Psychology and Aging

ERP Evidence That Auditory–Visual Speech Facilitates Working Memory in Younger and Older Adults

Jana B. Frtusova, Axel H. Winneke, and Natalie A. Phillips

Online First Publication, February 18, 2013. doi: 10.1037/a0031243

CITATION

Frtusova, J. B., Winneke, A. H., & Phillips, N. A. (2013, February 18). ERP Evidence That Auditory–Visual Speech Facilitates Working Memory in Younger and Older Adults. *Psychology and Aging*. Advance online publication. doi: 10.1037/a0031243

Psychology and Aging

0882-7974/13/$12.00 DOI: 10.1037/a0031243

ERP Evidence That Auditory–Visual Speech Facilitates Working Memory in Younger and Older Adults

Jana B. Frtusova
Concordia University

Axel H. Winneke
Jacobs University Bremen

Natalie A. Phillips
Concordia University

Auditory–visual (AV) speech enhances speech perception and facilitates auditory processing, as measured by event-related brain potentials (ERPs). Considering a perspective of shared resources between perceptual and cognitive processes, facilitated speech perception may render more resources available for higher-order functions. This study examined whether AV speech facilitation leads to better working memory (WM) performance in 23 younger and 20 older adults. Participants completed an ***n***-back task (0- to 3-back) under visual-only (V-only), auditory-only (A-only), and AV conditions. The results showed faster responses across all memory loads and improved accuracy in the most demanding conditions (2- and 3-back) during AV compared with unisensory conditions. Older adults benefited from the AV presentation to the same extent as younger adults. WM performance of older adults during the AV presentation did not differ from that of younger adults in the A-only condition, suggesting that an AV presentation can help to counteract some of the age-related WM decline. The ERPs showed a decrease in the auditory N1 amplitude during the AV compared with A-only presentation in older adults, suggesting that the facilitation of perceptual processing becomes especially beneficial with aging. Additionally, the N1 occurred earlier in the AV than in the A-only condition for both age groups. These AV-induced modulations of auditory processing correlated with improvement in certain behavioral and ERP measures of WM. These results support an integrated model between perception and cognition, and suggest that processing speech under AV conditions enhances WM performance of both younger and older adults.

Keywords: aging, speech perception, multisensory interaction, working memory, even-related potentials

It has been argued that cognitive and perceptual processing share overlapping resources (Schneider & Pichora-Fuller, 2000). Thus, the amount of resources that we need to allocate to perceptual processing can affect higher order functions such as working memory (WM) and vice versa. This becomes especially relevant in perceptually or cognitively demanding situations, where we may not have enough resources to perform effectively in both domains. For example, degraded sensory information requires more effortful perceptual processing, which may deplete resources necessary for further higher-order processing of information (Rabbitt, 1968). On the other hand, facilitation of perceptual processing could lead to enhancement of higher-order functions. This question is intriguing, given the argument that our pool of processing resources becomes more restricted due to age-related changes in sensory and cognitive systems (Hasher & Zacks, 1988; Salthouse, 1988). In the speech domain, it has long been known that the availability of both auditory and visual speech cues, which occurs during face-to-face interactions, leads to facilitation of auditory speech perception (Sumby & Pollack, 1954). The current study tests the idea that auditory–visual (AV) speech facilitation leads to better WM performance and examines possible age differences in this effect.

There are a number of models of WM (see Miyake & Shah, 1999). One of the most enduring is that of Baddeley and Hitch (1974), which proposes two modality-specific "slave" systems responsible for temporarily storing information (i.e., "phonological loop" and "visuospatial sketch pad"), plus a "central executive" responsible for manipulating the temporary information to accomplish the current task at hand. WM performance depends on the processes of all three components. Therefore, facilitating percep-

---

Jana B. Frtusova and Natalie A. Phillips, Department of Psychology/Centre for Research in Human Development, Concordia University, Montreal, Quebec, Canada; Axel H. Winneke, Jacobs Center on Lifelong Learning & Institutional Development, Jacobs University Bremen, Bremen, Germany.

The study was supported by grants awarded to Natalie A. Phillips from the Canadian Institutes of Health Research (Grant MOP-97808) and the Concordia University VPRGS Seed/Accelerator Funding Program (2009). We thank the Natural Sciences and Engineering Research Council of Canada for the Alexander Graham Bell Canada Graduate Scholarship, and Concordia University for the Special Entrance Award awarded to Jana B. Frtusova. We would also like to express gratitude to the Centre for Research in Human Development for the financial and technical support to the Cognitive Psychophysiology Laboratory where the research was conducted. Finally, we thank the members of the Cognitive Neurophysiology Lab for their contribution of time during piloting and testing, especially Dr. Shanna Kousaie.

Earlier reports of these data were previously presented at the Cognitive Aging Conference, Atlanta, Georgia, April 15–18, 2010, and the 18th Annual Cognitive Neuroscience Society Meeting, San Francisco, California, April 2–5, 2011.

Correspondence concerning this article should be addressed to Natalie A. Phillips, 7141 Sherbrooke Street, West, Montreal, Quebec, H4B 1R6, Canada. E-mail: Natalie.Phillips@concordia.ca

tual processes should impact modality-specific stores and consequently improve WM processing.

We first present a brief overview of how an AV presentation affects speech perception. This is followed by a short review of age-related changes in perceptual abilities important for speech perception and age-related changes in WM. Lastly, we will present existing knowledge on how AV speech perception influences WM.

AV Speech Perception

We are often required to comprehend speech under noisy conditions, such as when having a conversation on busy streets. Research has demonstrated that adding visual speech information to auditory speech information can substantially enhance speech perception, especially in suboptimal listening conditions (e.g., Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954; Sommers, Tye-Murray, & Spehar, 2005). One explanation for this AV enhancement is that visual cues provide complementary information that helps to resolve the ambiguity of acoustically similar phonemes (Summerfield, 1987). Importantly, behavioral improvements during AV speech are accompanied by reduced (e.g., Besle et al., 2008; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2005; Winneke & Phillips, 2011) and earlier electrical brain responses (e.g., van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007; Winneke & Phillips, 2011), as measured by auditory event-related brain potentials (ERPs). More specifically, the auditory N1, which is an ERP component related to the detection of auditory stimuli and the encoding of auditory stimulus properties (Eggermont & Ponton, 2002; Näätänen & Picton, 1987), is sensitive to multisensory interactions such that it occurs earlier and is smaller in amplitude during bimodal (AV) compared with unimodal (auditory-only [A-only]) speech presentation.[1] The fact that the brain activity is faster and is reduced in amplitude, and that performance is enhanced, suggests that the brain uses resources more efficiently when processing AV speech information. It has also been argued that visual cues facilitate auditory processing (i.e., N1 latency shifts) by helping to predict the nature (van Wassenhove et al., 2005) or timing (Stekelenburg & Vroomen, 2007) of the subsequent auditory input. Given the assumption of shared resources between perceptual and cognitive processes (Schneider & Pichora-Fuller, 2000), the facilitation provided by bimodal, as opposed to unimodal, information could have positive effects on subsequent higher order functions, such as WM.

AV Speech Perception and Aging

Many sensory and cognitive abilities decline with age and can yield deficits in speech perception (Committee on Hearing and Bioacoustics, 1988). According to Grant, Walden, and Seitz's (1998) model, speech perception involves a number of bottom-up and top-down processes, including auditory and visual sensory–perceptual processing as well as higher order cognitive functions, such as memory and linguistic abilities. Thus, age-related decline in sensory abilities, higher order cognition, and/or the ability to integrate auditory and visual information may contribute to difficulties that older adults experience during speech perception.

At the lowest level of processing, speech perception is affected by the quality of the sensory speech signal. Older adults often suffer from presbycusis, an age-related sensory–neural hearing loss that leads to an increase in hearing threshold for high-frequency sounds (Pichora-Fuller, 2003). Auditory processing also becomes slower and there is a loss of temporal neural synchrony, which causes difficulty in perceiving stimuli that are highly dependent on temporal cues, such as speech stimuli (Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007; Schneider, 1997; Schneider & Pichora-Fuller, 2001). In the visual domain, older adults experience decline in visual acuity and contrast sensitivity (e.g., Burg, 1966; Owsley, Sekuler, & Siemsen, 1983), which may contribute to their poorer performance on lip-reading (visual-only [V-only]) compared with younger adults (e.g., Sommers et al., 2005).

Despite poorer lip-reading skills, older adults benefit from AV speech presentation. For example, Sommers and colleagues (2005) estimated AV speech benefits after controlling for the quality of sensory input. This was defined as the benefit derived from the additional visual cues (visual enhancement) or auditory cues (auditory enhancement) relative to unisensory auditory and visual performance, respectively. Both measures indicated that the AV condition led to similar benefits in older and younger adults. Importantly, older adults were found to rely on visual speech cues more than younger adults, possibly in order to compensate for age-related decline in hearing (e.g., Cienkowski & Carney, 2002; Thompson, 1995; Thompson & Malloy, 2004). At the electrophysiological level, older adults show even larger facilitation of early auditory ERP responses than younger adults during a word categorization task involving background noise (Winneke & Phillips, 2011). These results support the argument that older adults not only benefit from an AV presentation during speech perception but also, in certain instances, benefit to an even greater extent than younger adults.

With regard to WM, normative decline has been demonstrated in older adults (e.g., Bopp & Verhaeghen, 2005; Vaughan, Basak, Hartman, & Verhaeghen, 2008), although the degree may depend on task complexity. For example, there seems to be greater decline on tasks that put more demand on attentional control, such as computation spans or letter-number sequencing, than on tasks that require only short-term memory (STM) or simple reordering, such as digit span forward and digit span backward (see Bopp & Verhaeghen, 2005). The *n*-back task is a complex WM task that requires the constant updating of WM while matching a currently presented stimulus with a target stimulus held in WM (Watter, Geffen, & Geffen, 2001). The task is sensitive to aging. Older adults' responses are generally slower and less accurate than those of younger adults (e.g., Mattay et al., 2006; Vaughan et al., 2008), with accuracy being more affected by age-related changes at the 2-back than the 1-back load (e.g., Nyberg, Dahlin, Stigsdotter Neely, & Bäckman, 2009; Van Gerven, Meijer, Prickaerts, & van der Veen, 2008; Verhaeghen & Besak, 2005).

Taken together, there is considerable evidence showing that both sensory and WM processing is compromised with aging. This

---

[1] Both van Wassenhove and colleagues (2005) and Winneke and Phillips (2011) showed that the amplitude of the auditory N1 differs from the amplitude of summed waveforms in the unimodal conditions (A+V). This indicates that the N1 amplitude reduction during AV speech reflects a genuine multisensory interaction, rather than the addition of two independent sensory stimuli.

makes the issue of resource allocation during perceptual and WM processing very pertinent for older adults.

AV Speech and WM

As summarized by Schneider and Pichora-Fuller (2000), there is convincing evidence for the notion that an increased perceptual load has a negative effect on cognitive performance. In the speech perception domain, it is argued that if a significant amount of processing resources must be devoted to speech signal decoding (e.g., when resolving a degraded auditory input), then fewer resources will be available for high-order processing, such as storing words and meanings in WM (Schneider & Pichora-Fuller, 2000). This suggestion has been supported by studies that showed that hearing deficits (McCoy et al., 2005) or listening under background noise (Rabbitt, 1968) have a negative effect on the ability to remember presented information. Importantly, these memory effects were found to be independent of speech recognition; that is, participants did not remember less information simply because they could not perceive it correctly. Rather, it appeared that effortful perceptual processing, caused by hearing impairment or background noise, depleted resources that could be otherwise used for information storage in memory.

As previously mentioned, AV speech can facilitate speech perception. Thus, in contrast to hearing deficits or background noise, which make perceptual processing more difficult, AV speech could have a positive effect on higher order processing. In support of this notion, Pichora-Fuller (1996) presented young adults with spoken sentences with and without background noise. After each sentence, participants had to report the last word of the sentence. Following each set of sentences varying in number, participants were prompted to recall all last words from the set. To account for perception errors on WM scores, credit was given even for misperceived words. For sentences presented under background noise, participants recalled more words in the AV than A-only modality, suggesting that AV speech improves WM processing in perceptually effortful situations in younger adults. In contrast, Brault, Gilbert, Lansing, McCarley, and Kramer (2010) reported that an AV speech presentation improved speech perception but did not have a positive effect on WM. They used a running WM paradigm, in which participants listened to word lists of varying length. At the end of each list, participants had to report the last three words. Correct recall of the third-to-last and the second-to-last words was considered to reflect both WM performance and speech perception, whereas the correct recall of the last word was considered to reflect speech perception only. It was hypothesized that if the AV presentation has an influence on WM, rather than merely on speech perception, there should be an interaction between the modality of presentation and the serial position of the word in the set. Results showed that AV speech presentation improved word recall in comparison to A-only condition for participants with mild hearing loss who were good lip readers or when speech stimuli were presented in noise. However, the AV benefit was equal for each of the three last words, leading to the conclusion that AV presentation does not have an effect on WM per se (Brault et al., 2010). However, it is possible that Brault and colleagues (2010) failed to find a beneficial effect of AV presentation on WM because the task was insensitive to subtle improvements in WM and/or because their small sample size precluded finding a reliable interaction.

In the current study, we employed the *n*-back task, which includes varying levels of WM load and has been found to be a valid (e.g., Gevins & Smith, 2000) and reliable measure of WM (e.g., Hockey & Geffen, 2004). In this task, participants have to maintain and update a running amount of information in WM and determine whether a current stimulus matches a stimulus presented exactly *n*-trials before. Moreover, both reaction time (RT) and accuracy can be measured. As noted, the *n*-task was found to be sensitive to aging effects, including decline in both processing speed and WM load capacity (Mattay et al., 2006; Nyberg et al., 2009; Van Gerven et al., 2008; Verhaeghen & Besak, 2005; Vaughan et al., 2008). Importantly for our purposes, the *n*-back task has also been validated in electrophysiological paradigms by examining the P3 ERP component. The P3 is elicited during the processing and categorization of task-relevant stimuli (Polich, 2007). Its latency is considered to reflect the timing of mental processes, whereas the P3 amplitude is considered to reflect the intensity of processing (Kok, 2001). Watter and colleagues (2001) showed that as memory demands increased (i.e., higher *n*-back conditions), the amplitude of P3 decreased, whereas the latency of the P3 remained constant. Similarly, Segalowitz, Wintink, and Cudmore (2001) found that the amplitude of the P3 decreased by about 1 μV per each successive load. This sensitivity, plus the ability to measure WM performance at multiple levels of load, makes the *n*-back task a useful measure of WM.

Present Study

Considering an integrated system between perception and cognition, facilitated speech perception should lead to more available resources for higher-order functions and thus better WM memory performance. In the current study, we expected that facilitated perception of speech tokens during AV presentation would have a positive effect on subsequent WM processes. Thus, we expected better WM performance in the AV condition compared with unimodal (A-only and V-only) conditions. We also expected that this improvement would be directly related to the sensory facilitation afforded by the AV modality as measured by ERPs. Specifically, we predicted that (a) an AV presentation would lead to facilitated speech perception, operationally defined as a decrease in the amplitude of the auditory N1 during AV speech compared with A-only speech; and (b) this facilitated speech perception would be related to improvement in WM performance in AV compared with A-only condition, both at the behavioral level (i.e., increase in accuracy and decrease in RT) and at the electrophysiological level (i.e., increase in the amplitude and decrease in the latency of P3). Thus, participants with a larger AV facilitation effect were expected to have a larger WM improvement than participants with smaller AV facilitation. Furthermore, due to a more restricted pool of processing resources available to older adults, they were expected to benefit from the AV presentation more than younger adults.

Method

Participants

Twenty-three younger adults, recruited through local advertisement and the Concordia University participant pool, and 20 older adults, recruited through an existing laboratory database, participated. In order to be eligible, participants had to (a) be right-handed, (b) be dominant English speakers, meaning that they learned English before the age of 5 years and have used it as their primary language ever since, (c) be in good self-reported health with no significant medical or neurological conditions, and (d) have normal or corrected-to-normal vision and normal hearing. Demographic information is presented in Table 1.

Vision screening consisted of the MARS Letter Contrast Sensitivity test (by MARS Perceptrix) with participants scoring within age-appropriate ranges (Haymes et al., 2006). Auditory acuity was screened by measuring pure tone averages (PTA; average hearing threshold for frequencies of 500, 1000, 2000 Hz) by using a Welch Allyn, a.m. 232 Manual Audio Meter. In order to be included, a participant's PTA could not exceed 25 dB HL (Katz, 1985). In reality, these values were much lower (see Table 1). Lastly, to ensure intact cognitive functioning, older adults were assessed using the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) and had to attain a score of 26 or more to be included in the experiment. Participants provided informed consent and were compensated for their time either by payment of $10 per hour or by receiving student participant pool credits. The study was approved by the Concordia University research ethics board.

Stimuli

The stimuli consisted of nine single syllable digits (1, 2, 3, 4, 5, 6, 8, 9, 10) spoken by a woman with a neutral facial expression. Digit 7 was omitted because it is disyllabic and thus disproportionately distinguishable from other digits. Stimuli were digitally audio-video recorded with an auditory digitization rate of 48,000 Hz in a recording studio. Each showed the full face, head, and shoulders of the speaker against a green background. Each digit was presented as a short video clip, edited in Adobe Premiere (Video codec, Windows Media Video 9; frame size, 500 px × 388 px; frame rate, 29.97 fps; Audio codec, Windows Media Audio; sample rate and size, 44,100 Hz 16-bit). The appearance of the speaker's face occurred nine frames before the first obvious lip movement and the video clip ended nine frames after the lips stopped moving. The average lag time across different digits between the onset of the lip movement and first sound was 395.33 ms (SD = 103.24). The average trial length was 2010 ms (SD = 160 ms). The intertrial interval was 2400 ms. The software Praat (version 5.1.30; Boersma & Weenink, 2010) was used to set the sound intensity of each digit at an approximately equal level. The auditory speech stimuli were presented at a mean intensity of 70.2 dB (SPL; SD = 2.59 dB) using EARLINK tube ear inserts (Neuroscan, El Paso, Texas). Imperceptible triggers were embedded in the video files at the onset of lip movement (i.e., visual triggers) and the speech signal (i.e., auditory triggers), with onsets identified *a priori* through visual and auditory inspection of the speech waveform. These triggers consisted of short transistor-transistor logic (TTL) pulses, which signaled the onset of a stimulus to the recording electroencephalogram (EEG) amplifiers. The videos were presented against a black background on a 15-in. CRT monitor positioned 60 cm from the participant.

Each AV video contained both auditory and visual channels (i.e., participants could both hear and see the speaker uttering digits). Stimuli in the A-only and V-only conditions were derived from these videos. For V-only stimuli, the auditory channel was deleted (i.e., participants could see but not hear the person speaking the digit). For the A-only condition, the video channel was deleted (i.e., participants could hear but not see the person speaking the digit). Thus, the A-only and V-only stimuli were identical to AV stimuli except for the absence of one modality. During the A-only condition, a black screen with a white dot in the middle was presented in order to maintain participants' eye fixation. In all modality conditions, the imperceptible triggers signaling the onset of lip movement and the onset of the auditory speech signal were present so that visual and auditory ERPs could be measured and compared across modalities. Inquisit (version 2.0; Millisecond Software, 2008) was used for stimulus presentation.

Procedure

Each participant completed the *n*-back task under three conditions: A-only, V-only, and AV. The order of A-only, V-only, and AV conditions was counterbalanced across participants. The participants' task was to decide whether the currently presented digit matched the digit presented in the previous trial (1-back condition), 2 trials before (2-back condition), or 3 trials before (3-back condition), or to decide whether the currently presented digit matched a "target" digit defined at the beginning of a block (0-back condition). There were 100 trials in each *n*-back condition and participants indicated their response by holding a computer mouse in both hands and pressing a "Match" or a "Non Match" button with their thumbs. The left- and right-hand "match/non-match" responses were counterbalanced across participants. In

Table 1
Demographic Information Describing Means and Standard Deviations of Younger and Older Adults

	Younger adults	Older adults	p value
Number	23 (15 females)	20 (15 females)	
Age (Years)	24.4 (SD = 3.70)	69.1 (SD = 6.17)	
Education (Years)	16.5 (SD = 2.11)	14.2 (SD = 2.60)	p = .002
Cognitive functioning (MoCA; max. 30)	N/A	27.5 (SD = 1.28)	
Hearing threshold (dB HL)[a]	2.5 (SD = 3.28)	13.4 (SD = 6.34)	$p < .001$
Vision threshold: binocular[b]	1.8 (SD = 0.05)	1.7 (SD = 0.06)	$p < .001$

Note. MoCA = Montreal Cognitive Assessment.
[a] Pure tone average (i.e., average of hearing sensitivity at 500, 1000, 2000 Hz) for the left and right ear combined. [b] Score on Mars Letter Contrast Sensitivity Test.

40% of the trials, the current digit matched the digit presented *n*-trials before, whereas in 60% of trials it did not.

Before starting the experiment, participants were first trained to identify the V-only stimuli (i.e., speech-reading). Participants had to correctly identify (i.e., by naming them out loud) all the digits presented in both a sequential order and random presentation before proceeding with the experiment. This training took approximately 1 min because participants learned it quickly and it rarely had to be repeated. Next, participants practiced the key assignment for "Match" and "Non Match" responses on the computer mouse. For this task, they were given an AV "target" digit at the beginning of the practice block, and they needed to indicate whether the digits from trial to trial matched the target digit or not. There were 10 trials during this practice and the task was essentially identical to the AV 0-back condition. Following this practice, the experiment began. Before each new *n*-back condition, there were 10 practice trials to make sure that the participant understood the task. During these practice trials, participants would hear a low frequency beep (250 Hz), lasting for 100 ms, whenever they made a mistake.

In the 0-back conditions, participants were asked to remember a target digit (either 3, 6, or 9) and to decide whether the current trial matched the predefined target digit. Thus, the condition controlled for stimulus input and motor responses, but there was no working memory load. Different target digits were assigned to different sensory modalities and the digit-modality combinations were counterbalanced across participants. In 1-, 2-, and 3-back conditions, there were five "warm up" trials at the beginning of each block to allow participants to get used to the new *n*-back level. Responses during warm-up trials were not included in the analyses. Each block consisted of a semirandom order of digits that maintained a 40/60 "Match"/"Non Match" ratio. Within each *n*-back level, there were three different blocked trial sequences, one for each sensory modality. Thus, a participant heard different trial sequences in each sensory modality and the modality-sequence combinations were counterbalanced across participants. Each participant completed the 0- to 3-back conditions in ascending order. The order of modality conditions was counterbalanced across participants, and participants were randomly assigned to a particular modality order. For each participant, the modality order was the same across all *n*-back levels and participants completed all three modalities at a given *n*-back level before moving to the next higher *n*-back condition.

Two behavioral measures were collected: (a) reaction time (RT) measured in milliseconds (ms) and operationalized as the amount of time between the onset of the auditory trigger and the participant's button response[2]; (b) accuracy, or the percentage of correct responses. Any response that occurred earlier than 200 ms was excluded from the analysis due to the fact that such early responses were unlikely to reflect a valid response to the current trial.

EEG Data Acquisition

A Biosemi ActiveTwo EEG system was used to measure electrical brain activity while participants were performing the task. The brain activity was recorded from 64 channels arranged according to the International 10–20 system (Jasper, 1958). Vertical and horizontal eye movements were monitored by electrooculograms (EOGs) with electrodes positioned above and below the left eye and beside the outer canthi of each eye. An EEG was recorded at a sampling rate of 2048 Hz, with a high-pass filter of .16 Hz and a low-pass filter of 100 Hz. The files were then downsampled off-line to 512 Hz.

The BioSemi data format was converted to the Neuroscan continuous data format using Polygraphic Recording Data Exchange (PolyRex; Kayser, 2003). Before the conversion, the EEG recordings were rereferenced off-line in PolyRex to a linked left and right earlobe reference. Subsequent data processing steps were performed using Scan software (version 4.3.1; Compumedics Neuroscan, 2003). Excessive ocular artifacts, such as eyeblinks, were corrected using a spatial filter technique (NeuroScan Edit 4.3.1 manual).

The files were epoched into a −100 to 1000 ms window around the onset of each auditory trigger (defined as 0 ms) in order to compute averaged auditory ERPs. The auditory trigger occurred at the same time point in each modality for a given digit, even in the V-only modality where no audible sound was present.[3] To assess the effect of an AV interaction, the amplitude and latency of the auditory N1 evoked during the AV condition was compared with the amplitude and latency of the summed ERP activity at the same time point in the A-only and V-only conditions (yielding an A+V waveform). For the A-only condition, this was the onset of the auditory signal. For the V-only condition, this time point was at the onset of the auditory signal, had it been audible. This allowed us to assess the electrical brain response in the V-only condition at exactly the same time points as in the A-only and AV conditions.

The EEG waveforms were baseline corrected according to the prestimulus period (−100 to 0 ms before the auditory trigger). A trial was rejected if horizontal EOG activity exceeded ±75 μV or if it contained activity exceeding ±100 μV in any of the active EEG electrodes around the center of the head (i.e., Fz, F1, F2, FCz, FC1, FC2, Fz, F1, F2, CPz, Pz, P1, P2, POz, PO1, PO2, Oz, O1, O2). Because the P3 is more prominent for less-frequent stimuli, only the correct match responses were used in the analyses. Consequently, we chose to analyze the N1 responses to match trial only. The mean number of correct "match" trials that passed artifact rejection was 29.7 (*SD* = 7.99) out of a possible maximum of 40. The waveforms were then averaged separately for each modality and *n*-back condition. The averages were then filtered off-line over the range of 1 to 30 Hz, using a zero phase shift band pass filter.

Four measures were obtained from the ERP components, namely, the amplitude (measured in μV) and the latency (measured in ms) of the N1 and the P3. The amplitude of the N1 was measured as an absolute difference between the peak of the N1 and the trough of the preceding P1 deflection. The P1 was scored as the

---

[2] The RT measured from the onset of the auditory trigger resulted in an underestimation of true RT in V-only and AV trials because the onset of visual speech information preceded the auditory trigger. To measure RT in this manner was important, however, in that it allowed for comparison of the V-only and AV conditions to the A-only condition, where the first cue about which digit was presented came from the later-occurring auditory speech information.

[3] This was necessary because the literature shows that AV speech results in clear modulation of auditory ERPs in contrast with visual ERPs. Thus, we were specifically interested in the modulation of the auditory N1 by the presence of visual speech cues and this logically necessitated that we compare conditions at this time point.

most positive peak occurring between 40 and 70 ms, whereas N1 was scored as the most negative peak occurring between 80 and 200 ms. The P3 amplitude was measured relative to the 0 μV prestimulus baseline and defined as the most positive peak occurring between 300 and 600 ms. For all ERP components, semiautomatic peak detection was performed by Scan software (version 4.3.1; Compumedics Neuroscan, 2003) on averaged waveforms of each individual, followed by manual check and adjustment if necessary. Latencies of both the N1 and P3 were operationalized as the components' peak relative to the onset of the auditory trigger.

Results

The data were analyzed using SPSS (version 16). Our *a priori* comparisons involved main effects of modality, *n*-back load, and age, and interactions of the first two factors with age. These were tested in ANOVA, followed by tests of simple effects on the interaction(s). For within-subject factors with more than one degree of freedom in the numerator, the Greenhouse-Geisser nonsphericity correction was used and, according to convention (Jennings, 1987), the uncorrected degrees of freedom and Greenhouse-Geisser epsilon (ε) values are reported. The reported *p* values and the mean square error (*MSE*) values are adjusted according to the Greenhouse-Geisser correction. Unless otherwise specified, all results reported are significant at $\alpha = .05$.

Behavioral Results

Repeated measures ANOVAs with modality (V-only, A-only, and AV) and load (0-,1-, 2-, 3-back) as within-subject factors and age (younger and older) as a between-subject factor were performed separately for accuracy and RT.

Accuracy. In order to account for the possibility of response bias, the accuracy data were calculated as d' and the results are plotted in Figure 1. There was a main effect of load, $F(3, 123) = 207.7$, $MSE = 1.58$, $p < .01$, $\varepsilon = .65$, $\eta_p^2 = .84$, such that accuracy decreased with each increase in memory load for both age groups. There was also a main effect of age, $F(1, 41) = 11.8$, $MSE = 2.64$, $p < .01$, $\eta_p^2 = .22$, which was qualified by an Age × Load interaction, $F(3, 123) = 3.1$, $MSE = 1.58$, $p = .05$, $\varepsilon = .65$, $\eta_p^2 = .07$. Older adults were less accurate than younger adults in the 2-back and 3-back conditions, but not in the 0-back and 1-back conditions. Importantly, there was a main effect of modality, $F(2, 82) = 8.1$, $MSE = .71$, $p < .01$, $\varepsilon = .93$, $\eta_p^2 = .16$, showing that participants were more accurate in the AV than in the A-only and V-only conditions. In order to test whether the AV modality would indeed facilitate WM performance when demands were highest, we compared performance at each level of WM load between the unisensory and the AV conditions. As expected, the simple effect of modality was not significant at either the 0- or 1-back loads, both Fs $< .67$, $p > .53$, but was significant at the 2-back load, $F = 9.2$, $p < .01$, and the 3-back load, $F = 8.2$, $p < .01$. The interaction between modality, load, and age was not significant, $F(6, 246) = 2.0$, $MSE = .59$, $p = .09$, $\varepsilon = .79$, $\eta_p^2 = .05$.

Reaction time. The RT data for correct match trials are plotted in Figure 2. Analyses revealed a main effect of load, $F(3, 123) = 123.5$, $MSE = 25980.57$, $p < .01$, $\varepsilon = .73$, $\eta_p^2 = .75$, with RT increasing with each increase in memory load, and a main effect of age, $F(1, 41) = 4.0$, $MSE = 140780.84$, $p = .05$, $\eta_p^2 = .09$, with older adults being slower than younger adults. There was a main effect of modality, $F(2, 82) = 46.7$, $MSE = 13337.68$, $p < .01$, $\varepsilon = .95$, $\eta_p^2 = .53$, which was qualified by a Modality × Load interaction, $F(6, 246) = 8.0$, $MSE = 6726.77$, $p < .01$, $\varepsilon = .76$, $\eta_p^2 = .16$. Participants benefited from the AV presentation at all levels of memory load compared with the unisensory presentations, and they were faster in the A-only compared with the V-only condition during 0-back and 1-back loads (i.e., 0-back, AV < A < V; 1-back, AV < A < V; 2-back, AV < A = V; 3-back, AV < A = V). The interaction between modality, load, and age was not significant, $F(6, 246) = 1.0$, $MSE = 6726.77$, $p = .40$, $\varepsilon = .76$, $\eta_p^2 = .03$.

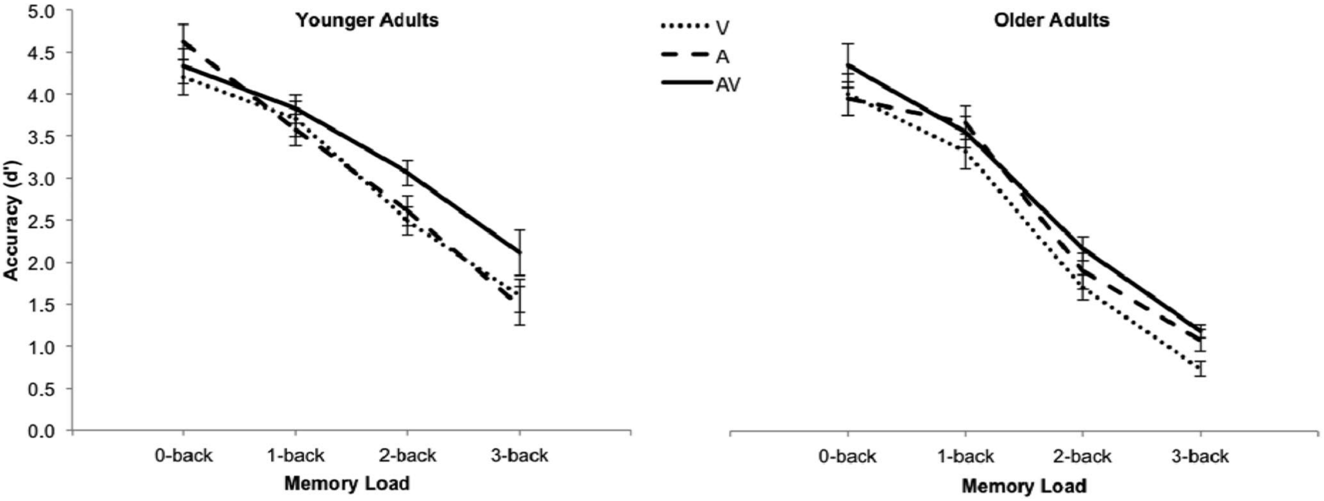

Figure 1. Mean accuracy (calculated as d') and standard error bars for younger (left panel) and older (right panel) adults in visual-only (V-only), auditory-only (A-only), and auditory–visual (AV) conditions at 0-, 1-, 2-, and 3-back memory load. Note the reduction in accuracy with increasing memory load. In addition, note better performance in AV condition compared with V-only and A-only conditions during the 2-back and 3-back loads.

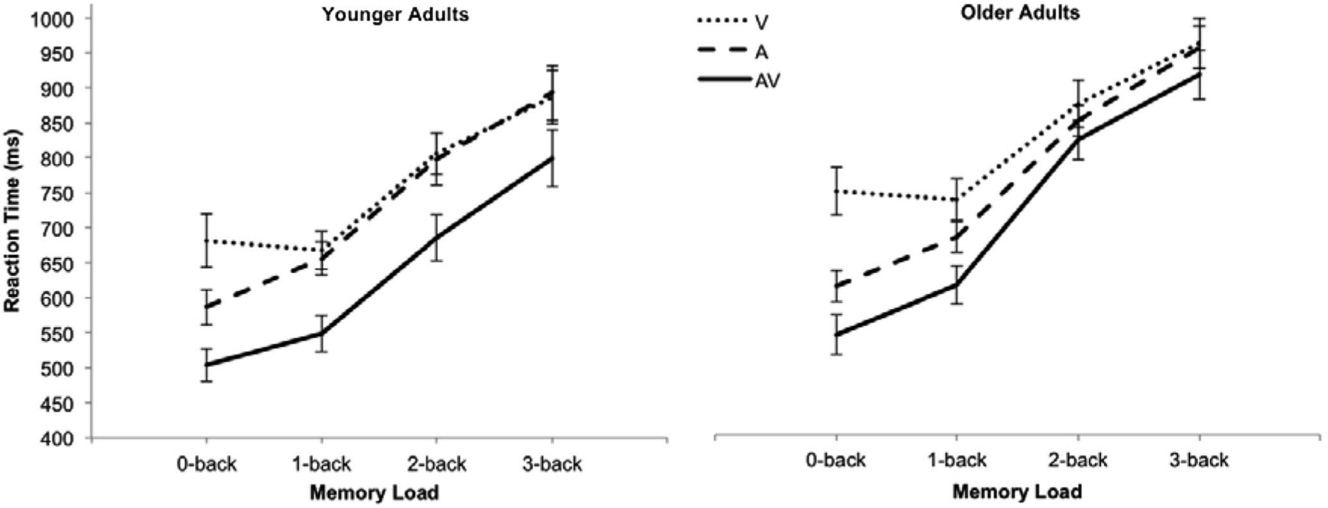

Figure 2. Mean reaction times and standard error bars for younger (left panel) and older (right panel) adults in visual-only (V-only), auditory-only (A-only), and auditory–visual (AV) conditions. Note the increase in reaction time with increasing memory load. In addition, note the faster responses in AV compared with V-only and A-only conditions across all *n*-back loads.

Electrophysiological Results

The auditory N1. Given that the auditory P1-N1 complex is most prominent at the midcentral electrodes (Näätänen & Picton, 1987), and that no hemispheric differences were detected in our previous work (Winneke & Phillips, 2011), the values from the Cz electrode were used for the analyses. To examine the effect of modality, load, and age on the amplitude and latency of the auditory N1, separate repeated measures ANOVAs were performed, with modality (A-only, A+V, and AV) and load (0-,1-, 2-, 3-back) as within-subject factors and age (younger and older) as a between-subject factor. The V-only condition was not included in the analyses, as we were interested in comparing the facilitation of auditory speech processing (i.e., the dominant modality for processing speech) when visual speech cues were absent or present (A-only vs. AV condition). The sum of responses during unimodal conditions (A+V) was used to test whether the responses elicited during the AV condition reflected a true multisensory interaction. If the ERP signals during the AV condition did not differ reliably from the summed activity in the A-only and V-only conditions (i.e., A+V), it would indicate that ERPs evoked in the AV condition merely reflected the summed activity of two independent processes. In contrast, if the AV condition differed from the A+V waveform, it could be inferred that providing visual cues affected auditory processing, and thus that multisensory interaction had occurred.

Auditory N1 amplitude. Mean amplitudes of the auditory N1 across modality, load, and age are shown in Table 2, and the average waveforms for the different modalities are presented in Figure 3.[4] Results revealed a main effect of age, F (1, 41) = 11.7, $MSE = 40.89$, $p < .01$, $\eta_p^2 = .22$, which was qualified by a Load × Age interaction, F (3, 123) = 4.5, $MSE = 5.21$, $p = .01$, $\varepsilon = .88$, $\eta_p^2 = .10$. The amplitude of the N1 was larger for older than for younger adults in 1-, 2-, and 3-back loads. There was also a main effect of modality, F (2, 82) = 34.1, $MSE = 2.67$, $p < .01$, $\varepsilon = .80$, $\eta_p^2 = .45$, which was qualified by a Modality × Age interaction, F (2, 82) = 8.2, $MSE = 2.67$, $p < .01$, $\varepsilon =. 80$, $\eta_p^2 = .17$. For younger adults, the amplitude of N1 was smaller in the AV waveform than in the A+V waveform, but AV and A-only waveforms did not differ. In contrast, for older adults, the amplitude of the N1 was smaller in the AV waveform compared with both the A-only and A+V waveforms (see Table 2).

Auditory N1 latency. Mean latencies of the N1 across different modality, load, and age are shown in Table 3. There was a main effect of age, F (1, 41) = 8.2, $MSE = 886.79$, $p = .01$, $\eta_p^2 = .17$, with the auditory N1 peaking later in older than younger adults (see Figure 3). Importantly, there was a main effect of modality, F (2, 82) = 8.8, $MSE = 252.19$, $p < .01$, $\varepsilon = .74$, $\eta_p^2 = .18$, with the N1 peaking earlier in the AV condition than in the A-only and A+V conditions.

The P3. In order to examine the effect of modality, load, and age on the amplitude and latency of P3, separate repeated measures ANOVAs were performed, with modality (A-only, and AV) and load (0-,1-, 2-, 3-back) as within-subject factors and age (younger and older) as between-subject factor. The maximum peak of the P3 is usually detected at midposterior electrodes between 300 to 1,000

---

[4] Inspection of the waveforms in Figure 3 revealed that modulations in the ERP waveforms in the AV condition were evident even earlier than the N1 component. In order to evaluate this, this earlier positive-going component (the P1) was analyzed using a repeated measures ANOVA, with modality (A-only, A+V, and AV) and load (0-,1-, 2-, 3-back) as within-subject factors and age (younger and older) as a between-subject factor. The results revealed a main effect of age, $F(1,41) = 22.2$, $MSE = 9.11$, $p < .001$, $\eta_p^2 = .35$, which was qualified by a Modality × Age interaction, $F(2, 82) = 7.4$, $MSE = 2.72$, $p = .002$, $\varepsilon =. 82$, $\eta_p^2 = .15$. For younger adults, there was no difference between the modality conditions, whereas for older adults, the P1 was smaller in AV compared with the A-only and the A+V condition. Thus, older adults showed an earlier multisensory interaction than younger adults, replicating our previous findings (Winneke & Phillips, 2011).

Table 2

Mean Amplitudes (μV) and Standard Deviations (in Parentheses) of N1 Amplitude for Younger and Older Adults at the Cz Electrode Site

Load	Modality		
	A-only	A+V	AV
Younger adults			
0-back	4.4 (1.93)	5.4 (2.38)	5.8 (2.80)
1-back	4.7 (2.26)	5.8 (2.69)	4.3 (1.54)
2-back	4.5 (2.20)	4.2 (1.75)	3.5 (1.28)
3-back	4.6 (1.53)	5.4 (2.31)	3.9 (1.73)
Older adults			
0-back	6.4 (2.51)	6.8 (2.65)	5.4 (2.78)
1-back	6.8 (3.03)	6.9 (2.86)	5.6 (2.52)
2-back	7.2 (3.24)	7.8 (3.73)	5.5 (2.53)
3-back	7.5 (2.90)	7.8 (3.45)	5.8 (3.07)

Note. A-only = auditory-only; A+V = auditory + visual; AV = auditory–visual.

ms after presentation of the stimuli (Friedman, Kazmerski, & Fabiani, 1997; Watter et al., 2001), and therefore the values from the Pz electrodes were chosen for the analysis.[5]

P3 amplitude. Mean P3 amplitudes as a function of modality, load, and age are shown in Table 4. There was a main effect of load, $F(3, 123) = 13.7$, $MSE = 3.37$, $p < .01$, $\varepsilon = .87$, $\eta_p^2 = .25$, such that P3 amplitudes in the 0- and 1-back loads were larger than in the 2-back and 3-back loads (see Figure 4). There was also a trend for P3 amplitudes to be larger in 0-back than in 1-back loads ($p = .07$), but 2-back and 3-back loads did not differ ($p = .18$). There was a main effect of age, $F(1, 41) = 4.9$, $MSE = 17.48$, $p = .03$, $\eta_p^2 = .11$, with larger P3 amplitudes in younger than in older adults (see Figure 4). The main effect of modality was not significant, $F(1, 41) = .01$, $MSE = 2.88$, $p = .94$, $\eta_p^2 = .00$.

P3 latency. Mean latencies across different modality, load, and age are shown in Table 5. The results showed a significant main effect of modality, $F(1, 41) = 12.3$, $MSE = 10233.59$, $p < .01$, $\eta_p^2 = .23$, with the P3 peaking earlier during the AV condition compared with the A-only condition (see Figure 5). Neither the main effect of load, $F(3, 123) = .9$, $MSE = 6892.25$, $p = .44$, $\varepsilon = .80$, $\eta_p^2 = .02$, nor age, $F(1, 41) = 2.7$, $MSE = 25870.91$, $p = .11$, $\eta_p^2 = .06$, was significant.

Correlations Between Electrophysiological and Behavioral Results

To examine whether resources saved at the perceptual level during multisensory speech processing might be associated with better WM performance, correlations between the facilitation of perceptual processing and improvement in WM performance were analyzed. The facilitation of perceptual processing was defined as the reduction in the amplitude of the N1 in the AV condition relative to the A-only condition (i.e., A-only minus AV). The WM improvement was defined as larger accuracy and P3 amplitude values, and smaller RT and P3 latency values, in the AV condition compared with the A-only condition. Overall, we reasoned that a larger N1 amplitude reduction reflects larger AV facilitation and should thus lead to larger improvement in WM performance at the behavioral level (i.e., higher accuracy and lower RT) as well as at the electrophysiological level (i.e., larger amplitude and earlier latency of P3) in the AV condition in comparison with the A-only condition. To illustrate, let's say that Participant A had an N1 amplitude of 5.7 μV in the A-only condition and 4.0 μV in the AV condition (a difference of 1.7 μV), and Participant B had an N1 amplitude of 5.7 μV in the A-only condition and 5.0 μV in the AV condition (a difference of 0.7 μV). In this case, Participant A showed a larger multisensory facilitation compared to Participant B and thus was expected to have a higher accuracy and P3 amplitude, a lower RT and P3 latency in the AV condition compared with the A-only condition. In general, a larger reduction in N1 should correlate positively with behavioral and P3 measures of WM improvement in the AV condition (i.e., the difference between A-only and AV). The results of the correlations are presented in Table 6.

The results suggest that, in many cases the AV-related sensory improvement was reliably related to WM improvement. Specifically, for the young adults, N1 amplitude reduction correlated positively with improvement in RT and P3 latency in 1-back condition and with improvement in accuracy and RT in 2-back condition. Additionally, there was a trend toward a positive correlation with improvement in P3 latency in the 2-back condition. For the older adults, N1 amplitude reduction showed a trend toward a positive correlation with improvement in accuracy in the 3-back condition. Overall, amongst the significant correlations or those showing trends toward significance, all were in the predicted direction.

Discussion

This study supports our hypothesis that the facilitation of perceptual processing, afforded by AV speech perception, is associated with better WM performance. Both younger and older adults performed better when they were asked to complete a WM task in the AV modality compared with A-only and V-only modalities. In addition, there was some support for the hypothesis that more efficient perceptual processing during AV speech, indicated by a smaller auditory N1 during the AV condition compared with the A-only condition, correlated with better WM functioning at both the behavioral as well as the electrophysiological level. This suggests that more efficient perceptual processing may facilitate higher order cognitive functions.

Behavioral Results

As expected, a decline in accuracy and an increase in RT was observed with each successive increase in WM load, regardless of modality.[6] This was true for both younger and older adults and demonstrated the effectiveness of the *n*-back manipulation. In

---

[5] The analysis including five midline electrodes (FCz, Cz, CPz, POz, and Pz) confirmed that P3 amplitude reached the maximum peak at the Pz electrode. There was a main effect of electrode, $F(4, 160) = 208.4$, $MSE = 7.69$, $p<.001$, $\varepsilon= .38$, $\eta_p^2 = .84$, with a significant decrease in the amplitude of P3, moving from posterior to frontal locations. This effect was evident in both age groups and therefore only the results from Pz electrode are reported in the text.

[6] The only exception was the V-only condition, where reaction times during 0-back and 1-back were not different from each other. There was, however, a significant increase of reaction time in the 2-back and a subsequent significant increase in the 3-back condition.

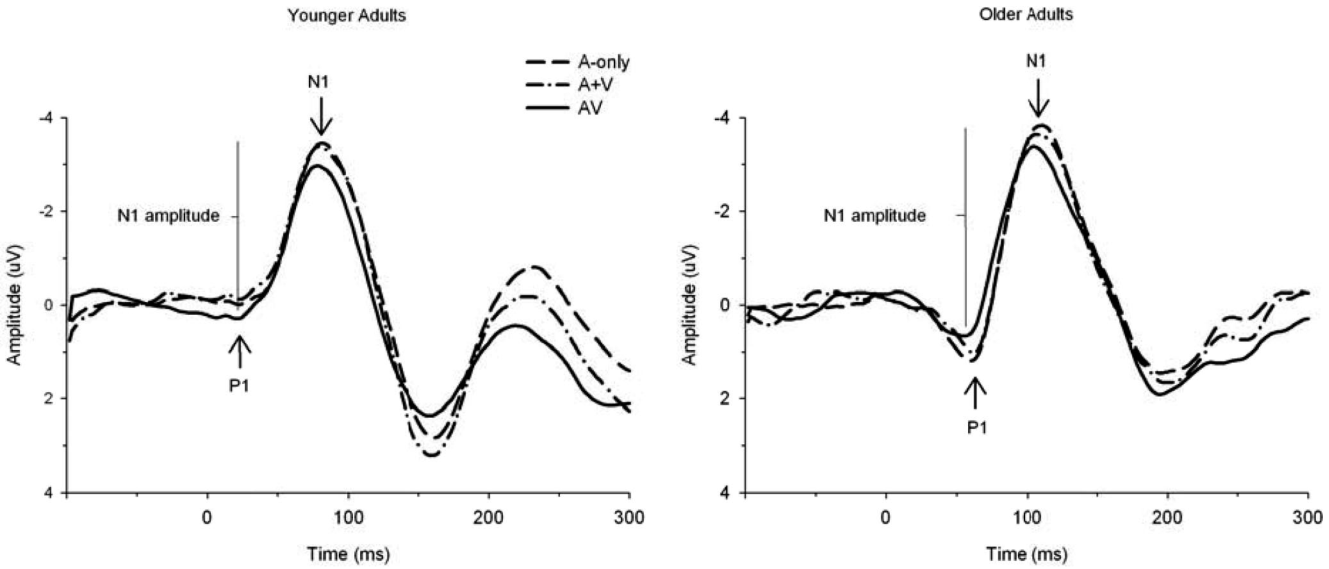

Figure 3. Grand average waveforms of the auditory N1 at Cz electrode for younger adults (left panel) and older adults (right panel) during auditory-only (A-only), auditory + visual (A+V), and auditory–visual (AV). The waveforms are averaged across *n*-back loads. The 0 ms on the x-axis denotes stimulus onset. Note the positive peak occurring at around 50 ms poststimulus (the P1) and the negative peak occurring at around 100 ms poststimulus (the N1). The difference in P1-N1 microvolt values was taken as the measure of N1 amplitude.

addition, in agreement with previous literature (e.g., Nyberg et al., 2009; Van Gerven et al., 2008; Verhaeghen & Besak, 2005), compared with younger adults, older adults performed more slowly across all WM loads and less accurately under certain WM loads. For the conditions that required no (i.e., 0-back) or a relatively low (i.e.,1-back) WM load, older adults performed similarly to younger adults, whereas for the more demanding loads (i.e., 2-back and 3-back), the older adults were less accurate. These results suggest that during low WM loads, older adults can be as accurate as younger adults, but require more time to complete the task due to a decline in speed of processing (e.g., Diamond et al., 2000).

Table 3
Mean Latencies (Ms) and Standard Deviations (in Parentheses) of N1 Latency for Younger and Older Adults at the Cz Electrode

	Modality		
Load	A-only	A+V	AV
Younger adults			
0-back	99.1 (14.29)	99.0 (17.76)	97.2 (15.27)
1-back	96.0 (12.62)	99.5 (15.35)	92.2 (11.56)
2-back	96.2 (15.60)	98.9 (17.54)	88.8 (11.74)
3-back	93.9 (12.14)	92.7 (14.64)	90.8 (12.42)
Older adults			
0-back	101.4 (12.14)	106.3 (18.07)	101.5 (14.84)
1-back	106.4 (16.85)	106.3 (17.34)	101.1 (19.86)
2-back	102.8 (16.30)	107.1 (20.78)	97.5 (13.82)
3-back	105.4 (11.74)	102.7 (12.22)	95.9 (14.33)

Note. A-only = auditory-only; A+V = auditory + visual; AV = auditory–visual.

Turning to our central prediction that the perceptual facilitation afforded by AV speech would result in better WM performance, we found an improvement on both behavioral measures of WM in the AV condition compared with the A-only and V-only conditions. For accuracy scores, the enhancement effect during AV presentation relative to A-only or V-only presentation was evident in conditions that required higher WM loads (i.e., 2-back and 3-back). This finding suggests that even though unimodal conditions may be sufficient when WM load is low (i.e., 0- and 1-back memory load), an AV speech presentation becomes especially useful when the demands on WM capacity increase and more resources are required to complete a task accurately. This observation supports the suggestion that the facilitation of perceptual processing becomes important when perceptual and higher-order processing would otherwise exceed the capacity of available resources. In terms of RT, both younger and older adults benefited from the AV presentation at all levels of WM load, suggesting that the AV information facilitates processing speed independently of WM load.

The lack of an interaction between modality and age on any of the behavioral measures suggests that both age groups benefited from AV presentation to a similar extent. This is notable, given that the older adults in the current study had significantly lower contrast sensitivity and higher pure tone hearing thresholds (i.e., poorer visual and auditory functioning, respectively) compared with younger adults. This, together with similar findings (e.g., Sommers et al., 2005; Winneke & Phillips, 2011), suggests that, despite age-related sensory declines, older adults benefit from AV speech presentation. However, this study is unique in showing that such an early multisensory benefit improves later WM performance in older adults. Moreover, presenting stimuli in the AV

Table 4
Mean Amplitudes (μV) and Standard Deviations (in Parentheses) of P3 at the Pz electrode in Younger and Older Adults

	Modality	
Load	A-only	AV
Younger adults		
0-back	5.9 (2.52)	5.9 (2.56)
1-back	5.4 (2.47)	6.2 (2.47)
2-back	4.3 (1.59)	4.5 (2.22)
3-back	4.8 (2.59)	4.3 (2.04)
Older adults		
0-back	5.3 (1.93)	4.7 (2.51)
1-back	4.2 (2.21)	4.1 (2.32)
2-back	3.5 (1.49)	3.5 (1.46)
3-back	3.8 (1.85)	4.1 (1.63)

Note. A-only = auditory-only; A+V = auditory + visual; AV = auditory–visual.

modality appeared to facilitate the speed and accuracy of the older adults' WM performance to the extent that it no longer differed from the performance of younger adults in the A-only condition,[7] the dominant modality for speech perception.

It is important to bear in mind that, in this study, stimuli were not masked and were presented at a clearly audible intensity level for both younger and older adults, as evidenced by their near-perfect scores in the 0-back load during the A-only modality. Thus, the observed AV benefit on WM performance was not simply the result of improved stimulus identification. Instead, our results suggest that the facilitation of perceptual processing has a positive effect on subsequent memory processing, supporting the observation by McCoy and colleagues (2005). Importantly, our study demonstrated that the facilitation of perceptual processing can be beneficial, even in relatively ideal listening conditions.

Electrophysiological Results

We predicted that AV presentation would facilitate speech perception, as reflected by changes in the auditory N1, as an index of early sensory signal processing. The auditory N1 occurred earlier during AV trials compared with A-only and A+V trials. Thus, the study replicated previous findings of faster auditory processing during AV presentation in younger adults (e.g., Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005; Winneke & Phillips, 2011) and in older adults (Winneke & Phillips, 2011), suggesting the speeding of auditory processing through the presence of additional visual speech cues during AV speech perception.

Equally important, we also observed a reduction of N1 amplitude (i.e., a facilitation effect) in the AV modality compared with the unisensory modality. However, there were two nuances to this effect. First, the AV amplitude interaction occurred earlier in the older than in the younger adults. That is, the evidence of AV interaction (i.e., difference between AV and A+V waveform amplitudes) was evident during the timing of the N1 level for younger adults but, for older adults, it was clearly notable even earlier, at the timing of the preceding P1 component (see Footnote 4). This finding is consistent with our previous work (Winneke & Phillips, 2011).

Second, the AV facilitation of these early sensory components (i.e., difference between AV and A-only amplitude) was larger in the older than the younger adults. A likely explanation for the smaller auditory facilitation in younger adults (despite the facilitation of their behavioral performance) is that, given the use of unmasked auditory stimuli, the auditory signal was very clear and there was little to interfere with auditory processing in a group with very low hearing thresholds. In other words, the AV modality did not have as strong a beneficial effect on auditory processing in younger adults because they were already close to a maximum level of efficiency during the A-only condition. In contrast, the older adults in this study had higher (although clinically normal) hearing threshold than the younger adults. This age-related decline in hearing likely caused the A-only condition to be more effortful for older adults (Pichora-Fuller, Schneider, & Daneman, 1995; Tun, McCoy, & Wingfield, 2009). As noted, it has been demonstrated that AV speech is especially useful when speech perception is challenged by background noise (e.g., Sommers et al., 2005) or auditory impairment (e.g., Grant et al., 1998). This is consistent with the overall pattern found here—that the AV benefit was larger (as measured by electrophysiology) in the older than in the younger adults. In order to observe electrophysiological AV benefits in younger adults, speech perception may need to be more challenging. In fact, recent work in our lab has shown that when stimuli were presented in background noise during the same WM task used here, young adults displayed a significant reduction in auditory N1 amplitude during AV presentation compared with both A-only and A+V waveforms (Frtusova, Amarsi, & Phillips, 2011).

Importantly, we ruled out that the N1 modulations in the AV condition were simply due to the simultaneous presentation of stimuli through two sensory modalities. In the current study, both the amplitude and latency of auditory N1 during AV speech presentation differed from the sum of amplitudes and latencies during unimodal presentations (i.e., A+V), indicating that the response during the AV condition represents a genuine multisensory interaction rather than an artifact resulting from processing information separately in two independent sensory channels (see also van Wassenhove et al., 2005 and Winneke & Phillips, 2011).

Before turning to the effect of modality on the latency and amplitude of P3, we will discuss how the results of this study replicated classic P3 effects in order to provide support for the validity of observed P3 measures. The expected distribution of the P3 across the scalp has been observed, with the P3 amplitude generally reaching its maximum at the posterior-midline electrode (i.e., Pz) and getting progressively smaller as it moved from posterior to frontal electrodes. Recall, the amplitude of the P3 reflects the amount of resources available to perform stimulus categorization during the WM processing requirements (Watter et al., 2001). Thus, higher WM loads are expected to yield smaller P3

---

[7] Although we did not obtain a significant Age × Modality interaction, inspection of Figures 1 and 2 suggests that the performance of older adults in the AV condition was similar to that of young adults in the A-only condition. To test this possibility, we computed a post hoc *t*-test between these two cells in our design and, indeed, the mean performances did not differ reliably for either d' or RT in any of the loads (all ps $> .05$). Nevertheless, this result should be interpreted cautiously because it is a null effect based on a post hoc analysis.

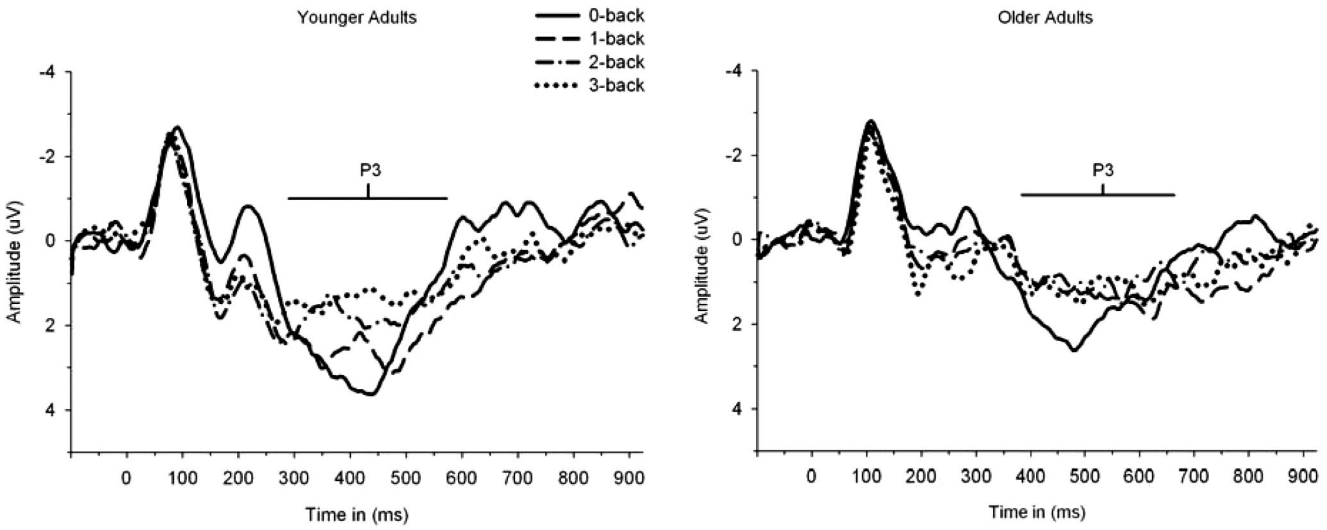

Figure 4. Grand average waveform of the P3 at the Pz electrode for younger adults (left panel) and older adults (right panel) during 0-back, 1-back, 2-back, and 3-back conditions collapsed across different modality conditions. The broad positive peak occurring between 300 and 600 ms after the onset of stimulus (i.e., 0 ms) represents the P3. Note the reduction in P3 amplitude in the higher working memory load conditions (2-back and 3-back) compared with the lower working memory load conditions (0-back and 1-back) in both age groups.

amplitudes, as processing resources are drawn away from stimulus categorization in order to support WM. The P3 in this study was smaller in high WM demand conditions (2-and 3-back) compared with low WM demand conditions (0-and 1-back), demonstrating this basic effect.

In terms of participant age, the P3 amplitude was smaller in older compared with younger adults, replicating the results of previous studies (e.g., Friedman, Simpson, & Hamberger, 1993; Vesco, Bone, Ryan, & Polich, 1993; Walhovd, Rosquist, & Fjell, 2008). Importantly, the P3 was affected in a similar manner by the task requirements in both younger and older adults, as indicated by the fact that increasing WM load affected the latency and amplitude of the P3 similarly in both age groups.

Table 5
Mean Latency (Ms) and Standard Deviations (in Parenthesis) of P3 at the Pz Electrode Site in Younger and Older Adults

	Modality	
Load	A-only	AV
Younger adults		
0-back	486.5 (68.86)	428.3 (90.03)
1-back	494.6 (76.17)	439.3 (110.70)
2-back	484.3 (86.21)	432.0 (98.31)
3-back	462.8 (78.41)	428.4 (101.06)
Older adults		
0-back	501.3 (90.76)	441.9 (70.55)
1-back	496.6 (86.40)	495.0 (119.97)
2-back	496.8 (89.03)	477.6 (88.79)
3-back	500.3 (98.10)	473.4 (107.09)

Note. A-only = auditory-only; A+V = auditory + visual; AV = auditory–visual.

With respect to our prediction that AV speech presentation would lead to better WM performance, as reflected by electrophysiological data, a modality effect was observed for P3 latency but not P3 amplitude. The P3 occurred earlier in AV compared with the A-only condition, suggesting faster cognitive processing during AV presentation. This may be because faster sensory processing, as evidenced by an earlier N1 in the AV condition, allows stimulus categorization processes to take place more quickly, leading to an earlier P3.[8]

If perceptual and higher order cognitive functions do indeed share processing resources, then the facilitation of one function should leave more resources available for higher order functions. We directly tested this hypothesis by examining the relationship between the facilitation of perceptual processing during AV speech presentation and improvement in WM performance. Recall that the facilitation of perceptual processing was reflected by a reduction in the amplitude of the auditory N1 during the AV condition compared with the A-only condition. We hypothesized that this reduction would be correlated with better performance on the WM task (defined as more accurate and/or faster responding), and with faster and less effortful higher-order processing (reflected by a shorter latency and/or larger amplitude of P3) in the AV condition relative to the A-only condition. Some support for these predictions was provided by the positive relationship between the reduction in N1 amplitude and (a) an improvement in accuracy in 2-back condition for younger adults and a marginal effect in 3-back condition for older adults, and (b) an improvement in the processing speed (RT and P3 latency) in the 1-back and 2-back conditions for young adults. Although supportive, these results

---

[8] This notion is in agreement with a significant correlation between P3 latency and RT during the AV condition (r = .34, p <.001).

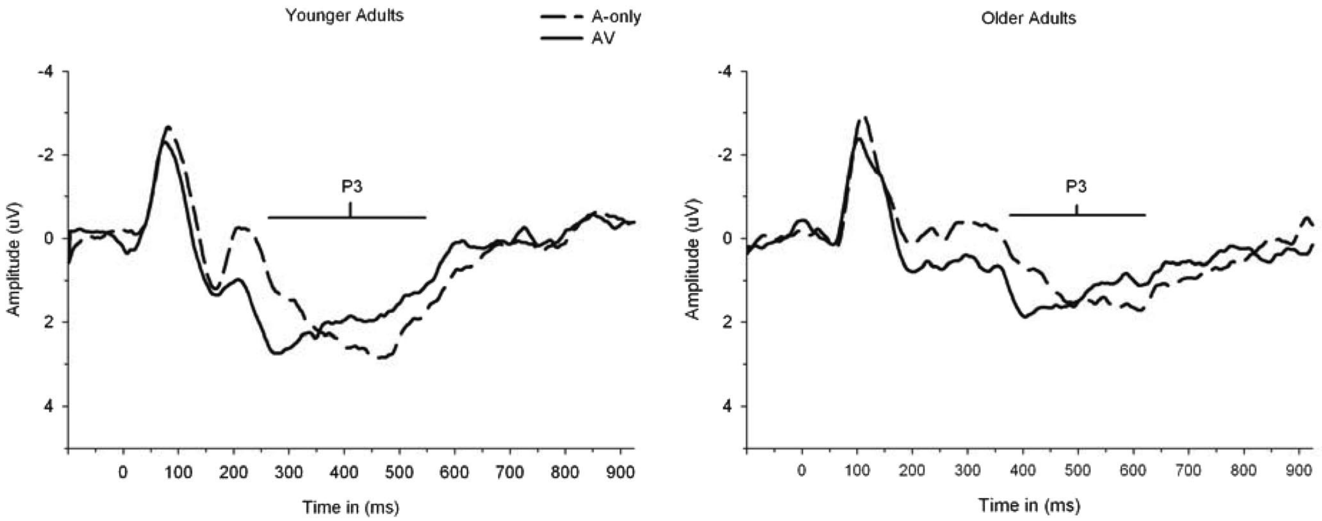

Figure 5. Grand average waveform of the P3 at the Pz electrode site for younger adults (left panel) and older adults (right panel) during auditory (A-only) and auditory–visual (AV) conditions collapsed across memory load conditions. The broad positive peak occurring between 300 and 600 ms after the onset of stimulus (i.e., 0 ms), represents the P3. Note that the P3 peaked earlier during AV condition compared with A-only condition in both age groups and earlier in younger than in older adults.

must be taken with caution, as some are statistical trends and will require replication with a larger sample size and appropriate control for Type I error. Nevertheless, these initial results suggest that fewer neural resources were allocated to auditory processing in the AV mode, increasing the efficiency of perceptual processing and leaving more resources available for later WM processing.

The current study had several limitations that need to be considered and addressed in future research. First, as mentioned, we did not fully replicate our previous N1 amplitude reduction in younger adults (Winneke & Phillips, 2011), probably due to the clear audibility of the stimuli. However, we have subsequently demonstrated that when younger adults are tested under background noise, we observe the facilitation effect of N1 amplitude during the AV condition relative to the A+V waveform (as in the current study) as well as the A-only waveform (Frtusova et al., 2011). Second, the present sample consists of relatively high-functioning older adults with minimal sensory loss. Future studies could examine the generalizability of the findings to a broader population of older adults, especially those who have experienced larger sensory decline or who have cognitive impairments as these populations may benefit from AV speech to an even higher extent.

Table 6
Zero-Order Correlations for Young and Older Adults Between the Auditory N1 Amplitude Reduction (A-Only Minus AV) and Improvement (AV Versus A-Only) in WM Indices (d′, RT, P3 amplitude, P3 Latency)

	Accuracy (*d*′)	RT	P3 amplitude	P3 latency
Younger adults				
0-back	.06	−.08	−.30	.31
1-back	.27	.37[a]	.30	.36[a]
2-back	.36[a]	.52[b]	.18	.34[c]
3-back	.26	.24	−.11	.19
Older adults				
0-back	.25	−.12	−.07	−.19
1-back	.04	.17	.27	−.03
2-back	.14	.07	−.17	.13
3-back	.35[c]	.32	.13	.14

Note. A-only = auditory-only; A+V = auditory + visual; AV = auditory–visual; RT = reaction time; WM = working memory.
[a] Value significant at $\alpha = .05$ (one-tailed). [b] Value significant at $\alpha = .01$ (one-tailed). [c] A trend toward significant value with $\alpha \leq .07$ (one-tailed).

Conclusions and Implications

This study is the first to report how AV speech presentation benefits WM performance in younger and older adults, using a well-established WM task and concurrently recorded ERPs. The study demonstrated that the presence of bimodal AV speech cues help younger and older adults to enhance their WM performance relative to clear, easily perceptible A-only input. The AV presentation mode appeared to improve the WM performance of older adults such that it no longer differed from that of younger adults (in the A-only condition). Moreover, as the load on WM increased, the benefits that younger and older adults derived from AV speech became even more prominent. Finally, we demonstrated that the benefit on WM performance was related to the degree of AV perceptual facilitation, as measured by auditory electrophysiological responses as early as 100 ms after stimulus onset. These results provide support for an integrated information-processing model (e.g., Schneider & Pichora-Fuller, 2000), in which processing resources are shared between perceptual and cognitive systems. The implications are especially important for older adults who experience mild changes in both perceptual processing and WM, and suggest that multimodal communication is of particular benefit in this population for a variety of reasons.

References

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). New York, NY: Academic Press.

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M. H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *The Journal of Neuroscience, 28,* 14301–14310. doi:10.1523/JNEUROSCI.2875-08.2008

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.30) [Computer software. Amsterdam, The Netherlands: Author.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 60,* 223–233. doi:10.1093/geronb/60.5.P223

Brault, L. M., Gilbert, J. L., Lansing, C. R., McCarley, J. S., & Kramer, A. F. (2010). Bimodal stimulus presentation and expanded auditory bandwidth improve older adults' speech perception. *Human Factors, 52,* 479–491. doi:10.1177/0018720810380404

Burg, A. (1966). Visual acuity as measured by dynamic and static tests: A comparative evaluation. *Journal of Applied Psychology, 50,* 460–466. doi:10.1037/h0023982

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

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

Compumedics Neuroscan. (2003). Scan (Version 4.3.1) [Computer software]. El Paso, TX: Author.

Diamond, B. J., DeLuca, J., Rosenthal, D., Vlad, R., Davis, K., Lucas, G., . . . Richards, J. A. (2000). Information processing in older versus younger adults: Accuracy versus speed. *International Journal of Rehabilitation & Health, 5,* 55–64. doi:10.1023/A:1012911203468

Eggermont, J. 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

Friedman, D., Kazmerski, V., & Fabiani, M. (1997). An overview of age-related changes in the scalp distribution of P3b. *Electroencephalography & Clinical Neurophysiology: Evoked Potentials, 104,* 498–513. doi:10.1016/S0168-5597(97)00036-1

Friedman, D., Simpson, G., & Hamberger, M. (1993). Age related changes in scalp topography to novel and target stimuli. *Psychophysiology, 30,* 383–396. doi:10.1111/j.1469-8986.1993.tb02060.x

Frtusova, J. B., Amarsi, J., & Phillips, N. A. (2011). Sensory processing affects working memory performance in noisy environment: An event-related potential study. *Psychophysiology, 48,* S107. doi:10.1111/j.1469-8986.2011.01259.x

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10,* 829–839. doi:10.1093/cercor/10.9.829

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory–visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory–visual integration. *Journal of Acoustical Society of America, 103,* 2677–2690. doi:10.1121/1.422788

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 193–225). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60041-9

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

Hockey, A., & Geffen, G. (2004). The concurrent validity and test–retest reliability of a visuospatial working memory task. *Intelligence, 32,* 591–605. doi:10.1016/j.intell.2004.07.009

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalograpy and Clinical Neurophysiology, 10,* 371–375. doi:10.1016/0013-4694(58)90053-1

Jennings, J. R. (1987). Methodology: Editorial policy on analyses of variance with repeated measures. *Psychophysiology, 24,* 474–475. doi:10.1111/j.1469-8986.1987.tb00320.x

Katz, J. (Ed.), 1985). Handbook of clinical audiology (3rd ed.) Baltimore, MD: Williams & Wilkins.

Kayser, J. (2003). Polygraphic Recording Data Exchange – PolyRex (Version 1.2) [Computer software]. New York, NY: New York State Psychiatric Institute, Department of Biopsychology. Retrieved from http://psychophysiology.cpmc.columbia.edu/PolyRex.htm

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38,* 557–577. doi:10.1017/S0048577201990559

Mattay, V. S., Fera, F., Tessitore, A., Hariri, A. R., Berman, K. F., Das, S., . . . Weinberger, D. R. (2006). Neurophysiological correlates of age-related changes in working memory capacity. *Neuroscience Letters, 392,* 32–37. doi:10.1016/j.neulet.2005.09.025

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58,* 22–33. doi:10.1080/02724980443000151

Millisecond Software. (2008). Inquisit (Version 2.0) [Computer software]. Seattle, WA: Author.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139174909

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the 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 Geriatrics Society, 53,* 695–699. doi:10.1111/j.1532-5415.2005.53221.x

Nyberg, L., Dahlin, E., Stigsdotter Neely, A., & Bäckman, L. (2009). Neural correlates of variable working memory load across adult age and skill: Dissociative patterns within the fronto-parietal network. *Scandinavian Journal of Psychology, 50,* 41–46. doi:10.1111/j.1467-9450.2008.00678.x

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research, 23,* 689–699. doi:10.1016/0042-6989(83)90210-9

Pichora-Fuller, M. K. (1996). Working memory and speechreading. In D. Stork & M. Hennecke (Eds.), *Speech reading by humans and machines: Models, systems and applications* (pp. 257–274). Berlin, Germany: Springer-Verlag.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42,* 59–67. doi:10.3109/14992020309074625

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,* 593–608. doi:10.1121/1.412282

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A stimulation of auditory aging. *Hearing Research, 223,* 114–121. doi:10.1016/j.heares.2006.10.009

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118,* 2128–2148. doi:10.1016/j.clinph.2007.04.019

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *The 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

Salthouse, T. A. (1988). Resource-reduction interpretations of cognitive aging. *Developmental Review, 8,* 238–272. doi:10.1016/0273-2297(88)90006-8

Schneider, B. (1997). Psychoacoustics and aging: Implications for everyday listening. *Journal of Speech-Language Pathology and Audiology, 21,* 111–124.

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

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age related changes in temporal processing: Implications for speech perception. *Seminars in Hearing, 22,* 227–240. doi:10.1055/s-2001-15628

Segalowitz, S. J., Wintink, A. J., & Cudmore, L. J. (2001). P3 topographical change with task familiarization and task complexity. *Cognitive Brain Research, 12,* 451–457. doi:10.1016/S0926-6410(01)00082-9

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

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

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–50). Hillsdale, NJ: Lawrence Erlbaum.

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

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging, 24,* 761–766. doi:10.1037/a0014802

Van Gerven, P. W., Meijer, W. A., Prickaerts, J. H., & van der Veen, F. M. (2008). Aging and focus switching in working memory: Excluding the potential role of memory load. *Experimental Aging Research, 34,* 367–378. doi:10.1080/03610730802274165

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *PNAS Proceedings of the National Academy of Science of the United States of America, 102,* 1181–1186. doi:10.1073/pnas.0408949102

Vaughan, L., Basak, C., Hartman, M., & Verhaeghen (2008). Aging and working memory inside and outside the focus of attention: Dissociation of availability and accessibility. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology, and Cognition, 15,* 703–724. doi:10.1080/13825580802061645

Verhaeghen, P., & Basak, C. (2005). Aging and switching of the focus of attention in working memory: Results from modified n-back task. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58,* 134–154. doi:10.1080/02724980443000241

Vesco, K. K., Bone, R. C., Ryan, J. C., & Polich, J. (1993). P300 in young and elderly subjects: Auditory frequency and intensity effects. *Electroencephalography and Clinical Neurophysiology/Evoked Potential Section, 88,* 302–308.

Walhovd, K. B., Rosquist, H., & Fjell, A. M. (2008). P300 amplitude age reductions are not caused by latency jitter. *Psychophysiology, 45,* 545–553. doi:10.1111/j.1469-8986.2008.00661.x

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology, 38,* 998–1003. doi:10.1111/1469-8986.3860998

Winneke, A. H., & Phillips, N. A. (2011). Does audio-visual speech offer a fountain of youth for old years? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology and Aging, 26,* 427–438. doi:10.1037/a0021683

Received July 12, 2012
Revision received November 6, 2012
Accepted November 9, 2012 ■