than the unisensory A-only and the summed $A + V$ responses. This is in line with previous findings (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). It is interesting to note that this facilitation of auditory processing speed was larger in OAs (33 ms) than in YAs (19 ms). According to van Wassenhove and colleagues (2005), N1 latency shifts to AV stimuli depend on the degree of predictability of the visual speech cue. With this in mind, our findings suggest that OAs were more apt than YAs in extracting useful information from visual speech cues to predict or supplement the upcoming spoken utterance. Alternatively, it is possible that the pronounced latency shift of the N1 and the amplitude reduction of the preceding P1 in the OAs could be because of heightened attention to visual speech cues (Thompson & Malloy, 2004). However, the majority of research has shown that stimuli that receive more attention serve to *increase* the amplitude of P1-N1 responses (e.g., Hansen & Hillyard, 1980; Näätänen, 1982), whereas we observed a P1-N1 amplitude *decrease* for AV trials. Therefore, while it is quite possible that attention serves to enhance performance at some point in the processing stream, we believe that the present early ERP results are more likely due to the presence of visual speech cues serving to increase the predictability or facilitate the processing of the auditory signal, rather than being because of attention effects *per se*.

Our results suggest that OAs, compared with YAs, are not better lipreaders per se but rather are better "lip/speech integrators." One explanation for this could be impoverished auditory functioning. The hearing thresholds, although clinically normal, were higher in the OAs than YAs. It is interesting to note that hearing level predicted the size of N1 latency shifts from A-only to AV trials in all participants, regardless of age. In other words, participants with poorer auditory functioning exhibited a more pronounced speeding of auditory processing at the neural level when visual speech cues were available. Our interpretation is that individuals with less optimal hearing compensate for diminished auditory function by making more efficient use of visual speech cues. The idea that OAs rely to a larger extent on additional visual speech cues is supported by other studies on AV speech perception in OAs (Cienkowski & Carney, 2002; Thompson & Malloy, 2004). Future research might also investigate whether the magnitude of the AV speech benefit in OAs varies as a function of stimulus properties, such as the identity of the speaker (e.g., older versus younger speakers) or the degree to which the visual speech cue predicts the upcoming auditory signal.

What might be the functional significance of our findings? Both the RT and accuracy findings revealed AV speech benefits in conjunction with electrophysiological results that showed an amplitude reduction of the auditory N1 in response to AV speech trials. This indicates that fewer neural resources were expended to achieve better performance, suggesting that AV_{speech} was processed more efficiently than auditory or visual speech alone in

---

[7] It is possible that the lack of an age difference on the N1 amplitude could be because of the fact that we adjusted the auditory S/N ratio of the older adults to match that of the younger adults. This was important to to equate the auditory perceptual load for both groups, bringing the perceived intensity in both age groups to the same level. This allowed us to evaluate the multisensory gain in the AV condition by comparing the two groups on the unisensory auditory baseline line. Regardless, we note that age differences in N1 amplitude are only variably reported in the literature (e.g., Anderer, Semlitsch, & Saletu, 1996).

both YAs and OAs. The idea of efficiency is intriguing as it leads to some interesting implications. Assuming that the brain has only a finite amount of neural resources available to perform both sensory and cognitive processing (e.g., Just & Carpenter, 1992; Rabbitt, 1968), efficiency in processing is crucial. Speech perception in noisy environments is more effortful for OAs (CHABA [Committee on Hearing and Bioacoustics], 1988; Pichora-Fuller, Schneider, & Daneman, 1995; Schneider & Pichora-Fuller, 2000). If signal processing is effortful, more processing resources have to be devoted to perceptual processing. This, in turn, leads to fewer resources available for higher level processing such as working memory (WM). Research has shown that WM performance declines with age in general (Park et al., 2002; Wingfield & Tun, 2001) and especially for auditory stimuli presented in background noise (Pichora-Fuller et al., 1995; Schneider & Pichora-Fuller, 2000). If $\mathrm{AV}_{\mathrm{speech}}$ signals make speech processing more efficient at the sensory level, which was demonstrated in the current study, resources that are not used could be recruited to improve higher level processes such as WM (Just & Carpenter, 1992; Schneider & Pichora-Fuller, 2000). Current research in our laboratory indicates that this is indeed the case (Frtusova, Winneke, & Phillips, 2010).

Conclusion

This study demonstrated that AV speech perception remains intact in older age and facilitates speech perception in a noisy environment. Despite lower lip reading performance, OAs performed as well as YAs during AV speech trials. It is interesting to note that despite a similar pattern in behavioral measures, the electrical brain responses indicated that AV speech resulted in earlier multisensory interaction effects (around 65 ms) and relatively larger N1 latency shifts in OAs. The brain responses indicated that OAs used visual speech cues more effectively to improve auditory speech processing in the presence of background noise. One explanation for this age-related benefit is that the availability of visual speech cues compensated for less-than-optimal auditory processing. That is, the additional visual speech cues allowed the ears of older adults to function "younger." Furthermore, we found indications of more efficient auditory processing in AV conditions in the young and older participants which should have important implications for the relationship of sensory resource allocation and cognitive functioning.

References

Anderer, P., Semlitsch, H. V., & Saletu, B. (1996). Multichannel auditory event-related brain potentials: Effects of normal aging on scalp distribution of N1, P2, N2, and P300 latencies and amplitudes. *Electrical and Clinical Neurophysiology, 99,* 458–472. doi:10.1016/S0013-4694(96)96518-9

ANSI. (1989). *ANSI S3.6–1989: American national standard specification for audiometers*. New York: American National Standards Institute.

Besle, J., Fort, A., Depuelch, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20,* 2225–2234. doi:10.1111/j.1460-9568.2004.03670.x

Bilger, R. C., Nuetzel, M. J., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32–48.

Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.5.04). Retrieved from http://www.praat.org

Bucur, B., Allen, P. A., Sanders, R. E., Ruthruff, E., & Murphy, M. D. (2005). Redundancy gain and coactivation in bimodal detection: Evidence for the preservation of coactive processing in older adults. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 60,* 279–282.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14,* 427–438. doi:10.1006/nimg.2001.0812

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 363*(1493), 1001–1010. doi:10.1098/rstb.2007.2155

CHABA (Committee on Hearing and Bioacoustics). (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83,* 859–895. doi:10.1121/1.395965

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear & Hearing, 23,* 439–449. doi:10.1097/00003446-200210000-00006

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 33,* 497–505. doi:10.1080/14640748108400805

Eggermont, J., & Ponton, C. (2002). The neurophysiology of auditory perception: From single units to evoked potentials. *Audiology and Neuro-otology, 7,* 71–99. doi:10.1159/000057656

Erber, N. P. (2002). *Hearing, vision, communication and older people.* Clifton Hill, Victoria, Australia: Clavis Publishing.

Erwin, R. J., & Buchwald, J. S. (1987). Midlatency auditory evoked responses in the human and the cat model. *Electroencephalography and Clinical Neurophysiology, Supplement, 40,* 461–467.

Frtusova, J., Winneke, A., & Phillips, N. (2010, April). *The effect of audiovisual speech information on working memory in younger and older adults*. Poster presented at the Cognitive Aging Conference, Atlanta, Georgia.

Government of Canada. (2002). *Canada's aging population.* Minister of Public Works and Government Services Canada. Retrieved from http://www.hc-sc.gc.ca/seniors-aines

Grabow, J. D., Aronson, A. E., Rose, D. E., & Greene, K. L. (1980). Summated potentials evoked by speech sounds for determining cerebral dominance for language. *Electroencephalography & Clinical Neurophysiology, 49,* 38–47. doi:10.1016/0013-4694(80)90350-8

Grant, K. W. (2002). Measures of auditory-visual integration for speech understanding: A theoretical perspective. *Journal of the Acoustical Society of America, 112,* 30–33. doi:10.1121/1.1482076

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(24), 95–112. doi:10.1007/BF02289823

Grunwald, T., Boutros, N. N., Pezer, N., von, O. J., Fernandez, G., Schaller, C., & Elger, C. E. (2003). Neuronal substrates of sensory gating within the human brain. *Biological Psychiatry, 53,* 511–519. doi:10.1016/S0006-3223(02)01673-6

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28,* 240–244. doi: 10.1111/j.1469-8986.1991.tb00417.x

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography & Clinical Neurophysiology, 49,* 277–290. doi:10.1016/0013-4694(80)90222-9

Haymes, S. A., Roberts, K. F., Cruess, A. F., Nicolela, M. T., LeBlanc, R. P., Ramsey, M. S., . . . Artes, P. H. (2006). The letter contrast sensitivity test: Clinical evaluation of a new design. *Investigative Ophthalmology & Visual Science, 47,* 2739–2745. doi:10.1167/iovs.05-1419

Hugenschmidt, C. E., Mozolic, J. L., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport, 20,* 349–353. doi:10.1097/WNR.0b013e328323ab07

Inquisit 2.0. (2006). Inquisit 2.0 [computer software]. Seattle, WA: Millisecond Software.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. doi:10.1037/0033-295X.99.1.122

Kim, S. H., Frisina, R. D., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2006). Effect of age on binaural speech intelligibility in normal hearing adults. *Speech Communication, 48,* 591–597. doi:10.1016/j.specom.2005.09.004

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27,* 1155–1163. doi:10.1016/j.neurobiolaging.2005.05.024

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748. doi:10.1038/264746a0

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279. doi:10.1016/0010-0285(82)90010-X

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. *Psychological Bulletin, 92,* 605–640. doi:10.1037/0033-2909.92.3.605

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of component structure. *Psychophysiology, 24,* 375–425. doi:10.1111/j.1469-8986.1987.tb00311.x

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society, 53,* 695–699. doi:10.1111/j.1532-5415.2005.53221.x

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17,* 299–320. doi:10.1037/0882-7974.17.2.299

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*(1), 593–608. doi:10.1121/1.412282

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language and Hearing Research, 52,* 1073–1081. doi:10.1044/1092-4388(2009/07-0276)

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20,* 241–248. doi:10.1080/14640746808400158

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17,* 1147–1153. doi:10.1093/cercor/bhl024

Rugg, M. D., & Coles, M. G. H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 27–39). Oxford, United Kingdom: Oxford University Press.

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 155–219). Mahwah, NJ: Lawrence Erlbaum Associates.

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 & Hearing, 26,* 263–275. doi:10.1097/00003446-200506000-00003

Statistics Canada. (2005). Population projections for Canada. Retrieved from http://www.statcan.gc.ca/daily-quotidien/051215/dq051215b-eng.htm

Stein, B. E., & Meredith, M. A. (1993). *Merging of the senses.* Cambridge, MA: MIT Press.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19,* 1–10. doi:10.1162/jocn.2007.19.12.1964

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215. doi:10.1121/1.1907309

Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging, 10,* 215–228. doi:10.1037/0882-7974.10.2.215

Thompson, L. A., & Malloy, D. (2004). Attention resources and visible speech encoding in older and younger adults. *Experimental Aging Research, 30,* 241–252. doi:10.1080/03610730490447877

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology, 47*(Suppl. 2), S31–37. doi:10.1080/14992020802301662

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear & Hearing, 28,* 656–668. doi:10.1097/AUD.0b013e31812f7185

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 Science, 102*(4), 1181–1186. doi:10.1073/pnas.0408949102

Wingfield, A., & Tun, P. A. (2001). Spoken language comprehension in older adults: Interactions between sensory and cognitive change in normal aging. *Seminars in Hearing, 22,* 287–301. doi:10.1055/s-2001-15632

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science, 14,* 144–148. doi:10.1111/j.0963-7214.2005.00356.x

Received December 18, 2009
Revision received September 20, 2010
Accepted September 23, 2010 ■