The impact of auditory-visual speech perception on working memory

Jana B. Frtusova

A Thesis

in

The Department

of

Psychology

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Arts (Clinical Psychology) at
Concordia University
Montreal, Quebec, Canada

© Jana B. Frtusova

Library and Archives Canada
Published Heritage Branch
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada
Direction du Patrimoine de l'édition
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-71015-9
Our file Notre référence
ISBN: 978-0-494-71015-9

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Contributions of Authors

The study was conceived by Jana B. Frtusova, Dr. Axel Winneke, and Dr. Natalie Phillips. Jana B. Frtusova recorded the stimuli, recruited the participants, designed and conducted the experiment, processed the data, analyzed and interpreted the results. All this with help and under supervision of Dr. Natalie Phillips. Dr. Natalie Phillips and Dr. Axel Winneke came up with the idea for the research project and helped with the conceptualization of the study.

ABSTRACT

Adding visual speech information (i.e. lip movements) to auditory speech information (i.e. voice) can enhance speech comprehension in younger and older adults while at the same time it reduces electrical brain responses, as measured by event-related potentials (ERPs). Thus, the brain seems to allocate fewer resources to speech comprehension when audio-visual (AV) speech information is available. This study examined whether the brain resources saved at the perceptual level during AV presentation allow younger and older adults to perform better on a working memory task, and whether older adults benefit to the same extent as younger adults. Twenty older adults and 23 younger adults completed an n-back working memory task (0-, 1-, 2-, 3-back) under visual-only (V-only), auditory-only (A-only), and AV condition while ERPs were recorded. The results showed a decrease in reaction time across all memory loads and an improvement in accuracy for 2-back and 3-back during AV compared to the V-only and A-only conditions. In addition, ERP analysis from a sample of 12 younger and 12 older adults showed a smaller N1 amplitude for the older group during AV compared to A-only presentation. The attenuation of N1, however, did not correlate with behavioural data. Nor did it show a relationship with changes either in the latency or the amplitude of P3, an ERP that reflects working memory processes. Thus, despite clear behavioural improvements on the working memory task during AV speech presentation, a more direct relationship between facilitation of sensory processing and working memory improvement was not identified.

Acknowledgements

I would like to thank Dr. Natalie Phillips for her incredible support during this project, and for the patience with which she has helped me to learn about the new and exciting area of brain imaging research. I would also like to thank her, as well as Dr. Axel Winneke, for their insights in the conceptualization of this study and in resolving many technical obstacles encountered during the process of this research. This project could not have been completed without their help.

I would also like to thank all members of the Phillips Lab for their contribution of time during the piloting and testing, and sharing of knowledge during the whole process. Special thanks to Shanna Kousaie for sharing her knowledge of ERP technology, and Cristina McHenry, for her help with editing this paper.

Finally, I would like to thank the Natural Sciences and Engineering Research Council of Canada, Concordia University, and the Centre for Research in Human Development, for their financial and technical support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AD=	Alzheimer’s disease
ANOVA =	analysis of variance
A-only=	auditory-only
AV=	auditory-visual
cm=	centimeter
EEG=	electroencephalography
EOG=	electrooculogram
ERP=	event-related potential
MCI=	mild cognitive impairment
mm=	milimeter
ms=	millisecond
n=	number
OA=	older adult
RT=	reaction time
S/N=	signal-to-noise ratio
V-only=	visual-only
μV=	micro-volt
YA=	younger adults

The Impact of Auditory-Visual Speech Perception on Working Memory

When we interact with the outside world, sensory information enters our bodies through different sensory modalities. Sensations coming from different sensory organs can interact and affect our perceptual experiences and understanding of the outside world. For example, every time we taste food, the interaction between smell and taste allows us to perceive thousands of different flavours, such as strawberry or coffee, despite the fact that only a few taste sensations can be perceived directly from the tongue. Thus, the interaction between different sensations is a part of our everyday experience and contributes to the richness of our surroundings. Speech also plays an important role in our lives. It allows us to learn and communicate with others and helps to prevent social isolation. This same type of multi-sensory interaction affects our perception of speech. During auditory-visual speech (AV), where we can both hear and see a person speaking, the visual speech information (i.e., lip, tongue, and face movements) interacts with auditory speech sounds and affects the content, as well as the quality, of what we hear (e.g., McGurk & MacDonald, 1976; Sumby & Pollack, 1954).

In addition to sensory processing, higher order functioning, such as working memory, also affects our speech comprehension (Just & Carpenter, 1992). For example, when talking to other people we need to perceive the spoken words, understand the meaning of each word and, based on the grammatical structure, reach a decision regarding the meaning of the sentence as a whole. Despite the fact that auditory and visual sensory processing, as well as memory, are recognized to play an important role in speech comprehension (Grant, Walden, & Seitz, 1998), the relationship between multisensory integration and working memory capacity has not received much attention in the research literature.

The purpose of the current project was to examine the behavioural and electrophysiological correlates between AV speech perception and working memory in younger and older adults. More specifically, this thesis examined the relationship between performance on a working memory n-back task and mode of speech presentation: visual-only (V-only), auditory-only (A-only), and AV. The n-back task is a working memory task, in which participants need to decide whether the currently presented stimulus matches the one presented n-trials before (e.g., 1 trial before in 1-back or 2-trials before in 2-back). In addition, the relationship between electrical brain responses and behavioural measures was explored with the purpose of examining the reallocation of resources from sensory processes toward working memory performance.

An overview of the literature begins with a discussion of basic findings relevant to AV speech perception, including studies that have used behavioural and/or electrophysiological measures. A review of studies that concentrates on assessment of working memory performance under different sensory modalities, and a discussion presenting the findings on speech perception in older population, follows. The section concludes with the presentation of the rationale, a brief description of the design, and the hypotheses for the results of the current study.

AV Speech Perception

The AV interaction can be seen as the influence of visual information on auditory perception (Tuomainen, Andersen, Tiippana, & Sams, 2005). Even though the AV interaction has been demonstrated for both speech and non-speech stimuli (e.g., Stekelenburg & Vroomen, 2007; Tuomainen, et al., 2005; Winneke & Phillips, submitted), it has been proposed that the interaction between auditory and visual speech cues differs from the interaction of auditory and visual non-speech cues. In other words, it has been suggested that AV speech is special and represents a specific mode of perception (Tuomainen et al., 2005). Others have argued against this position, suggesting that the basic mechanisms behind AV speech and non-speech perception do not differ (e.g., Stekelenburg & Vroomen, 2007; Winneke & Phillips, submitted). Since this argument has not yet been resolved, and given the purpose of the current project, the following discussion concentrates mainly on studies that have specifically assessed AV speech perception.

The effect of AV interaction on speech perception is clearly evident when there is incongruence between the information coming from auditory and visual modalities. This was demonstrated by McGurk and MacDonald (1976) who showed that simultaneously presenting one syllable auditorily (e.g., hearing someone say /ba/) and another syllable visually (e.g., seeing someone saying /ga/) results in the perception of an entirely different syllable (in this example /da/). This illusionary effect has become known as the McGurk effect and illustrates that although audition can be considered a primary modality in speech (Easton & Basala, 1982), vision also plays an important role.

The McGurk effect nicely demonstrates the strong effect of visual cues on speech comprehension but it is not a common experience to encounter incongruent auditory and visual speech information in our every day lives. In contrast, we are often required to comprehend speech under suboptimal conditions, such as when we are trying to have a conversation on busy streets or in restaurants full of other people talking. Research has demonstrated that adding visual speech information to auditory speech information can substantially enhance speech comprehension in the suboptimal conditions created by background noise (e.g., Callan et al., 2003; Schwartz, Berthommier, & Savariaux, 2004; Sommers, Tye-Murray, & Spehar, 2005; Sumby & Pollack, 1954; Summerfield, 1979) or auditory impairment (e.g., Bergeson & Pisoni, 2004; Grant et al., 1998; Rouger et al., 2007). One explanation for the AV enhancement effect in speech comprehension is that visual cues provide additional information that helps to resolve the ambiguity of acoustically similar phonemes, such as /v/ and /b/, by delivering information about the place of articulation (Summerfield, 1987). That is, observing the organization of upper and lower lips, and placement of the tongue helps to distinguish between the phonemes that sound similar. In addition to information derived from the lips and tongue, availability of other visual information, such as eyes, forehead, and head movements, also influences speech perception (Davis & Kim, 2006; Munhall & Vatikiotis-Bateson, 1998). In fact, it has been shown that these types of visual information help improve comprehensibility of speech in noisy environments, possibly by providing cues for appropriate parsing of speech signals (Davis & Kim, 2006).

Despite numerous studies that have examined multisensory interaction during AV speech, the exact neural processes involved in this phenomenon remain unclear. Brain imaging research using functional magnetic resonance imaging and magnetoencephalography found that adding visual speech information affects the neural processing in the sensory specific cortices (i.e. auditory cortex) as well as in multimodal regions, such as the superior temporal sulcus and superior temporal gyrus (Calvert et al., 1999; Calvert, Campbell, & Brammer, 2000; Möttönen, Schürmann, & Sams, 2004; Sekiyama, Kanno, Miura, & Sugita, 2003; Wright, Pelphrey, Allison, McKeown, &

McCarthy, 2003). The question remains about where the interaction first occurs. According to one hypothesis, the interaction of AV speech information first occurs in multisensory regions and then modulates processing in the sensory specific areas through feedback projections (Calvert et al., 1999; Calvert et al., 2000; van Atteveldt, Formisano, Goebel, & Blomert, 2004). This assumption however, is not supported by the results from a magnetoencephalography study conducted by Möttönen et al. (2004) who investigated the time course of multisensory integration in different cortical regions and found that modulated responses in auditory cortices preceded the modulated responses in the superior temporal sulcus. The existence of direct connections between primary auditory and primary visual cortex observed in macaque monkeys (Clavagnier, Falchier, & Kennedy, 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002) provides further support for the possibility that AV interaction can occur in sensory specific cortices, even before the information reaches higher multisensory areas. Moreover, studies using single cell recordings in cats suggest that AV integration may even occur sub-cortically. For example, Stein & Meredith (1993) identified multisensory neurons that respond to AV stimuli in the midbrain, more specifically in the superior colliculus. Similar results have been reported for humans. Calvert, Hansen, Iversen, and Brammer (2001) examined the AV interaction using non-speech stimuli by measuring blood oxygenation level dependent responses and observed an interactive AV effect in the frontal areas, superior temporal sulcus, as well as in the superior colliculus. It is possible, however, that the interaction effect observed in the human midbrain is the result of feedback received from higher sensory areas rather that a sign of an early interaction.

One way to determine exactly when the interaction of AV speech information occurs is to use electroencephalography (EEG), a recording of ongoing electrical activity in the brain. Event related potentials (ERPs) are derived from EEG and they represent small changes in the electrical activity of the brain triggered by external or internal events. ERPs reflect discrete stages of processing, from sensory to cognitive levels, and they allow us to track the flow of information processing that occurs at these different levels. The ERPs allow detection of rapidly oscillating electrical activity from multiple locations on the scalp by using small electrodes placed around the head. The oscillating electrical brain activity is observed in waveforms with peaks and troughs that represent changes in the voltage. Cognitive processing is reflected in deflections of these waveforms. The measures that can be obtained from ERPs are: the amplitude of the wave (measured in μV), reflecting the amount of perceptual or cognitive processing involved in the task; the latency of the wave peak or trough (measured in ms), reflecting the timing of sensory or cognitive processing; and the topography of the wave, reflecting the distribution of electrical activity associated with sensory or cognitive processing across the scalp. Decades of research have associated different ERP components with specific cognitive functions, such as sensory processing. The components are usually described in terms of their polarity, with *N* referring to negativity and *P* referring to positivity, and the latency of the amplitude. For example, N100 refers to the negativity reaching the maximum amplitude at around 100 ms. Because ERP measures reflect discrete stages of processing, they allow us to determine the level of information processing, either sensory or cognitive, at which the AV benefit actually occurs. For more about general information on ERP methodology Handy (2005) or Luck (2005) is recommended. Before discussing the results of ERP studies examining AV speech, three

ERP components that were of interest in the current study and previous ERP research are described in more detail.

In response to the presentation of auditory stimuli, the brain elicits sensory-driven ERPs known as *P1* and *N1*. The auditory P1 is positivity that occurs around 50 ms after the onset of sound and is usually largest in frontocentral electrodes (Luck, 2005). The P1 is followed by N1, which is negativity consisting of several subcomponents (Luck, 2005; Näätänen & Picton, 1987). As suggested by Näätänen and Picton (1987) the first subcomponent is generated in the cortex of the supratemporal plane and peaks around 100 ms after stimulus onset, with the amplitude reaching its maximum in frontocentral locations. This component may be enhanced by attention and is affected by the intensity of the stimuli. That is, the higher the intensity, the higher the amplitude. The second subcomponent is biphasic, with the positive peak reaching its maximum at around 100 ms and the negative peak reaching its maximum at 150 ms. The generator of this subcomponent is considered to be the temporal gyrus and the maximum amplitude can be detected at mid-temporal electrodes. This subcomponent may be affected by factors such as expectancy of the stimuli (Näätänen & Picton, 1987). The third subcomponent is negativity peaking at around 100 ms. The exact location of the generator for this component is not known but Näätänen and Picton (1987) suggested that the areas of the cortex mainly responsible for motor functioning may be involved. This component reaches its maximum amplitude at the vertex and lateral central electrodes. Despite the fact that N1 reflects early sensory processing, it has been shown that its amplitude can be affected by the difficulty of target discrimination (Fitzgerald & Picton, 1984; Wang, Song, Qu, & Ding, 2010). The amplitude of N1 has been shown to decrease with increasingly challenging tasks.

The third component of interest in the current study is *P3*, which is a broad positive component occurring between 300 and 1000 ms after presentation of the stimulus (Luck, 2005; Friedman, Kazmerski, Fabiani, 1997). In the literature, two subcomponents of P3 are reported: P3a and P3b. The P3a is elicited by distinct, infrequent stimuli presented among frequent stimuli without the requirement of a response from the participant. In contrast, the P3b is elicited during task-relevant processing of stimuli (see review by Polich, 2007). In this paper, P3 always refers to the P3b subcomponent and no further distinction between these two terms will be made.

There is an increasing consensus that rather than being a unitary brain potential, the P3 reflects a summation of activity from different neural generators across the brain (Johnson, 1993; Picton, 1992), each related to processing of different information (Johnson, 1993). The P3 latency is thought to reflect timing of mental processes while the P3 amplitude is thought to reflect intensity of processing (Kok, 2001). The P3 latency is affected by the time required to categorize the stimulus and is thus affected by the quality of sensory processing (Luck, 2005). In comparison, the P3 amplitude is affected by factors such as probability of target stimulus, with an infrequent stimulus eliciting a larger P3 (Luck, 2005; Picton, 1992); attention, with the amplitude decreasing as attention is directed away from the eliciting stimuli (Kok, 2001); and task difficulty, with the amplitude of P3 decreasing with increased task difficulty (Kok, 2001). Considering that more difficult tasks require more attention, the conflicting effect of attention and task difficulty on the amplitude of P3 limits its utility as a measure of processing capacity and mental workload (Kok, 2001). In spite of this, several studies have validated P3 as a measure of working memory (e.g., Segalowitz, Wintink, & Cudmore, 2001; Singhal & Fowler, 2005; Watter, Geffen, & Geffen, 2001) showing that the amplitude of P3 decreases as working memory load increases. In terms of scalp distribution, the P3 usually reaches the maximum peak at the central posterior sites but with increased age this topography changes to more equipotent distribution across the scalp midline (Friedman et al., 1997).

Many studies have used the ERP method to examine AV interaction during speech perception (e.g., Besle et al., 2008; Besle, Fort, Delpuech, & Giard, 2004; Pilling, 2009; Reale et al., 2007; Stekelenburg & Vroomen, 2007; van Wassenhove, Grant, & Poeppel, 2005; Winneke & Phillips, submitted). Interestingly, despite the previously mentioned findings that visual cues improve comprehensibility of speech, these ERP studies found reduced (Besle et al., 2008; Besle, et al., 2004; Pilling, 2009; Reale et al., 2007; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005; Winneke & Phillips, submitted) and faster brain responses (e.g., van Wassenhove et al., 2005; Stekelenburg & Vroomen, 2007; Winneke & Phillips, submitted) during AV speech presentation. More specifically, compared to the amplitude and latency of N1 when the unimodal (i.e., A-only or V-only) speech stimuli are presented, the bimodal (AV) speech perception is associated with earlier onset and smaller amplitude of auditory N1. Thus, the brain seems to use fewer resources to process AV speech information even though it is able to produce better behavioural outcomes.

One series of studies that used ERPs to examine AV integration was conducted by van Wassenhove and her colleagues (2005). In one of the experiments, participants were presented with syllables in either A-only, V-only, or AV modality and they needed to decide which syllable they heard, saw or both heard and saw. The results showed that during AV presentation, N1 showed both temporal facilitation (i.e., N1 occurred earlier during AV speech than during A-only speech) and amplitude reduction (i.e., the amplitude of N1 was smaller during AV speech than during A-only speech), suggesting more efficient neural processing during AV speech perception.

It has been shown that adding visual speech stimuli to auditory speech sounds can particularly enhance speech comprehension in a noisy environment (Sumby & Pollack, 1954). Thus, Winneke and Phillips (submitted) examined the enhancement effect of AV modality at both behavioural and electrophysiological levels by using speech stimuli surrounded by a 20-speaker babble. In contrast to van Wassenhove et al. (2005), they used whole words (names of living and non-living objects such as tree or rock) rather than syllables, since words represent more ecologically valid speech stimuli. They presented the words in A-only, V-only, and AV conditions and the participants' task was to decide whether the presented word was the name of living or non-living object. The results showed that under the AV condition participants responded both faster and more accurately than in A-only or V-only condition alone. Behavioural improvements were accompanied by decreased amplitude and latency of the auditory N1 in the AV condition compared to the A-only condition as well as compared to summed responses elicited during the A-only and V-only conditions (A+V), confirming the results of van Wassenhove and colleagues' experiment. Importantly, both van Wassenhove and colleagues (2005) and Winneke and Phillips (submitted) showed that the amplitude of auditory N1 differs from the amplitude of summed responses in the unimodal conditions (A+V), suggesting that the N1 amplitude reduction during AV speech presentation reflects a genuine multisensory interaction, rather than an artefact resulting from processing two independent sensory stimuli.

Researchers have tried to identify what kind of information contained in visual stimuli was responsible for the earlier onset and suppression of auditory N1 during AV presentation. Van Wassenhove and her colleagues (2005) suggested that this temporal facilitation and amplitude reduction of auditory N1 in AV condition represent two distinct computational stages of multisensory integration. By looking at the relationship between correct identification in the V-only condition and changes in N1, they showed that the N1 latency facilitation in AV speech perception depends on the degree to which visual information (i.e., lip reading) allows prediction of subsequent auditory input. Thus, the more ambiguous the visual information was, the less it was associated with temporal facilitation of the auditory N1. On the other hand, the degree with which visual information can predict auditory information was not found to be associated with amplitude reduction. Rather, the decrease in the amplitude seems to support the deactivation hypothesis, which states that activation of one modality inhibits activation of the other modality, and this helps to reduce processing of redundant information conveyed in bimodal presentation of the same stimuli (van Wassenhove et al., 2005).

Another study that examined which information during AV speech is responsible for modulation and speeding of the auditory N1 was conducted by Stekelenburg and Vroomen (2007). They hypothesized that if visual stimuli predict what the person will hear then the changes in auditory N1 during AV speech should be present for congruent AV presentations (i.e., seeing the lips pronounce /fu/ paired with the sound of /fu/) but not for incongruent AV presentations (i.e., seeing the lips pronounce /bi/ paired with the sound of /fu/). Conversely, if the changes in auditory N1 happen because visual stimuli help to predict when the auditory information will occur, then the changes should occur when visual stimuli contain anticipatory motion (such as seeing somebody speak) but not if anticipatory motion is missing (such as seeing two hands hold and subsequently tear a sheet of the paper). The results showed that both congruent and incongruent AV conditions were associated with changes in auditory N1, but when the visual information did not include anticipatory motion no changes in auditory N1 were observed. Based on these results, the authors proposed that the earlier onset and suppression of the auditory N1 occurring during AV presentation are due to temporal information contained in visual stimuli (i.e., it helps to predict when the sound will occur). It is important, however, to notice that the stimuli lacking anticipatory motion in this study were not speech stimuli, and therefore there are limitations to the conclusions that can be derived from this study in regards to sensory processing that occurs during AV speech perception.

In conclusion, AV speech has been found to improve speech comprehension while at the same time enhancing neural processing , rendering it more efficient, both in terms of timing and amount of neural resources used. There are other factors, however, including working memory, which also affect speech perception. The following section discusses an interesting relationship between speech comprehension and working memory capacity, as well as the effect of AV speech on working memory.

AV Speech and Working Memory

In addition to being affected by the amount and quality of sensory information, speech comprehension is also highly affected by working memory capacity. Working memory allows for storage of propositions as well as integration of ideas across successive words, and it plays a role in syntactic analysis (Just & Carpenter, 1992). Hitch and Baddeley (1976) demonstrated that increasing working memory demands decreases the speed of language processing. In their task, participants needed to verify, as fast as possible, grammatical structure of sentences while saying nothing or simultaneously repeating either the word "the", or a counting sequence from 1 to 6, or a sequence of six random digits. The results showed that the articulation of simple words or simple sequences (i.e., counting from 1 to 6) did not interfere much with the language task. However, repeating a sequence of random digits, which requires more working memory capacity, slowed the language processing significantly. The effect was especially evident for more grammatically complex sentences. This supports Just and Carpenter's (1992) suggestion that both working memory and language comprehension use the same fund of resources for cognitive processing.

Research has demonstrated that providing AV speech information improves higher order processing such as memory. For instance, Pichora-Fuller (1996) demonstrated that when participants were allowed to both listen and watch a person speaking (AV condition), as compared to only listen (A-only condition), they could not only recognize more words from the speech, but they also recalled more words later on. Improvements in recognition were especially pronounced in conditions where contextual information was limited (low context sentences) and when speech was hard to understand due to low signal to noise ratio (S/N). Similarly, even though recall in the AV condition and A-only conditions were similar in high S/N, it was significantly better in the AV condition compared to the A-only condition in low S/N. Based on these findings, it may be assumed that in conditions where speech comprehension is challenged (i.e., a low S/N), more resources need to be allocated to comprehension, thereby decreasing the availability of resources for information storage. Thus, similarly to Baddeley and Hitch's (1974) findings, there seems to be a trade off effect between language comprehension and memory.

The AV speech benefits were also shown during immediate recall task. Thompson (1995) tested immediate recall of sentences in younger and older adults. The meaningful and non-meaningful sentences were presented in A-only, AV, and auditory + iconic gestures condition. The results showed that immediate recall of non-meaningful sentences was better in AV compared to A-only condition for both younger and older adults. For meaningful sentence, older but not younger adults benefited from AV compared to A-only speech, suggesting that visual speech cues can be especially beneficial for older adults.

It has been proposed that the advantage of bimodal presentation on recall may be explained in terms of dual code theory (Thomson & Paivio, 1994). According to this theory, items are coded, and can be retrieved, independently in verbal and non-verbal subsystems. Thus, items presented in the AV condition are easier to recall than items presented in V-only or A-only condition because they are encoded twice: once in non-verbal (i.e., visual) mode and once in verbal (i.e., auditory) mode (Mastroberardino, Santagelo, Botta, Marucci, & Olivetti Belardinelli, 2008). This suggestion came out of research done by Thomson and Paivio (1994) who examined the effect of modality on free recall. In their study, environmental stimuli were presented in three conditions: V-only (line drawing of living and non-living objects such as cat or car), A-only (sounds of living and non-living objects such as cat or car) and AV (line drawings and corresponding sounds presented simultaneously). The results showed that recall of AV items was better than recall of A-only or V-only items. In addition, the results showed an additive effect; the probability of recall for AV items was similar to expected probability of recall from both unimodal conditions combined. This supports the dual code theory proposition that each modality additively contributes to memory performance by improving encoding. The results of the same study also showed that two semantically identical (but not visually identical) pictures paired with corresponding sounds improve free recall more than two visually identical pictures paired with corresponding sounds. Since the first condition allows more distinct encoding than the second condition, it is possible that it is the distinctiveness of the two codes in different modalities that contributes to better recall in the AV condition.

According to dual code theory, different modalities help to create several memory traces of the semantically identical item. However, the dual code theory explanation for better recall in the AV condition has been challenged by observations that AV presentation improves working memory performance even if the auditory and visual stimuli are not semantically related. For example, Santegelo, Mastroberardino, Botta, Marucci, and Olivetti Belardineli (2006) found that when Chinese ideograms were paired with musical fragments (AV condition), the reaction time on a working memory n-back task improved compared to when ideograms (V-only condition) or musical fragments (A-only condition) were presented alone. Based on these results, the authors suggested that the improved performance in the AV condition could be due to an early multisensory facilitation process (Mastroberardino et al., 2008). This suggestion is consistent with previously mentioned observations that AV speech perception is associated with a lower ERP response at the sensory level, as measured by the reduced amplitude and latency of auditory N1 (Stekelenburg & Vroomen, 2007; van Wassenhove, 2005; Winneke & Phillips, submitted). Thus, it may be argued that providing AV stimulation facilitates sensory processing and "frees" resources that can be used for higher order cognitive tasks, such as working memory.

Many of the theories mentioned above are based on non-speech stimuli. In fact, there are only a handful of studies examining how AV speech presentation affects working memory and the results of these studies are inconsistent. Similarity to Pichora-Fuller's (1996) findings, Phillips et al. (2009) found that participants could recognize more previously presented words if they had been presented in the AV modality compared to A-only or V-only modalities. Pisoni, Saldaña, and Sheffert (1996) also found that AV speech can lead to improved memory performance. In their experiment, participants were presented with 30 lists, each containing 10 words. The lists were presented in either the A-only or AV modality. After each list, participants were asked to recall as many of the words from the list as they could. The results showed that AV presentation led to a better primacy effect. In other words, participants recalled more words that were first or second on the list in the AV condition than in the A-only condition. No differences were found between modality for the words presented later on in the lists. Similarly to Thomson and Paivio (1994), the authors of this study argued that AV speech leads to more elaborate encoding and transfer of information into long-term memory.

In their second study, Pisoni et al. (1996) investigated the AV speech benefit for an immediate short-term memory and interestingly, found that AV presentation can actually lead to interference with memory performance. In this experiment, participants were presented with sequences of letters, ranging from 4 to 9 letters, and the experimenters measured the number of correctly recalled sequences. The results showed that participants recalled more correct sequences in the A-only than in the AV condition. The authors suggested that during AV speech, visual and auditory information are processed separately, which creates higher demands on a common set of resources and negatively impacts working memory capacity. Altogether, Pisoni's results suggests that AV speech presentation helps information to be better encoded and transferred into long term memory but that this more elaborate encoding interferes with short term memory capacity.

Pisoni's et al. (1996) findings from the second study are intriguing for several reasons. First, according to Baddeley's model of working memory, visual and auditory information are processed by two independent processors: visual information by the visuo-spatial sketch-pad and the auditory information by the phonological loop (Baddeley, 1992). Since the two systems are independent, interference between them should not occur. In fact, it has been suggested that because of the independence between the two processors, material presented in more than one modality should increase working memory capacity (Mastroberardino et al., 2008). The second reason is that findings from brain imaging studies suggest facilitation rather than interference of processing during AV presentation compared to unimodal presentation (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005; Winneke & Phillips, submitted). Thirdly, the findings of Pisoni et al. (1996) contradict not only results which show memory enhancement when speech sounds are accompanied by articulatory information (Phillips et al., 2009; Pichora-Fuller, 1996) but also the results of studies that show memory enhancement when speech sounds are accompanied by written words or pictures. For example, Goolkasian and Foos (2005) asked participants to recall items presented either in unimodal conditions (picture, printed word, or spoken word) or bimodal conditions (combination of picture and printed word, picture and spoken word, spoken word and printed word). The results showed that participants recalled more words in bimodal conditions than in unimodal conditions, and they recalled more words in bimodal AV conditions (combination of a picture and spoken word or printed word and spoken word) than in bimodal visual conditions (combination of picture and written word). Thus once again, the results supported Thomson and Paivio's (1994) suggestions that bimodal stimulation contributes to better encoding, and that more distinctive codes (i.e. bimodal AV stimuli) lead to better encoding than less distinctive codes (bimodal visual stimuli).

The discrepancies between the results of studies assessing the effect of AV presentation on memory are hard to evaluate due to a relative lack of research in this area. Discrepancies may be the result of differences in task requirement (immediate as opposed to delayed recall) or differences in stimuli presentation (a combination of speech sounds and articulatory information compared to a combination of speech sounds with printed words or pictures). The current study used a working memory n-back task, which is a commonly used and well-researched working memory task. In this task, participants need to decide whether the currently presented stimulus matches the one presented 1, 2, or 3 trials before. Variants of this task include the visual-spatial n-back, in which participants need to remember the location of the stimulus, or the verbal n-back, in which participants need to remember the identity of the stimuli (e.g., digit or letter).

The validity of the n-back task as a measure of working memory ability was measured in the study by Gevins and Smith (2000). They found that higher accuracy and lower reaction time on the 0- and 2-back visual-spatial n-back correlated with higher scores on the digit span, a subtest from the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) that is used to assess working memory capacity. In addition, reaction time on the n-back task correlated with the overall IQ score on WAIS-R, suggesting that faster processing speed may play a role in general cognitive ability.

A different study by Hockey and Geffen (2004) investigated concurrent validity and test-retest reliability of the n-back task with 70 university students. Participants were tested twice with testing sessions one week apart, on a visual-spatial variant of the n-back task at 4 levels of difficulty (i.e., 0-,1-,2- and 3-back). Reaction time and accuracy on the n-back were correlated with scores on the computerized Multidimensional Aptitude Battery to see if the n-back task is a reliable and valid measure of the cognitive processes believed to underlie intelligence. The results showed high test-retest reliability for reaction time and moderate test-retest reliability for the accuracy scores. In addition, faster reaction time during 0-, 1- and 2-back was predictive of higher full-scale IQ scores, while higher accuracy during 2- and 3-back were predictive of performance IQ scores, suggesting that the n-back task is sensitive to individual differences in cognitive abilities. In addition, the results also suggest that faster processing during the n-back is associated with higher general cognitive ability, whereas greater working memory capacity is associated with higher accuracy on performance tasks. Furthermore, it seems that faster processing enhances performance on the n-back task at the easier levels but as the task gets more difficult, working memory capacity seems to supersede mental speed (Hockey & Geffen, 2004).

Another study, done by Kane, Conway, Miura and Colflesh (2007) found only a weak correlation between performance on a verbal n-back task (2- and 3-back) and another measure of working memory called operation span task (OSPAN), in which participants need to verify mathematical equations while also trying to remember random words for later recall. The n-back and OSPAN, however, accounted for independent variance on the scores of fluid intelligence measured by Ravens Advanced Progressive Matrices (Raven, Raven, & Court, 1998), suggesting that the two measures do not reflect the same underlying construct. This finding points to the generalizability of conclusions about working memory that may derive from studies using the n-back task. The main challenge remains that consensus about which processes working memory exactly involves has not been reached and therefore different tasks that measure the construct differ in terms of what they require participants to do.

Nevertheless, the n-back task continues to be widely used in experimental studies, including those using neuroimaging. Watter and colleagues (2001) used the n-back task to investigate ERPs related to cognitive processes (matching and memory) that are involved in this task. Their findings showed that as memory demands increased (higher n-back condition), the amplitude of P3 decreased, while the latency of P3 remained constant. The authors explained that the observed decrease in P3 amplitude is an indication of the relocation of attention and processing capacity from the matching task, toward the primary working memory task.

Similarly to Watter and colleagues' study (2001), Segalowitz and colleagues

(2001) found that as working memory load increased during the n-back task, the amplitude of P3 at the posterior-central electrodes decreased by about 1 μV per each successive load. Interestingly though, the amplitude of P3 at fronto-central electrodes increased as the memory load increased. The authors suggested that this is because more complex tasks require more controlled processing, which is executed by the frontal lobes. Thus, it seems that P3 at the posterior sites may reflect working memory capacity, whereas P3 at the frontal sites may reflect attention control.

The current study investigated whether AV speech presentation enhances performance on the working memory n-back task in younger and older adults. Before presenting the details of the experiment, however, I will discuss how AV speech processing is affected by aging.

AV Speech Perception and Aging

Many sensory and cognitive abilities that are important for AV speech perception decline with age (Schneider & Pichora-Fuller, 2000) and thus it is not surprising that declines in speech perception have been identified in older adults (CHABA, 1988). According to the Grant and colleagues' (1998) model, speech perception can be affected by auditory and visual sensory information as well as higher order cognitive functions, such as memory and linguistic abilities. Considering this model and the results of studies that have shown the influence of AV interaction (e.g., Callan et al., 2003; Sommers et al., 2005; Sumby & Pollack, 1954; van Wassenhove et al., 2005; Winneke & Phillips, submitted) and recognition memory (e.g.; Phillips et al., 2009; Pichora-Fuller, 1996) on speech perception, problems that older adults experience may result from inadequate encoding of auditory and visual speech information, inadequate integration of AV speech information, decline in memory and linguistic skills, or a combination of any of these factors (Phillips et al., 2009).

At the lowest level of processing, speech perception is affected by the quality of sensory speech information. Older adults often suffer from presbycusis, the age-related cochlear hearing loss that leads to an increase in hearing threshold for high-frequency sounds (Pichora-Fuller, 2003). Temporal processing is also affected by aging; hearing processing becomes slower and more asynchronous (Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007; Schneider & Pichora-Fuller, 2001). Both high-frequency sounds and timing play an important role in speech perception (e.g., Wingfield, Tun, & McCoy, 2005; Pichora-Fuller, 2003) and therefore the well-documented age-related decline in speech comprehension in noisy environments (e.g., Abel, Sass-Kortsak, & Naugler, 2000; Pichora-Fuller, Schneider, & Daneman, 1995; Tun & Wingfield, 1999) or when speech rate is increased (e.g., Wingfield, 1996) can be secondary to the decline in perceptual processing (Abel, et al. 2000; Pichora-Fuller, 2003). In addition to hearing decline, older adults also experience decline in visual abilities, including visual acuity and contrast sensitivity (e.g., Schneider & Pichora-Fuller, 2000), which may contribute to their poorer performance on speech-reading as compared to younger adults (e.g., Cienkowski & Carney, 2002; Sommers, et al., 2005).

Despite poorer performance on speech reading, older adults seem to benefit from AV speech presentation. For example, Sommers and colleagues (2005) found that after controlling for quality of sensory input, older adults were able to benefit from AV speech presentation to the same extent as younger adults. The quality of sensory input was controlled by matching participants on hearing threshold by adjusting the volume of background babble (i.e., changing S/N) to cause a 50% error rate in the A-only condition and then using the same S/N in the AV condition. Participants' task was to identify syllables, words, or sentences presented in A-only, V-only, or AV modalities and the purpose of the study was to examine the extent to which adding visual speech cues to auditory speech cues improves identification of the stimuli (i.e., the visual enhancement effect). Results showed that the visual enhancement effect was the same for younger and older adults, suggesting that older adults are able to successfully use visual speech cues despite their poorer performance on speech reading, as shown by their performance in the V-only condition.

Similar results have also been found in another study that used a syllable identification task (Cienkowski & Carney, 2002). The study looked at age differences in the McGurk effect, presenting both younger and older participants with syllables in A-only, V-only, or AV modalities. The results showed that the McGurk effect did not differ between younger and older adults, suggesting that both groups were able to fuse incongruent visual and auditory stimuli to a similar degree. Interestingly though, when participants failed to fuse the incongruent AV stimuli, younger adults seemed to rely more on auditory cues whereas older adults seemed to rely more on visual cues. It is important to mention that older adults tested in this study had higher hearing threshold than younger adults. Thus, it is possible that older adults relied more on visual cues because of unclear auditory input caused by a decline in hearing.

Similarly, in a study conducted by Thompson (1995), younger and older participants completed a syllable identification task in A-only or AV condition. The experimenters varied the ambiguity of the speech stimuli by synthesizing the syllables along the continuum from /ba/ to /da/. The results showed that the more ambiguous the stimuli were, the more both younger and older adults relied on visual cues, supporting the suggestion that visual speech information helps to resolve the ambiguity in spoken language (Summerfield, 1987). Importantly, older adults were found to rely on visual speech cues significantly more than younger adults.

The suggestion that older adults rely on visual cues to a higher extent than younger adults was further supported by the study conducted by Thompson and Malloy (2004). In their study, younger and older participants were presented with videos in which a speaker presented a story. During the presentation, dots appeared in several locations on the speaker's face, including eyes and mouth. The task was to mark the location of the dots as well as answer the questions about presented stories. The results showed that the difference between correctly identified dots in the mouth and eye areas was greater in older than in younger adults. The authors interpreted these results as evidence that during AV speech presentation, older adults allocate more of their limited attention resources to the mouth area compared to younger adults. Even though visual acuity and hearing screening were not performed in this study, it is possible that older adults allocated more attention to mouth area in order to use visual cues as a compensatory method to offset the decline in hearing.

The above mentioned studies provide support for the suggestion that older adults can successfully integrate auditory and visual speech information and benefit from AV presentation to the same extent as younger adults. However, the inverse-effectiveness hypothesis suggests that the benefit from multisensory presentation should increase as the effectiveness of each unisensory modality decreases (Stein & Meredith, 1993). Thus, according to this hypothesis, older adults should benefit from AV presentation even more than younger adults due to their decline in both auditory and visual sensory processing. Support for the inverse-effectiveness hypothesis was found by Laurienti, Burdette, Maldjian, and Wallace (2006) who investigated multisensory integration in younger and older adults using a discrimination task. Participants had to discriminate between a red and blue color that was presented in A-only (i.e., spoken color words), V-only (i.e., colored discs presented on the computer screen) or AV (i.e., combination of colored word with the colored disc) condition. The speed of responses made in the AV condition was compared to the summed probability of the unisensory responses, as predicted by the race model. The race model computes a predicted reaction time that would be obtained from the combination of two sensory modalities that do not interact with each other. This method is superior to comparing AV condition to each unisensory condition separately in that it allows exploration of the interaction between auditory and visual stimuli. Laurienti et al. (2006) showed that older adults benefited from AV presentation to a higher extent than younger adults, suggesting that multisensory presentation may be an effective strategy to compensate for age-related declines in unisensory processing.

Laurienti et al. (2006) used non-speech stimuli and thus the extent to which their results can be generalized to AV speech perception is limited. However, a partial support for the inverse-effectiveness hypothesis was found in one of the studies conducted in our laboratory (Winneke & Phillips, submitted). This study looked at speed and accuracy of younger and older adults who were asked to identify spoken words presented in A-only, V-only or AV modalities. In A-only and AV modalities, the words were presented in a multi-speaker background babble. In addition, the study also investigated neural processes underlying AV speech and how these processes differ between younger and older adults by measuring ERPs (i.e., the latency and the size of amplitude for auditory N1). The results showed that both younger and older adults responded faster and more accurately in AV condition compared to A-only or V-only condition. In terms of speed, both age groups benefited from AV presentation to the same extent. In terms of accuracy, however, older adults benefited from AV presentation, compared to V-only presentation, significantly more than younger adults. The improvement in the AV condition compared to A-only condition did not differ for the two age groups. In terms of electrophysiological measures, both groups showed a reduction in the amplitude of auditory N1 in the AV condition compared to the A-only condition and compared to summed responses in both unisensory conditions (i.e., A+V), suggesting that neural processing underlying AV speech perception does not differ among younger and older adults. However, examining the latency revealed that even though the auditory N1 occurred earlier for both groups in the AV condition compared to the A-only condition, the shift was significantly larger for older adults. Together with van Wassenhove's et al. (2005) proposition that the latency facilitation in AV speech perception depends on the degree to which visual information allows prediction of subsequent auditory input, these results suggest that older adults are more efficient than younger adults at extracting useful information from the visual cues and predicting subsequent auditory input.

Finally, according to the Grant and colleagues' (1998) model, speech perception is also affected by the top-down processes, including individual knowledge and working memory capacity. Individual knowledge includes word and world knowledge, language background, linguistic competence, reasoning ability, and processing speed. Some of these abilities, such as processing speed (e.g., Baudouin, Clarys, Vanneste, & Isingrini, 2009; Diamond et al., 2000) decline with age. On the other hand, skills such as word knowledge, as measured by vocabulary, remain stable or even improve as one grows older (e.g., Park et al., 2002; Verhaeghen, 2003). In addition, it has been shown that during word recognition tasks wherein the sentences are presented in background noise, older adults make better use of contextual information than younger adults (Pichora-Fuller et al., 1995). The authors suggest that this effect may occur because older adults have more experience in using contextual information since in everyday situations they are often exposed to suboptimal listening conditions, caused by the level of background noise, which would not be problematic for younger adults (Phillips et al., 2009; Pichora-Fuller et al., 1995).

In accordance with the limited resources hypothesis (Just & Carpenter, 1992), the more cognitive resources older adults invest into speech comprehension, the fewer cognitive resources they have for processing and storing the information conveyed in the speech. Thus, working memory capacity is adversely affected by increased demands at lower levels of speech processing (Pichora-Fuller, 2003). In other words, even though older adults may be able to use compensatory strategies, such as contextual information, to diminish the age-related decline in sensory processing, they need to put more effort into speech comprehension than younger adults. This more effortful listening affects their ability to remember what the speaker has said. This has been demonstrated in a study conducted by Pichora-Fuller and colleaques in 1995. In this experiment, younger and older adults were asked to identify and later recall final words for sentences that were either read or presented orally at different S/N levels. The results showed that even though both age groups recalled the same number of words when they read the sentences, older adults recalled fewer words than younger adults when the sentences were presented orally under all S/N conditions. This suggests that both age groups had similar working memory spans, but when older adults needed to invest more resources to auditory comprehension, their ability to store and retrieve spoken information was negatively affected.

Similar results have also been reported by Tun, McCoy and Wingfield (2009). In this experiment, four groups of participants (younger adults with good hearing, younger adults with poor hearing, older adults with good hearing, and older adults with poor hearing) were presented with lists of words and after each list they were asked to recall as many words as they could. The lists were presented at the intensity level that allowed all groups to correctly identify spoken words. In spite of this, older adults recalled fewer words than younger adults, and groups with poor hearing recalled fewer words than groups with good hearing. The authors interpreted the results as suggesting that the effortful listening caused by impaired hearing negatively affected the ability to remember information even though it was presented at a comprehensible level.

The previously mentioned finding of similar working memory spans between younger and older adults reported by Pichora-Fuller and colleagues (1995) is quite surprising considering the evidence for age-related decline in working memory that has been reported in other studies (e.g., Bopp & Verhaeghen, 2009; McCabe & Hartman, 2008; Salthouse, 1994; Vaughan, Basak, Hartman, & Verhaeghen, 2008). The discrepancy may be the result of unclear definitions of working memory and a broad range of tasks used to assess it in different studies. The requirements in different working memory tasks can range from simple retrieval of word-lists to the manipulating and sequencing of presented information (e.g., letter-number sequencing task). It was suggested that greater age-related declines may be seen in more complex working memory tasks, such as backwards memory span, compared to less complex working memory tasks, such as forward memory span (Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000).

Several studies have investigated age-related differences on a working memory n-back task, which is used in the current study. These studies found that older adults are generally slower and less accurate than younger adults (e.g., Vaughan et al., 2008) and that accuracy, but not reaction time, is disproportionally more affected when switching from 1-back to 2-back in older compared to younger adults (van Gerven, Meijer, Prickaerts, & van der Veen, 2008; Varhaeghen & Besak, 2005). The research also suggests that age-related differences in n-back performance can be attributed to larger costs of switching focal attention within working memory, rather than memory load difficulties (van Gerven et al., 2008).

In sum, older adults experience a decline in both sensory processing and working memory abilities. Despite this, they seem to be able to integrate and benefit from AV speech information to the same extent as younger adults. Even though no research has directly investigated the age difference in working memory during AV speech presentation compared to unisensory (V-only or A-only) presentation, it can be predicted that older adults could benefit from AV presentation during a working memory task to the same extent as younger adults. However, due to declines in sensory processing and working memory abilities, the overall performance on the n-back task can be expected to be poorer for older adults than for younger adults.

Present Study

Based on the observation that improvement in speech comprehension during AV speech presentation is accompanied by the consumption of fewer cognitive resources, as measured by the amplitude of the auditory N1 (van Wassenhove et al. 2005; Winneke & Phillips, submitted), the current study examined whether the brain is able to use the resources spared during sensory processing to facilitate a higher order cognitive task (i.e. working memory). In addition, the relocation of cognitive resources from sensory processing to working memory performance was compared between younger and older adults.

Performance on a working memory n-back task across 4 different memory loads (0-, 1-, 2-. 3-back) was measured in younger and older adults at behavioural and electrophysiological (using ERPs) levels. Working memory capacity was assessed by three behavioural measures obtained from the n-back task: reaction time, accuracy, and weighted n-back score, which was calculated by the multiplication of correct responses in 0-, 1-, 2-, 3-back memory load by 1, 2, 3, 4 respectively. The calculation of n-back scores is based on the assumption that correct responses in higher loads reflect better working memory capacity than correct responses at lower loads. Electrophysiological measures included the amplitude and latency of auditory N1 (reflecting sensory processing) and the amplitude and latency of P3 (reflecting working memory processing).

Based on the previous observations, it was hypothesized that when participants could both hear and see the person presenting digits they would be faster, more accurate, and have higher weighted n-back scores than when they could only hear or only see the person presenting digits. At the electrophysiological level, it was predicted that a larger decrease in the amplitude of auditory N1 between the AV and A-only conditions, and between the AV and A+V conditions would be associated with lower reaction time, higher accuracy, and a higher weighted n-back score. Further, it was also predicted that a larger decrease in the amplitude of N1 in the AV condition compared to A-only and A+V conditions would be correlated with greater amplitude and shorter latency of P3 in the AV condition. Thus, it was assumed that facilitation processes at the sensory level (as measured by a decrease in amplitude and latency of N1 in AV condition) can predict the capacity of working memory (as measured by amplitude and latency of P3 and behavioural observations).

In terms of memory load, it was predicted that as the memory load increases, the reaction time would be greater and accuracy would be lower. For the electrophysiological measures, it was predicted that as the memory load increases, the amplitude of P3 would decrease while the latency of P3 would stay the same (Watter et al., 2001).

In terms of age, it was predicted that older adults would benefit from AV presentation during the working memory task to the same extent as younger adults (Phillips et al., 2009). However, due to declines in sensory processing and working memory abilities, the overall performance on the n-back task was expected to be poorer for older adults than for younger adults in terms of accuracy, reaction time and weighted n-back score. More specifically, it was predicted that in the AV condition, compared to A-only or V-only conditions, older adults would show similar decreases in reaction time, and similar increases in accuracy and weighted n-back score as younger adults even though their overall reaction time would be greater, and accuracy and weighted n-back score would be lower than those of younger adults.

The interaction of age with memory load was also predicted for accuracy but not for reaction time. That is, older adults were expected to have accuracy similar to that of younger adults for lower memory loads that do not require much memory capacity (0-back and 1-back), but their accuracy was expected to be lower than that of younger adults for higher memory load conditions (2-back and 3-back) due to their decline in working memory capacity. In contrast, due to general decline in processing speed, the reaction time of older adults was expected to be lower that that of younger adults across all memory loads. These predictions are based on findings that even though older adults are generally slower than younger adults, they can be highly accurate (Diamond et al., 2000).

In terms of electrophysiological measures, it was predicted that older and younger adults would show similar decreases in the amplitude of auditory N1 in the AV condition compared to A-only and A+V conditions (Winneke & Phillips, submitted), and that this decrease would correlate with behavioural improvements as well as greater amplitude and shorter latency of P3 in AV condition.

Method

Participants

Twenty-seven younger adults and 23 older adults were recruited for the study.

Young participants were recruited through a participation pool in the Department of Psychology at Concordia University and through an advertisement posted on the McGill University web site. Older participants were recruited from an existing laboratory database. In order to be eligible for the experiment, participants had to meet the following requirements: first, their first language needed to be English. More specifically, they needed to have learned English before the age of 5 and use it as their primary language ever since. Second, they had to be right handed. Third, they had to be in good health. The health status was screened by a health questionnaire (see Appendix A) and exclusion criteria included: multiple risk factors for stroke, current depression, and past history of epilepsy, seizures, bypass surgery, head injury, coma, or neurological disorders that impact cognitive functioning. Fourth, all participants had to have normal or corrected to normal vision, and normal hearing. The vision screening consisted of the MARS Letter Contrast Sensitivity test (by MARS Perceptrix) and participants needed to score within the age-appropriate range as reported in Haymes and colleagues (2006). Specifically, all younger adults had to reach the contrast sensitivity score of 1.56 or higher, and all older adults had to reach the contrast sensitive score of 1.52 or higher in each eye. The auditory acuity screening consisted of measuring pure tone hearing thresholds for different frequencies (i.e., 250, 500, 1000, 2000, 4000 Hz) in both ears. In order to be included in the study, the mean of pure tone thresholds for 500, 1000, and 2000 Hz could not exceed -25 dB in either ear, and the hearing threshold in the left and right ear could not differ by more than 10 dB in more than 3 out of 5 frequencies (250, 500, 1000, 2000, 4000 Hz). Lastly, to screen for dementia and mild cognitive impairment, older adults were assessed using the Montreal Cognitive Assessment

(MoCA; Nasreddine et al., 2005). The MoCA was not administered to younger adults, since they were all 35 years old or younger and therefore cognitive decline in this age group had not been suspected.

Four younger adults and 3 older adults were excluded: 1 younger adult because she did not meet the language requirements, 1 younger adult and 2 older adults did not meet the visual screening requirements, 1 younger adult because of health screening requirements, 1 younger adult could not finish testing due to headache, and 1 older adult due to technical difficulties during the ERP recording. The final sample consisted of 23 younger and 20 older adults. Participants signed an informed consent (see Appendix B) and at the end of the experiment received a debriefing sheet, which briefly described the rationale and hypotheses of the study (see Appendix C). Participants were compensated for their time by either monetary value of 10 dollars per hour, or by receiving 3 participant pool credits (Concordia University students only). The study had been approved by the Concordia University research ethics board.

Demographic information about the final sample is presented in Table 1. The t-test analyses on demographic data showed that even though older adults were highly educated, they were less educated than the younger adults ($t(1, 41) = -3.3$; $p = .002$). Furthermore, despite the fact that all younger and older adults met the age-appropriate criteria for normal vision and hearing, older adults had lower contrast sensitivity than younger adults ($t(1, 41) = 4.2$; $p < .001$) as well as higher pure tone thresholds both in left ($t(1, 41) = 6.6$; $p < .001$) and right ($t(1, 41) = 7.7$; $p < .001$) ear.

Table 1

Demographic information describing means and standard deviations of younger (YA) and older (OA) adults who completed the study

	YAs (n = 23)	OAs (n = 20)
Age (Years)	24.35 (*SD* = 3.70)	69.10 (*SD* = 6.17)
Education (Years)	16.52 (*SD* = 2.11)	14.15 (*SD* = 2.60)
Cognitive Functioning (MoCA score)	N/A	27.45 (*SD* = 1.28)
Hearing: Right Ear (dB)	2.68 (*SD* = 3.72)	13.67 (*SD* = 5.56)
Hearing: Left Ear (dB)	2.86 (*SD* = 2.85)	13.08 (*SD* = 7.18)
Vision: Binocular*	1.76 (*SD* = 0.05)	1.69 (*SD* = 0.06)

* Score on Mars Letter Contrast Sensitivity Test

Stimuli

The stimuli consisted of 9 spoken digits (1, 2, 3, 4, 5, 6, 8, 9, 10) presented by a female speaker. Digit "seven" was omitted because it has two syllables and thus would be easily recognized among other digits. Stimuli were videotaped in the recording studio at the Department of Journalism at Concordia University. Each digit was presented as a short video clip edited in Adobe Premiere so that the onset of the face would occur 300 ms before the first obvious lip movement and the clip would end 300 ms after the lips stopped moving. The average lag time between the onset of lip movement and first sound across different digits was 395.33 ms (*SD* = 103.24; the exact lengths of the lags for each digit are reported in Appendix D). The average length of the whole trial was 2010 ms (*SD* =160 ms). The inter-stimuli period (between the end of one stimulus and the onset of the next stimulus) was 2400 ms. The auditory files were modified by Praat (version 5.1.30; Boersma & Weenink, 2010) to set the intensity of each digit at an approximately equal level (*M* = 74.51 dB, *SD* = .16 dB); the stimuli were also modified by using Adobe Audition (version 3.0) such that the person's voice occurred in the left channel while triggers signaling the onset of lip movement (i.e., visual triggers) and the onset of the speech signal (i.e., auditory triggers) occurred in the right channel. The left channel was then split into the left and right ear phones for binaural presentation; the triggers were not audible. The visual stimuli (120 x 93 mm) were presented on 15" CRT monitor positioned about 60 cm from the participant. The auditory stimuli were presented at the mean intensity of 70.2 dB (*SD* = 2.59 dB) using EARLINK tube ear inserts (Neuroscan, El Paso, TX, USA). Before each testing session, the calibration tone was used to assure that the stimuli were presented at the same intensity for each participant.

The AV trials contained both auditory and visual files and thus participants could both hear and see the person presenting digits. These files were subsequently modified to create A-only and V-only stimuli. For V-only stimuli, the auditory left channel was deleted while everything else stayed the same, and therefore participants could see but not hear the person speaking the digit. For the A-only condition, video frames were deleted, while everything else stayed the same, thus participants could hear but not see the person speaking the digit. During the A-only condition, a black screen with a fixation point in the middle was presented and participants were asked to keep looking at the fixation point. In all modality conditions, triggers signaling the onset of lip movement and the onset of the speech signal were present so that visual and auditory ERPs could be measured and compared across different modalities. Inqusit (version 2.0; Millisecond Software, 2008) was used for stimuli presentation.

Procedure

Each participant completed the n-back task under three conditions: A-only, V-only, and AV. In the A-only condition, participants heard a person speaking digits (but could not see the person), in the V-only condition participants saw the person speaking digits (but could not hear the voice), and in the AV condition participants could both hear and see the person speaking digits. The participants' task was to decide whether the currently presented digit matched the one that they were asked to remember (0-back condition) at the beginning of a block or to decide whether the currently-presented digit matched the digit presented in the previous trial (1-back condition), 2 trials before (in 2-back condition), or 3 trials before (in 3-back condition). There were 100 trials in each n-back condition and participants indicated their response by pressing a "Match" or a "Non Match" button on the computer mouse with left- and right-hand responses counterbalanced across participants. In 40% of the trials the target digit matched the currently presented digit (answer should be "Match") and in 60% of trials the target digit did not match the currently presented digit (answer should be "Non Match").

Before starting the experiment, participants practiced identifying the V-only stimuli (i.e., lip reading). First, the V-only stimuli were presented in order from 1 to 10 (except 7) and participants were asked to say aloud which digit was presented. Then, the V-only stimuli were presented in a random order, and again participants were asked to say aloud which digit was presented. Participants had to correctly identify all the digits in both the orderly and random presentation before proceeding with the experiment. If they made a mistake, the practice was repeated. This practice took about 1 minute and it never had to be repeated more than twice. Next, participants practiced the key assignment for "Match" and "Non Match" responses on the computer mouse. For this task, they were given a "target" digit at the beginning of the practice block, and they needed to decide whether the digits presented by the speaker matched the target digit or not. If they matched, they would press the "Match" button, and if they did not match they would press the "Non Match" button on the mouse. During the key assignment practice, AV stimuli were presented and thus the participant could both hear and see the speaker. There were 10 trials during this practice and the task was essentially identical to the AV 0-back condition. Following this practice the experiment would start. 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 buzzing noise whenever they made a mistake.

In different 0-back conditions, participants were asked to remember a digit (either 3 or 6 or 9), with different digits assigned to different sensory modalities and digit-modality combinations counterbalanced across participants. Thus, while one participant needed to remember digit 3 in the A-only condition, digit 6 in the V-only condition, and digits 9 in the AV condition, another participant needed to remember digit 9 in the A-only condition, digit 3 in the V-only condition, and digit 6 in the AV condition, etc. The counterbalancing was done to prevent any effect that may have been associated with a particular digit (i.e., some digits may be easier to lip read than others) to interact with the effect of modality presentation.

In 1-, 2-, and 3-back conditions, there were 5 "warm up" trials at the beginning of each block to allow participants to get used to the new n-back load as well as the new modality. Responses during warm up trials were not included in the data analyses. Each block consisted of a random order of digits but the number of "Match" and "Non Match" responses was controlled for such that the 40/60 ratio of "Match" to "Non Match" responses remained in each block. There were 3 different sequences of random digits and each sequence existed in each sensory modality. A participant heard different sequences in different sensory modalities and modality-sequence combinations were counterbalanced across participants. Each participant completed, in ascending order, 0-back, 1-back, 2-back, and 3-back condition. The order of modality conditions (A-only, V-only, and AV) was counterbalanced and participants were randomly assigned to groups with particular modality orders. The modality order was the same for each n-back condition and participants completed all 3 modality conditions before moving to the higher n-back condition. Each participant completed all 12 conditions (4 n-back A-only conditions, 4 n-back V-only conditions, and 4 n-back AV conditions).

Three behavioural measures were collected from participants' performance on the n-back task. First was the reaction time measured in milliseconds (ms) and operationalized as the amount of time between the onset of the auditory trigger and the participant's response[1]. Second was accuracy, or the number of correct responses (number of trials in which participants answered "Match" when the stimulus on the current trial matched the one presented *n* trial(s) before, and participants answered "Non Match" when the stimulus on the current trial did not match the one presented *n* trial(s) before). The final behavioural measure was the weighted n-back score operationalized as follows. For each correct answer, participants were given 1 point in the 0-back condition, 2 points in the 1-back condition, 3 points in the 2-back condition, and 4 points in the 3-back condition. The sum of the scores across different n-back conditions represented the participant's weighted n-back score for each modality. The weighted n-back scores was used to take into account that a correct response in higher memory loads required a greater working memory capacity than a correct response in lower memory loads. In other words, since higher n-back conditions were more demanding, a correct response reflected better working memory, and thus deserved more points, than a correct response in lower memory loads. Since the weighted n-back score already accounted for n-back load, it allowed direct comparison of modality conditions.

The data were analyzed using an analysis of variance (ANOVA) for repeated measures with Modality and Load as within subjects factors, and Age as a between subjects factors. Only correct "Match" responses were used in the analyses (see the next section for the rationale). Any response that occurred earlier than 200 ms was counted as incorrect due to the fact that such early responses could hardly reflect true cognitive processing and rather they may be either late responses to the previous trial or unintended responses.

While participants were performing the n-back working memory task, their electrophysiological measures were also recorded. The absolute amplitude of N1 (difference between the trough of P1 and the peak of N1) and the relative amplitude of P3 (the peak of P3 relative to the baseline) were measured and results were correlated with the three behavioural measures.

EEG Data Acquisition

The Biosemi ActiveTwo EEG system was used to measure the electrical brain activity while participants were performing the task. The brain activity was recorded from 72 channels arranged according to the International 10-20 system. Vertical and horizontal eye movements were monitored by electrooculograms (EOGs) with electrodes positioned both above and below the left eye and beside the outer canthi of each eye. EEG was recorded at the sampling rate of 2048 Hz, with the high pass filer of .16 Hz and low pass filter of 100 Hz. The files were then downsampled offline to 512 Hz in order to decrease the size of the files[2].

The BioSemi data format (i.e., BDF files) was converted to continuous data format (i.e., CNT files) using Polygraphic Recording Data Exchange (PolyRex; Kayser, 2003). Before the conversion, the EEG recordings were re-referenced offline in the PolyRex using the recordings from left and right earlobes. The subsequent analyses were performed using Scan (version 4.3.1) software (Compumedics Neuroscan, 2003) In order to obtain a sufficient number of trials, excessive ocular artefacts, such as eye blinks, were corrected before the files were further processed. This correction was done by identifying 60 to 80 blinks in each file; these blinks were then averaged and a template of a "typical" blink for a particular person was created. Based on this template, the program identified what values needed to be subtracted from each channel when the blink had occurred. The spatial filter was used to subtract appropriate values from each channel. The continuous EEG files were then divided into 1100 ms epochs, including a 100 ms prestimulus baseline interval.

The files were epoched to the onset of the auditory trigger. As mentioned previously, the auditory trigger signaled the onset of the speech signal and thus epoching files according to the auditory trigger allowed for the detection of auditory evoked potentials. The auditory trigger occurred at the same time point in all stimuli, even in V-only stimuli where no audible sound was present. This was done so that the ERP components occurring at exactly the same time could be compared across different modalities. In addition, in order to assess the effect of AV interaction, the amplitude and latency of the auditory N1 evoked during the AV condition was compared to the amplitude and latency of N1 in A+V condition (the sum of the auditory N1 evoked during A-only and V-only conditions). In order to obtain the sum, ERPs needed to be measured at the same time point in each modality.

After the files were epoched, they were baseline corrected according to pre-stimulus period (-100 to 0 ms before the auditory trigger). Next, the trials during which horizontal EOG activity exceeded +/- 75 μV were rejected. In addition, trials with artefacts exceeding +/- 100 μV in any of the active EEG electrodes around the centre of the head (i.e., Fz, F1, F2, FCz, FC1, FC2, Fz, F1, F2, CPz, Pz, P1, P2, POz, PO1, PO2, Oz, O1, O2) were also excluded if they occurred up to 100 ms before or up to 800 ms after the onset of the auditory trigger. The epoched files were then averaged separately for each modality and n-back condition, and separately for "Match" and "Non Match" trials. Since the P3 is affected by the frequency of the stimuli, and it is more prominent for less frequent stimuli, only "Match" responses from each condition were used in the analyses. The averages were then filtered offline over the range of 1-30 Hz, using a zero-phase-shift band pass filter.

Four measures obtained from the ERP components were of interest in the current study. These were: the amplitude of N1 and P3 (measured in μV), and the latency of N1 and P3 (measured in ms). The amplitude of N1 was measured as an absolute difference between the trough of P1 and the peak of N1. In comparison, the amplitude of P3 was measured as the amplitude of the peak relative to the 0 μV baseline. Latencies of both N1 and P3 were measured at the components' peaks relative to the onset of the auditory trigger.

Results

Behavioural Results

All behavioural data were analyzed by repeated measures ANOVAs. Significant main effects with more than two levels and significant interactions were further examined by pairwise comparisons. For within subject factors with more than 1 degree of freedom in the numerator, the Greenhouse-Geisser non-sphericity correction was used, and according to the convention for these types of analyses, the uncorrected degrees of freedom, mean square error (*MSE*), adjusted *p* values and Greenhouse-

Geisser epsilon (ε) are reported. Unless otherwise specified, all data presented are significant at $\alpha = .05$. In order to examine the effect of Modality, Load, and Age on accuracy, reaction time, and weighted n-back score, separate 3-way 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.

Accuracy. Accuracy represented the number of correct responses for "Match" trials. Before analyses, the outliers (1 younger adult during the A-only 0-back, 1 younger adult during the V-only and AV 2-back, and 1 older adult during the V-only 0-back) who scored below 3 SDs from the mean for a particular condition were replaced by group means[3]. The analyses revealed a main effect of Modality ($F(2, 82) = 14.1$; $MSE = 86.24$; $p < .001$; $\varepsilon = .88$); participants were significantly more accurate in AV compared to A-only or V-only, while the A-only and V-only conditions did not differ from each other. There was also a main effect of Load ($F(3, 123) = 145.6$; $MSE = 299.74$; $p < .001$; $\varepsilon = .59$). Pairwise comparisons revealed that with each increase in the memory load, the accuracy significantly decreased. Next, there was a significant main effect of Age ($F(1,41) = 5.4$, $MSE = 307.92$; $p = .03$) with older adults being less accurate than younger adults.

In addition, there was a significant interaction between Modality and Load ($F(6, 246) = 4.5$; $MSE = 85.55$; $p = .002$; $\varepsilon = .66$). In 0-back, there were no significant differences between modalities. In 1-back, participants were more accurate in A-only than in V-only, but the AV condition did not differ from the A-only or V-only conditions. In 2-back and 3-back, participants were more accurate in AV than in A-only and V-only, but the performance in unimodal conditions did not differ. These results indicate that AV presentation improves accuracy more in more demanding working memory tasks.

Lastly, there was a significant interaction between Load and Age ($F(3, 123) = 3.9$; $MSE = 299.74$; $p = .03$; $\varepsilon = .59$). Older adults performed with the same accuracy as younger adults in 0-back, 1-back, and 3-back but they were significantly less accurate in 2-back. These results together with exploration of group means (see Figure 1) suggest that 0-back and 1-back are relatively easy for both younger and older adults. In comparison, 2-back is more difficult for older than younger adults, and 3-back is difficult for both age groups. An interaction between Modality, Load, and Age was not significant ($F(6, 246) = 1.5$; $MSE = 85.55$; $p =.22$; $\varepsilon = .66$).

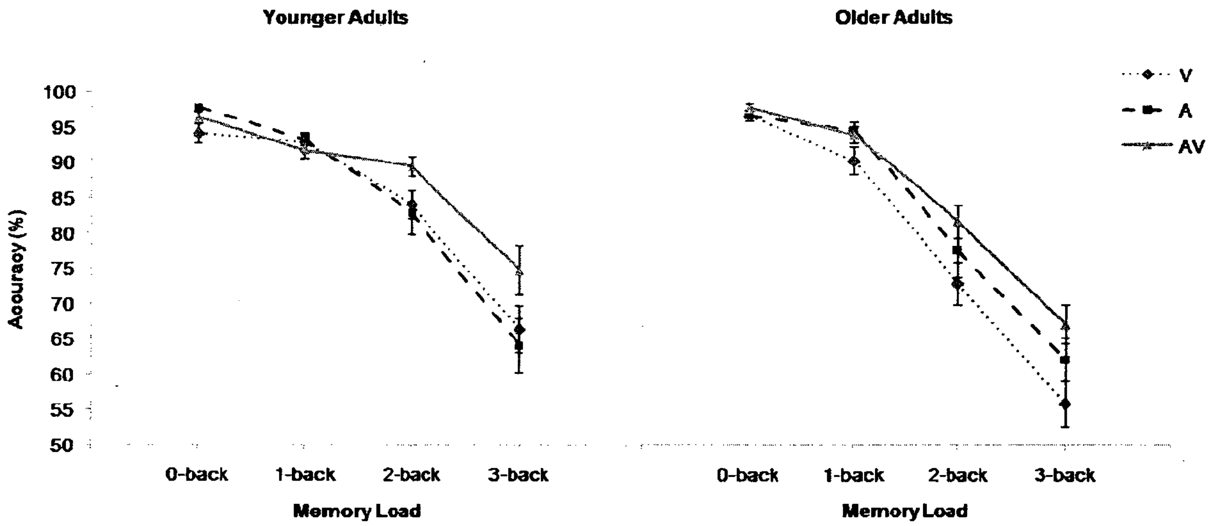


Figure 1: Mean percent of correct responses and standard error bars for younger and older adults in visual (V), auditory (A), and auditory-visual (AV) conditions.

Weighted n-back score. The weighted n-back score represented the number of correct responses for "Match" trials multiplied by 1, 2, 3 and 4 for 0-, 1-, 2- and 3-back respectively. Since the weighting of the scores already reflected memory load, scores were summed across n-back conditions, leaving a total score for each modality. The analyses revealed a main effect of Modality ($F(2, 82) = 14.7$; $MSE = 3731.04$; $p < .001$; $\varepsilon = .86$), with participants reaching a higher weighted n-back score in AV than in V-only or A-only; the V-only and A-only conditions did not differ. There was also a significant main effect of Age ($F(1,41) = 5.6$; $MSE = 15034.04$; $p = .02$), showing that older adults reached a lower n-back weighted score than younger adults. The interaction between Modality and Age was not significant ($F(2, 82) = 2.6$; $MSE = 3731.04$; $p = .09$; $\varepsilon = .86$), suggesting that Modality had a similar effect on the working memory performance in both age groups. See Figure 2.

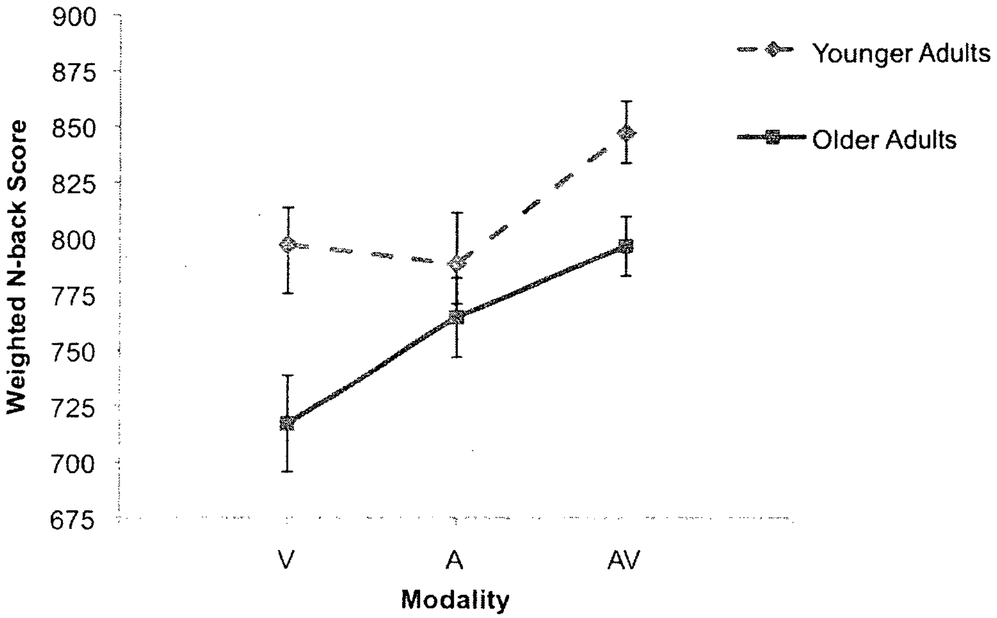


Figure 2: Mean weighted n-back scores and standard error bars for visual (V), auditory (A) and auditory-visual (AV) modality by younger and older adults.

Reaction time. The reaction time was measured as the time between the onset of the auditory trigger and the participant's response. Analyses revealed a significant main effect of Modality (*F*(2, 82)=45.8; *MSE*= 13020.03; *p*<.001; ε= .95); participants were faster in AV compared to A-only or V-only, and they were faster in A-only than in V-only. There was also a significant main effect of Load (*F*(3, 123)= 122.1; *MSE*= 27072.02; *p*< .001; ε= .71). Pairwise comparisons revealed that as the memory load increased, reaction time significantly increased across all memory load conditions. Next, there was a significant main effect of Age (*F*(1, 41)= 4.0; *MSE*= 133095.11; *p*= .05), with older adults being significantly slower than younger adults.

Furthermore, there was a significant interaction between Modality and Load (*F*(6, 246) = 8.3; *MSE* = 6675.78; *p* <.001; ε = .76). For 0-back and 1-back, participants were faster in AV than in A-only or V-only, and they were faster in A-only than in V-only. For 2-back and 3-back, participants were faster in AV than in A-only or V-only, but A-only and V-only conditions did not differ. These results suggest that in terms of reaction time, participants benefited from AV presentation at all levels of memory load. In contrast, even though they seemed to be faster in A-only than in V-only condition when demands on working memory were low, as the demands increased, the performance in the unimodal conditions became equal. The interaction between Modality, Load, and Age was not significant (*F*(6, 246)= 1.0; *MSE*= 6675.78; *p*= .42; ε= .76). See Figure 3.

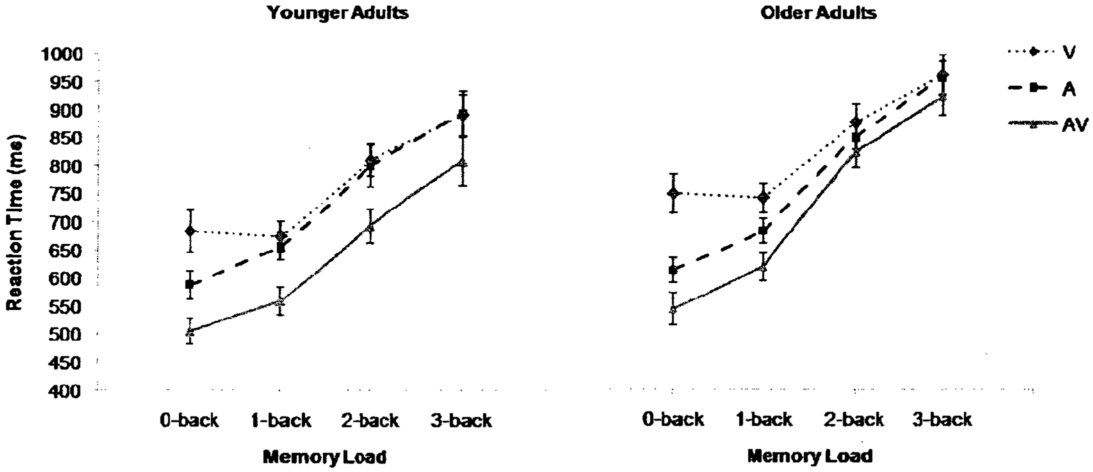


Figure 3: Mean reaction times and standard error bars for younger and older adults in visual (V), auditory (A) and auditory-visual (AV) conditions.

Electrophysiological Results

Due to time constraints, only 12 younger and 12 older adults were used for ERP analyses at this time. In order to prevent an order effect from interacting with data, two participants from each order of modality presentation were chosen. Participants with the least number of artefacts in their EEG recordings and the largest number of valid trials were selected for ERP analyses. Independent t-tests on demographic variables (age, years of education, hearing acuity, visual acuity, MoCA scores) showed that the participants chosen for the ERP analyses did not differ from the overall sample on any of these demographic characteristics. However, the chi-square analyses on the number of males and females revealed that there were more males in the whole sample than in the ERP sample ($\chi^2 = 4.8$; $p = .03$). All electrophysiological data were analyzed by repeated measures ANOVAs, and results are reported following the same rules as for behavioural data.

Analyses of behavioural data for ERP sample. In order to verify that the sample of participants chosen for ERP analyses did not differ from the overall sample behaviourally, analyses on the behavioural data from the ERP sample were performed. The analyses showed that in terms of accuracy, the same main effects and interactions were shown in the ERP and the total sample but the pairwise comparisons examining the main effect of Modality, interaction between Modality and Load, and interaction between Load and Age revealed a slightly different pattern of results. For the effect of Modality, participants in the ERP sample were more accurate in A-only and AV than in V-only condition, but they showed only a trend toward higher accuracy in AV compared to the A-only condition ($p = .07$). Also, examining the interaction between Modality and

Load showed that in 0-back, participants were less accurate in V-only than in AV and A-only but their performance in AV and A-only condition did not differ. In 1-back, participants performed the same in V-only compared to A-only or AV, but they performed better in A-only than in AV. In 2-back, participants were less accurate in V-only than in A-only and AV but they performed the same in AV and A-only. Lastly in 3-back, participants were less accurate in V-only than in A-only and AV, but they were also more accurate in AV compared to A-only condition. Exploring the significant interaction between Age and Load revealed that in 0-back and 1-back older adults did not differ from younger adults, but in 2-back and 3-back older adults were less accurate than younger adults. See Figure 4.

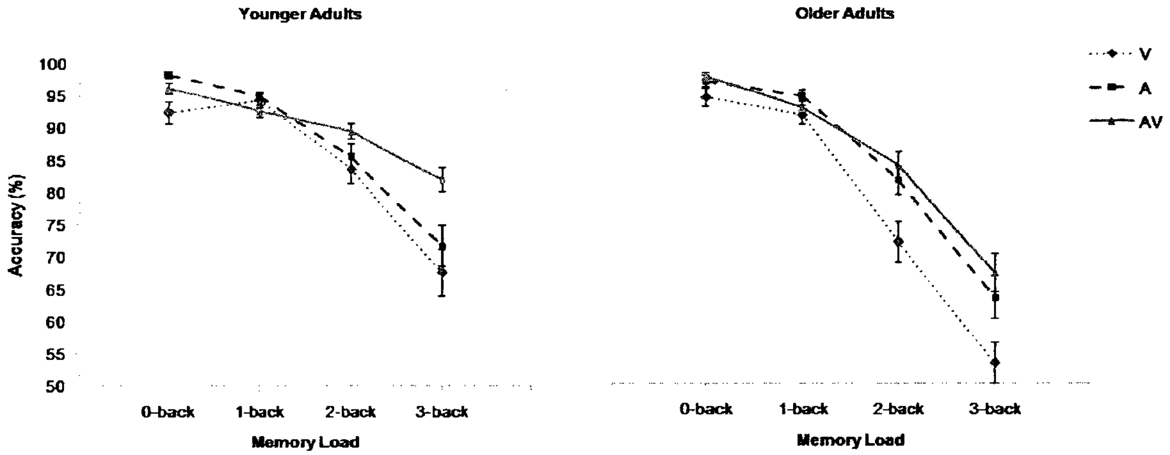


Figure 4: Mean percent of correct responses and standard error bars of younger and older adults included in the ERP analyses for visual (V), auditory (A) and auditory-visual (AV) conditions.

In terms of weighted n-back score, similar patterns of scores in terms of significant main effects of Modality and Age and a non-significant interaction between Modality and Age were identified but the pairwise comparisons examining the main effect of Modality revealed a slightly different pattern of scores. Specifically, participants reached a higher n-back score in AV compared to V-only and A-only conditions, but they also reached a higher score in the A-only compared to the V-only condition. See Figure 5.

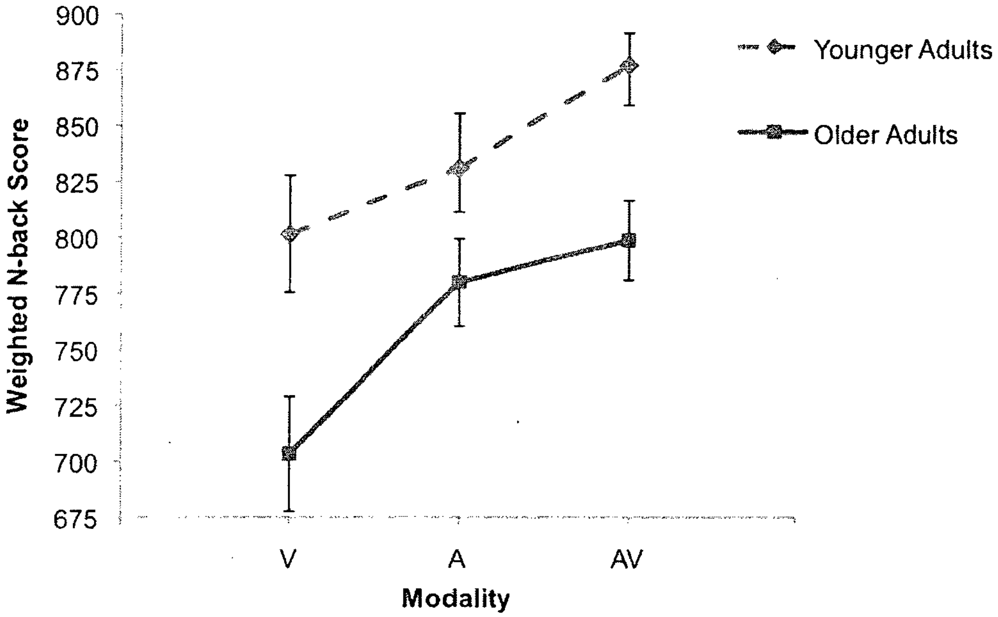


Figure 5: Mean weighted n-back scores and standard error bars for visual (V), auditory (A) and auditory-visual (AV) modality by younger and older adults included in ERP analyses.

In terms of reaction time, the ERP scores reflected a similar pattern to the whole sample for main effects and interactions. The only difference was a non significant Age effect ($F(1, 22) = 2.2$; $MSE = 96517.63$; $p = .15$). The pairwise comparisons also revealed a slightly different pattern of scores for the main effect of Load; the reaction time significantly increased as the memory load increased for all memory load levels except for 0-back compared to 1-back, where only a trend toward a significant difference was observed (p = .07). See Figure 6.

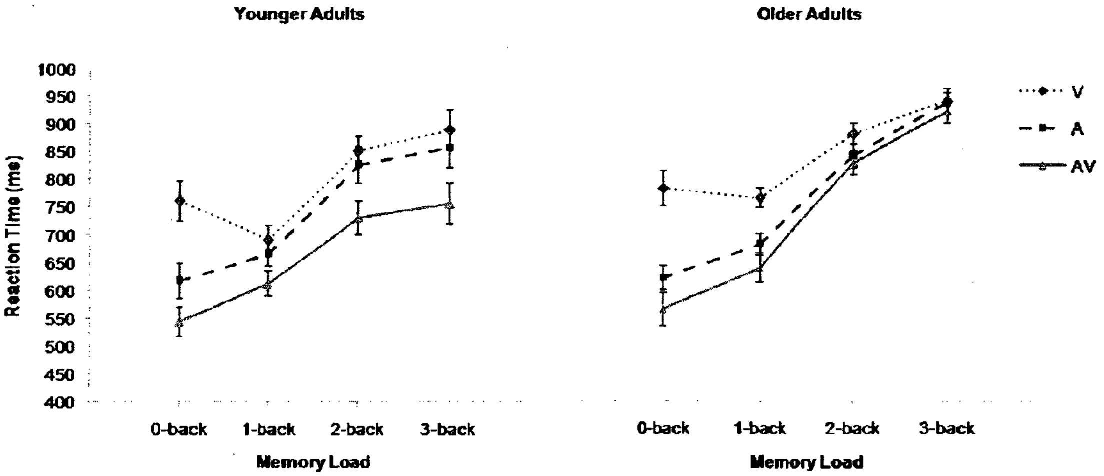


Figure 6: Mean reaction times and standard error bars of younger adults and older adults in the ERP analyses for visual (V), auditory (A) and auditory-visual (AV) conditions.

Amplitude and latency of auditory N1. In order to examine the effect of Modality, Load, and Age on the amplitude and latency of auditory N1, separate 3-way 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 or older) as a between subject factor. The V-only condition was not included in the analyses as only the ERPs triggered by the onset of the auditory speech signal were analyzed and compared across modalities. The sum of responses during unimodal conditions (A+V) was used to examine if the responses triggered during AV condition reflected a true multisensory interaction rather than an artefact created by presentation of two separate signals. More specifically, if the ERP signals during the AV condition would be the same as during the A+V condition, it could mean that ERPs triggered in the AV condition reflected two independent processes, and the difference between A-only and AV condition was due to the fact that the brain was processing two separate signals instead of one. In contrast, if the AV condition differed from the A+V, it could be assumed that providing visual cues affected auditory processing, and thus multisensory interaction had occurred.

The amplitude of auditory N1 was calculated as an absolute difference between the trough of P1 and peak of N1. The auditory P1-N1 complex is the most prominent in the mid-central electrodes and therefore the values from the Cz electrode were used for the analyses. Mean size of the amplitudes across different modality, load and age are shown in Table 2. Results revealed a significant main effect of Modality ($F(2, 44) = 12.0$; $MSE = 2.32$; $p < .001$; $\varepsilon = .83$). Specifically, the amplitude of N1 was smaller in AV than in A+V, and there was a trend for the N1 amplitude to be smaller in AV than in A-only condition ($p = .06$). Furthermore, the amplitude of N1 was smaller in A-only than in

A+V. See Figure 7. There was also a main effect of Age (*F*(1,22) = 9.0; *MSE* = 50.85; *p* = .01), showing that the amplitude of N1 was larger for older than for younger adults. This finding suggests that speech processing requires more processing capacity in older than younger adults, and it is consistent with the idea that speech processing is more effortful for older adults.

Next, there was a significant interaction between Modality and Age (*F*(2, 44) = 8.6; *MSE* = 2.32; *p* =.002; ε = . 83). For younger adults, the amplitude of N1 was smaller in A-only than in A+V but AV did not differ from A-only or A+V. In contrast, for older adults the amplitude of N1 was smaller in AV compared to A-only and A+V, but A-only and A+V did not differ. Lastly, there was a significant interaction between Load and Age (*F*(3, 66) = 3.4; *MSE* = 5.07; *p* = .03; ε = .86). The pairwise comparisons revealed that even though the amplitude of N1 was smaller in younger than in older adults across all memory loads, there were larger differences between younger and older adults in 2-back and 3-back than in 0-back and 1-back. Also, while for younger adults the amplitude of N1 did not seem to be affected by the memory load, for older adults it was smaller during less demanding compared to more demanding tasks. Specifically, the amplitude of N1 was smaller during 0-back compared to 2-back, and during 1-back compared to 2-back and 3-back. See Figure 8.

The latency of auditory N1 was calculated relative to the onset of the auditory trigger, which was signaling the onset of the auditory speech signal. Again, the data from Cz electrode were used for the analyses. Mean latencies of N1 across different modality, load and age are shown in Table 3. The results revealed a significant main effect of Modality (*F*(2, 44) = 3.6; *MSE* = 291.81; *p* = .05; ε = .74), with the auditory N1 peaking earlier during AV compared to A+V but not compared to A-only, and with no differences in the N1 latency between A-only and A+V conditions. See Figure 7. There was also a significant main effect of Age ($F(1, 22) = 8.1$; $MSE = 964.08$; $p = .01$), in that the auditory N1 occurred significantly earlier in younger than older adults.

Table 2

Mean amplitudes (μV) and standard deviations (in parenthesis) of N1 amplitude for younger and older adults at the Cz electrode

	Modality		
Load	A-only	A+V	AV
		Younger Adults	
0-back	3.76 (1.33)	5.00 (2.18)	5.01 (2.95)
1-back	3.69 (1.78)	4.60 (1.76)	4.76 (1.58)
2-back	3.81 (1.90)	4.13 (1.52)	3.51 (1.60)
3-back	4.33 (1.50)	5.09 (2.14)	3.76 (1.53)
		Older Adults	
0-back	6.65 (2.74)	6.92 (2.89)	5.44 (3.15)
1-back	6.48 (3.01)	6.16 (2.72)	5.81 (2.85)
2-back	7.89 (3.81)	8.65 (4.40)	5.78 (2.47)
3-back	7.59 (3.25)	7.85 (4.13)	6.50 (3.51)

Table 3

Mean latencies (ms) and standard deviations (in parenthesis) of N1 amplitude for younger and older adults at the Cz electrode

	Modality		
Load	A-only	A+V	AV
		Younger Adults	
0-back	97.70 (12.89)	98.27 (13.57)	96.63 (14.86)
1-back	93.40 (13.02)	102.19 (13.97)	94.48 (13.34)
2-back	95. 91 (15.44)	101.47 (19.03)	89.82 (14.50)
3-back	95.20 (13.07)	92.69 (15.90)	92.15 (13.33)
		Older Adults	
0-back	101.66 (13.20)	105. 24 (18.41)	101.50 (12.68)
1-back	112.40 (18.19)	112.57 (18.23)	108.99 (19.78)
2-back	104.59 (16.74)	110.61 (23.61)	100.36 (13.87)
3-back	110.45 (11.05)	106.38 (12.39)	100.36 (14.27)

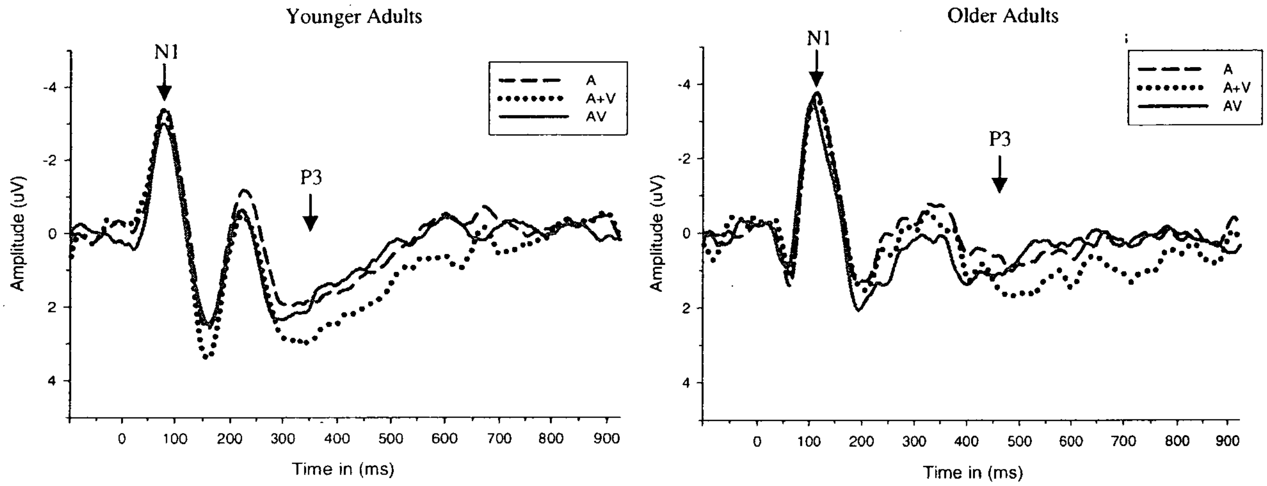


Figure 7: Grand average waveform at the Cz electrode during auditory (A), auditory + visual (A+V), and auditory-visual (AV) condition in younger (left graph) and older (right graph) adults. Note the reduction in N1 during AV condition compared to both A and A+V in older adults and earlier onset of N1 in AV compared to A+V condition in both groups.

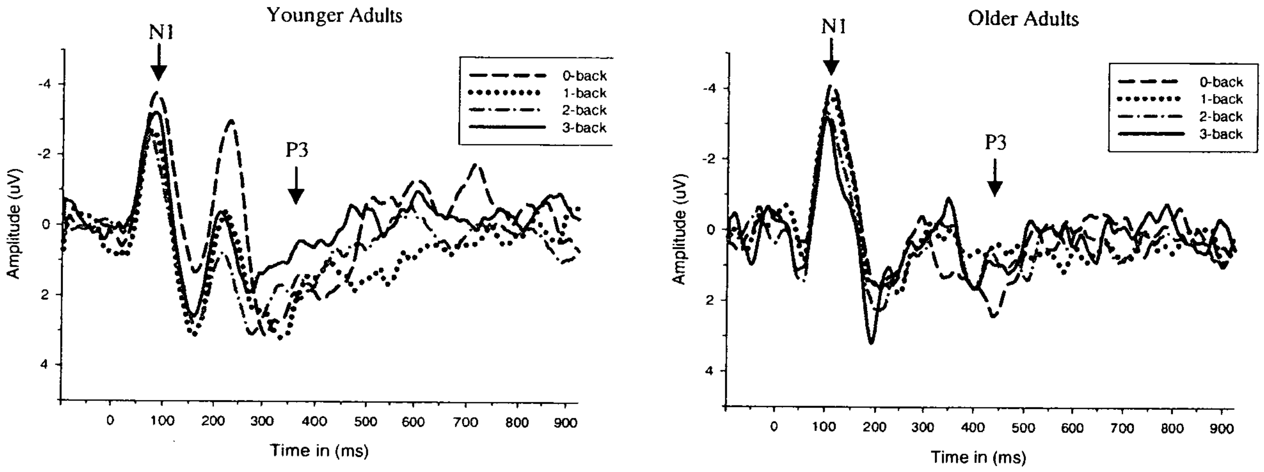


Figure 8: Grand average waveform at the Cz electrode during 0-back, 1-back, 2-back, and 3-back condition in younger (left graph) and older (right graph) adults. Note smaller N1 amplitude in younger adults across all memory loads, but especially pronounced differences in 2-back and 3-back conditions.

Amplitude and latency of P3. In order to examine the effect of Modality, Load, Electrode and Age on the amplitude and latency of P3, separate 4-way repeated measures ANOVAs were performed, with Modality (A-only, and AV), Load (0-,1-, 2-, 3-back), and Electrode (Fz, FCz, Cz, PCz, Pz) as within subject factors and Age (younger and older) as a between subject factor. The maximum peak of P3 is usually detected in mid-posterior electrodes (i.e., Pz); to assess expected distribution across the scalp, 5 midline electrodes (i.e., Fz, FCz, Cz, PCz, Pz) were chosen for the analyses. The modality effect compared the amplitude and latency of P3 during AV with P3 amplitude and latency during A-only. Since the ERPs were triggered to the onset of speech signal, neither V-only nor A+V were used in analyses since the components are shifted during V-only presentation, reflecting the onset of visual rather than auditory stimulus.

The amplitude of P3 was measured at the component's peak relative to the 0 μV baseline. Mean size of P3 amplitudes across different modality, load and age are shown in Table 4. The results revealed a significant main effect of Load ($F(3, 66) = 8.6$; $MSE = 10.82$; $p < .001$; $\varepsilon = .73$). Specifically, the amplitude of P3 was similar in 0-back compared to 1-back and 2-back compared to 3-back. But it was significantly larger in 0-back and 1-back compared to 2-back and 3-back (see Figure 9). There was also a significant main effect of Electrode ($F(4, 88) = 122.1$; MSE = 7.44; p < .001; $\varepsilon = .36$), showing that the amplitude of P3 significantly increased across the electrodes, moving from frontal to posterior areas, confirming the expected distribution across the scalp. Neither the main effect of Modality ($F(1,22) = .2$; $MSE = 8.67$; $p = .68$) nor Age ($F(1,22) = 1.6$; $MSE = 41.68$; $p = .21$) was significant.

There was, however, a significant interaction between Electrode and Age ($F(4, 88)$

= 4.1; *MSE* = 7.44; $p = .04$; $\varepsilon = .36$) showing that even though none of the comparisons between younger and older adults was significant, there were smaller differences between younger and older adults in Fz and FCz than in Cz, CPz, and Pz. Lastly, there was a significant interaction between Load and Electrode (*F*(12, 264) = 3.2; *MSE* = 2.55; $p = .02$; $\varepsilon = .34$) indicating that the effect of Load on the amplitude of P3 was similar across all the electrodes except for Fz, where there were no differences between any of the memory loads. This finding conforms to the expected posterior topography of the P3.

The latency of P3 was measured at the component's peak relative to the onset of the auditory trigger, which signaled the onset of the auditory speech signal. Mean latencies across different modality, load and age are shown in Table 5. The results showed a trend toward a significant main effect of Modality (*F*(1, 22) = 3.8; *MSE* = 51614.83; $p = .06$), indicating that the P3 occurred earlier in AV than in A-only (see Figure 10). Furthermore, there was a trend toward a main effect of Load (*F*(3, 66) = 2.7; *MSE* = 30921.31; $p = .07$; $\varepsilon = .83$), showing that the P3 occurred later in 1-back than in 0-back or 2-back. No other memory load conditions differed from each other (See Figure 9). The main effect of Age was not significant (*F*(1,22) = 3.3; *MSE* = 154735.15; $p = .08$).

Finally, there was a significant interaction between Load and Electrode (*F*(12, 264) = 2.8; *MSE* = 227.38; $p = .02$; $\varepsilon = .42$). There were no significant differences in the latency of P3 during 0-back, 1-back, and 2-back. In 3-back however, P3 occurred significantly later in Fz than in FCz, Cz, CPz and Pz, and significantly later in FCz than in CPz. No other means were significantly different from each other.

Table 4

Mean amplitudes (μV) and standard deviations (in parenthesis) of P3 at the Pz in younger and older adults

	Modality	
Load	A-only	AV
	Younger Adults	
0-back	5.46 (1.80)	6.16 (3.05)
1-back	5.56 (2.38)	5.81 (2.18)
2-back	4.55 (1.87)	4.48 (2.09)
3-back	4.76 (2.69)	4.25 (1.89)
	Older Adults	
0-back	5.26 (2.19)	4.64 (2.55)
1-back	4.40 (1.78)	4.07 (2.59)
2-back	3.52 (1.24)	3.37 (1.46)
3-back	3.72 (1.93)	4.07 (1.94)

Table 5

Mean latency (ms) and standard deviations (in parenthesis) of P3 at the Pz in younger and older adults

	Modality	
Load	A-only	AV
	Younger Adults	
0-back	474.65 (75.24)	436.30 (81.83)
1-back	475.73 (85.11)	480.75 (95.55)
2-back	457. 27 (79.84)	407.98 (90.28)
3-back	470.35 (85.20)	414.79 (86.73)
	Older Adults	
0-back	5.26 (2.19)	4.64 (2.55)
1-back	4.40 (1.78)	4.07 (2.59)
2-back	3.52 (1.24)	3.37 (1.46)
3-back	3.72 (1.93)	4.07 (1.94)

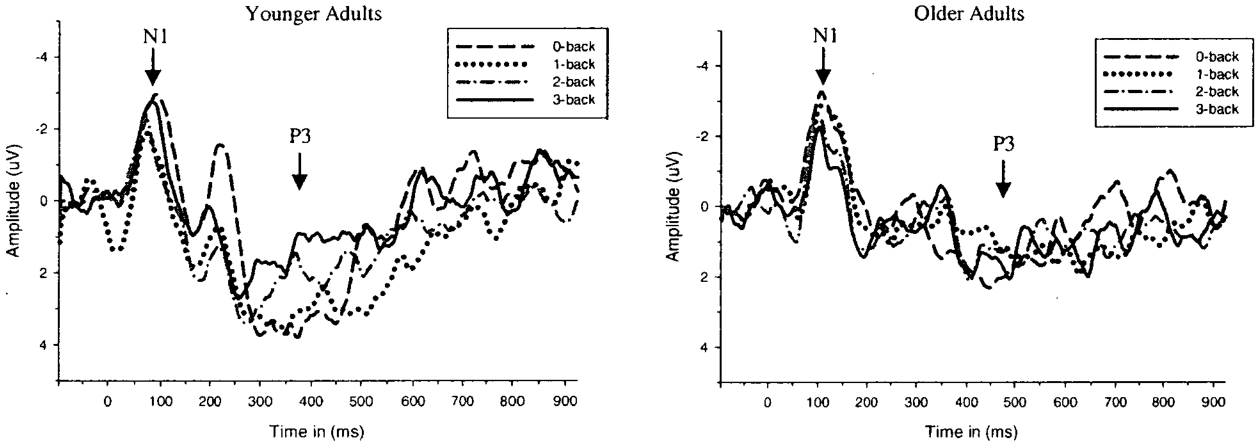


Figure 9: Grand average waveform at the Pz electrode during 0-back, 1-back, 2-back, and 3-back condition in younger and older adults. Note the reduction in P3 in the higher working memory load conditions (2-back and 3-back) compared to the lower working memory load conditions (0-back and 1-back).

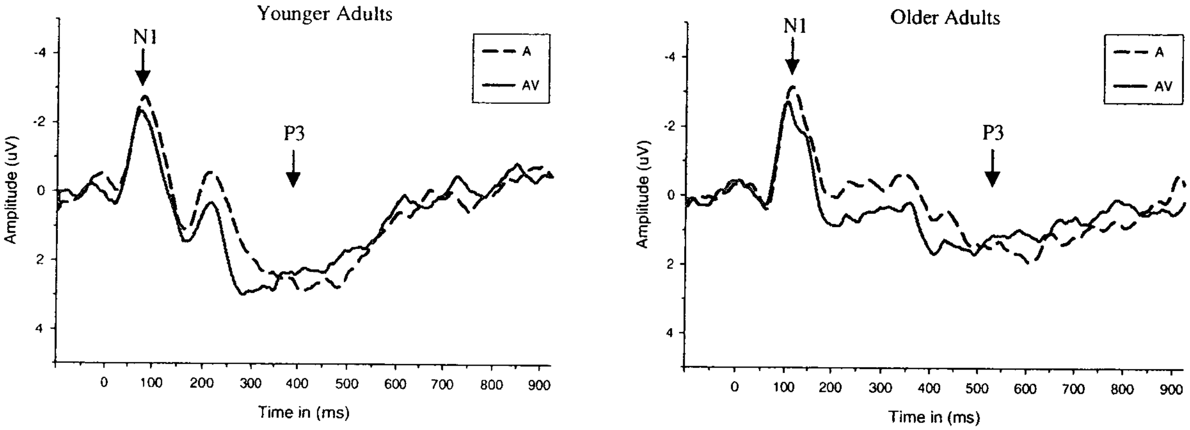


Figure 10: Grand average waveform at the Pz electrode site during auditory (A) and auditory-visual (AV) condition in younger and older adults. Note earlier P3 during AV condition compared to A condition in both age groups.

Correlations Between Electrophysiological and Behavioural Results

The relationship between sensory enhancement and improvements on the working memory task was examined by correlating a decrease in the amplitude and latency of N1 during AV compared to A-only and A+V conditions with the behavioural data as well as with the latency and amplitude of P3. The amplitude reduction in the AV condition was calculated as the difference in the amplitude of N1 during A-only or A+V conditions and the AV condition (i.e. A-only – AV; A+V - AV). Similarly, the latency reduction during AV condition was calculated as the difference in the latency of N1 during A-only or A+V conditions and AV condition. Since older adults but not younger adults showed an amplitude reduction during both AV and A+V conditions, the correlations were calculated separately for younger and older adults.

The results showed that there was no systematic relationship between the N1 amplitude reduction in the AV condition and any of the behavioural or P3 data (see Table 2). Interestingly, however, there was a significant positive correlation between accuracy and the decrease in N1 in AV compared to A-only and A+V during 2-back for older adults. The 2-back condition was the only condition in which older adults performed worse than younger adults in terms of accuracy. Thus, it seems that the facilitation of sensory processing may have helped older adults during this challenging task. The scatter plots showing significant correlations are presented in the Appendix E.

Similarly, there were no systematic correlations between latency reduction of N1 during AV and any of the behavioural or P3 data (see Table 3). OAs showed a significant correlation between latency reduction in N1 and latency of P3 during 3-back, suggesting that faster sensory processing may facilitate working memory processing. At this time, however, this conclusion should be considered with caution due to the lack of significant correlations during other n-back conditions and low accuracy in 3-back condition. The scatter plots showing significant correlations are presented in Appendix F.

Table 6

Zero-order correlations between the difference in auditory N1 amplitude during AV compared to A-only or A+V conditions and accuracy, reaction time (RT), weighted n-back score, P3 latency and P3 amplitude in younger (YA) and older (OA) adults

	Accuracy	RT	P3 amplitude	P3 latency
YAs: A-only - AV:				
0-back	.09	.15	-.04	-.32
1-back	-.01	.00	-.40	.24
2-back	.38	-.06	-.11	.04
3-back	.03	-.00	.60*	-.18
YAs: A+V – AV:				
0-back	.19	.34	-.29	-.15
1-back	-.60*	-.49	-.18	.14
2-back	.29	.38	-.20	.15
3-back	-.04	.25	.45	.08
OAs: A-only - AV:				
0-back	-.01	.00	-.40	.24
1-back	-.33	-.11	.17	-.52*
2-back	.56*	-.23	.49	.09
3-back	-.40	-.48	.15	-.29
OAs: A+V - AV:				
0-back	-.46	-.49	-.18	.14
1-back	-.05	-.26	-.16	-.38
2-back	.55*	-.35	.59*	.03
3-back	-.46	-.37	-.02	-.18

* value significant at $\alpha = .05$

Table 7

Zero-order correlations between the difference in auditory N1 latency during AV compared to A-only or A+V conditions and accuracy, reaction time, weighted n-back score, P3 latency and P3 amplitude in younger (YA) and older (OA) adults

	Accuracy	RT	P3 amplitude	P3 latency
YAs: A-only – AV:				
0-back	-.15	.32	-.38	.45
1-back	-.12	-.08	-.06	.04
2-back	.13	.17	-.13	.26
3-back	.11	-.28	-.35	.28
YAs: A+V – AV:				
0-back	.13	.36	-.46	.24
1-back	-.24	-.17	-.04	.26
2-back	.43	-.05	.31	.02
3-back	.25	-.29	-.08	.05
OAs: A-only - AV:				
0-back	.02	.11	-.42	.58*
1-back	.10	-.03	-.06	-.16
2-back	.29	-.22	-.39	.05
3-back	.06	.17	.21	.72**
OAs: A+V - AV:				
0-back	.08	-.12	-.47	-.23
1-back	.10	-.03	-.10	-.11
2-back	-.15	-.04	-.67**	.14
3-back	-.09	.18	.27	.58*

* value significant at α = .05
** value significant at α = .01

Discussion

This study has provided indirect evidence to support the suggestion that AV speech perception may facilitate working memory performance. Both younger and older adults performed better when they were asked to remember sequences of digits presented in AV compared to A-only or V-only modalities. A more direct relationship between the facilitation of sensory processing and improvement on a working memory task was not confirmed as no systematic relationship between decrease in N1 amplitude or latency during AV presentation and behavioural performance or P3 measures in AV condition was observed. Possible reasons that may have contributed to this unexpected finding are discussed further in the text. In the following sections I will discuss how the effect of n-back load, age, and modality affected behavioural performance and electrophysiological measures and how these results relate to previous findings.

Behavioural Results

As expected, a decline in accuracy and increase in reaction time was observed with each successive increase in working memory load, regardless of modality[5]. This was true for both younger and older adults and demonstrated the effectiveness of the task. More specifically, across successive memory loads, the n-back task requires that the participant hold more items in short term memory and keep track of their specific order; thus each successive n-back condition requires more working memory load, making the task longer and more difficult.

The predicted decrease in working memory performance with age was identified. Older adults performed more poorly than younger adults on all of the behavioural measures of working memory. These findings confirmed previously reported findings of age related declines in working memory (e.g., Bopp & Verhaeghen, 2009; McCabe et al., 2008; Salthouse 1994; Vaughan et al., 2008). In contrast, these findings do not agree with the finding of Pichora-Fuller and her colleagues (1995) who identified similar memory spans in younger and older participants when they were asked to read out loud sentences and identify the final words. Sentences were presented in sets of 2, 3, 4, 5, or 6, and at the end of each set, participants were asked to recall all final words. It is not clear whether participants needed to recall the final words in order of sentence presentation, which may contribute to the discrepancy between their findings and finding of the current study. This is because remembering the correct sequence of the digits is an essential part of the n-back task, except in 0-back load.

As predicted, there was an interaction between age and memory load for accuracy. For the memory conditions that required only attention (i.e., 0-back) or relatively low levels of working memory capacity (i.e.,1-back), older adults performed similarly to younger adults, whereas for more demanding tasks (i.e., 2-back) older adults were less accurate than younger adults. Interestingly, at 3-back, older adults did not differ from younger adults. These results suggest that, at least for relatively simple tasks, older adults can be as accurate as younger adults, even though they are slower due to their decline in speed of processing (Diamond et al., 2000). In the 2-back condition however, older adults showed a steep decline in accuracy and this decline was further exacerbated in 3-back condition. In comparison, younger adults showed a steep decline in accuracy during 3-back condition. This may suggest that as a group, the older adults reached their working memory capacity at a lower load (i.e., at 2-back) than younger adults (i.e., at 3-back).

Importantly, the lack of an interaction between modality condition and age on any of the behavioural measures suggests that both age groups benefited from AV presentation to a similar extent. As previously mentioned, older adults in the current study had significantly lower contrast sensitivity thresholds and significantly higher pure tone hearing thresholds compared to younger adults. This finding, together with similar findings observed in other studies (e.g., Cienkowski & Carney, 2002; Laurienti et al., 2006; Sommers et al., 2005; Winneke & Phillips, submitted) suggests that despite the age-related sensory declines, older adults benefit from AV speech presentation. The inverse-effectiveness hypothesis, which proposes that older adults should benefit from AV presentation to a greater extent than younger adults due to their decline in unimodal sensory processing, has not been supported in the current study. It is important to acknowledge though, that the prediction of no difference in AV benefit between younger and older adults presents a logical problem in hypothesis testing, as a null effect can hardly be proven with statistical tests of limited power. It is possible that subtle differences between age groups exist, but that they were not detected in the current study due to lack of power of the statistical tests.

According to prediction, there was improvement on all the behavioural measures of working memory in AV condition compared to A-only and V-only conditions. In terms of accuracy and weighted n-back scores, which were derived from accuracy scores, the current study demonstrated that both younger and older adults performed similarly during V-only compared A-only conditions, and both age groups benefited from AV compared to unimodal presentation. For accuracy scores, the enhancement effect during AV presentation relative to A-only or V-only presentation was evident in conditions that required more working memory processing load (i.e., 2-back and 3-back). This finding suggests that even though unimodal conditions may be sufficient when cognitive processing is relatively simple (i.e., 0-back and 1-back memory load), as the complexity of the task increases and more resources are required to complete the task accurately, AV speech presentation becomes especially useful.

In terms of reaction time, both younger and older adults benefited from the AV presentation at all levels of working memory load, suggesting that the AV modality facilitates processing speed independently of the cognitive demands of the task. In contrast, when comparing unimodal conditions participants were faster in A-only compared to V-only in less demanding conditions (i.e., 0-back and 1-back) but not in more demanding conditions (2-back and 3-back). Thus, even though there were processing speed benefits in the A-only condition compared to the V-only condition when demands on working memory were low, as the demands increased, the performance in the unimodal conditions became equivalent. It is important to remember that even though lip-reading is a challenging task, the stimuli that were used in the current study presented a very limited range of choices and participants were trained to lip-read the stimuli correctly before conducting the experiment. These factors may have contributed to the smaller discrepancy between A-only and V-only conditions.

Electrophysiological Results

There was a significant decrease in the amplitude of auditory N1 during AV condition compared to A+V condition and a trend toward a significant decrease in AV compared to A-only, suggesting facilitation of sensory processing during AV presentation. Upon further examination, it was revealed that this effect was mostly driven by the older adult group. The reason for the lack of effect in younger adults (despite the facilitation of their behavioural performance) may be that the auditory signal was very clear and there was nothing to interfere with auditory processing. In other words, the AV modality did not have a beneficial effect on the N1 amplitude in younger adults because the auditory processing during A-only condition already reached the maximum level of efficiency. It has been demonstrated that AV speech can be especially useful when speech comprehension is challenged by distraction from background noise (e.g., Callan et al., 2003; Schwartz, et al., 2004; Sommers et al., 2005; Sumby & Pollack, 1954; Summerfield, 1979) or auditory impairment (e.g., Bergeson & Pisoni, 2004; Grant et al., 1998; Rouger et al., 2007). Thus, impairment caused by the age related decline in hearing may have been enough to observe an AV speech benefit in older adults but due to high hearing acuity, younger adults may require further manipulation to make speech comprehension more challenging. One possibility would be to present the digits in background noise, which would likely enhance AV speech benefit on sensory processing in both age groups.

The auditory N1 occurred earlier during AV presentation compared to A+V presentation. Thus, the study replicated previous findings of faster sensory processing during AV presentation compared to A+V condition found in previous research (e.g., van Wassenhove et al., 2005; Winneke & Phillips, submitted), suggesting the speeding of sensory processing during multisensory interaction. The observation of faster sensory processing during AV condition compared to A-only condition (e.g., Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005; Winneke & Phillips, submitted) was not replicated in this study. This again may be related to the methodological design of the current study. Even though facilitation of sensory processing speed had been previously observed even without background noise (van Wassenhove et al., 2005), making the task more challenging could help magnify AV speech enhancement.

One challenge of multisensory research using ERP methodology is that the signal may change simply due to the presentation of stimuli through two or more sensory modalities simultaneously. In the current study, however, 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), suggesting that the AV condition represents a genuine multisensory interaction rather that an artefact resulting from processing two separate sensory information. That is, if the ERP signals during the AV condition would be the same as during the A+V condition, it could mean that ERPs triggered in the AV condition reflected two independent processes, and the difference between A-only and AV condition was due to the fact that the brain was processing two separate signals instead of one. In contrast, if the AV condition differed from the A+V, it could be assumed that providing visual cues affected auditory processing, and thus multisensory interaction had occurred. The effect of multisensory integrations on the amplitude and latency of auditory N1 observed in this study supports findings of multisensory integration observed in other studies, such as those by van Wassenhove and colleagues (2005) as well as previous studies conducted in our laboratory (Winneke & Phillips, submitted).

Overall, older adults showed a larger auditory N1 than younger adults, suggesting that they may be using more resources for the processing of sensory stimuli. The age difference was evident across all memory load conditions but it was especially pronounced in more challenging tasks (i.e., 2-back and 3-back). In addition, for younger adults the amplitude of N1 did not seem to be affected by the memory load but for older adults it was smaller during less demanding compared to more demanding tasks (0-back compared to 2-back, and during 1-back compared to 2-back and 3-back). Thus, it seems that the task difficulty not only affects higher order cognitive functioning, but it can also influence lower level sensory processing in older adults. Previous findings have shown an attenuation in N1 amplitude when the discrimination of target from non-target stimuli becomes more difficult (Fitzgerald & Picton, 1984; Wang et al., 2010). This seems contradictory to the findings of the current study, but the contradiction may be caused by task differences. Specifically, previous research has varied the level of difficulty at the sensory level, whereas the current study varied the level of difficulty at the higher order processing level, making the comparison of the results questionable. To our knowledge there is no study that varied the task difficulty in terms of memory load, and examined how this manipulation affects the auditory N1 amplitude.

The P3 showed the predicted decrease in the size of amplitude as the task became more challenging, replicating the results of previous studies (e.g., Segalowitz et al. 2001; Watter et al., 2001). More specifically, the amplitude of P3 during 0-back and 1-back was larger than during 2-back and 3-back. No difference between 0-back and 1-back was observed, suggesting that both tasks used approximately the same amount of working memory capacity. Similarly, no differences between 2-back and 3-back were observed, but in this case it may be that 3-back exceeded working memory capacity for both age groups and therefore further reduction in P3 did not occur. This is supported by the accuracy data, which showed that the mean accuracy in the 3-back condition was relatively low (68% for younger adults and 62% for older adults). Even though this performance is still above the chance level, it represents a substantial decrease compared to lower memory load conditions. The suggestion that 3-back condition may have exceeded working memory capacity for most of the participants is also supported by participants' reports, stating that the 3-back is very challenging and during the responding they relied more on whether or not they recently heard the digit rather than remembering the exact sequences.

Before discussing the effect of memory load on the latency and amplitude of P3, I will first discuss how the results of this study replicated classic P3 effects, including changes in the topography, age-related changes, and changes associated with increasing memory load. Replication of the classic P3 effects is important as it provides support for validity of observed P3 measures. The expected distribution of P3 across the scalp has been observed, with the P3 amplitude reaching its maximum at the posterior-central electrode (i.e., Pz) and getting progressively smaller as it moved from posterior to frontal electrodes. In terms of the age effect, no age differences were observed in the latency or amplitude of P3. Several studies have reported two foci, posterior and frontal, in the scalp distribution of P3 in older adults compared to predominantly posterior distribution in younger adults (e.g., Fabiani & Friedman, 1995; Friedman, Simpson, & Hamberger, 1994). These topographical differences between younger and older adults were not observed in the current study. This may be due to differences in task requirements. It has been suggested that larger topographical age differences in P3 are seen for Go/No-Go type tasks than for choice reaction tasks (Friedman et al., 1997), which is the type of task used in the current study.

The lack of age differences in the amplitude of P3 is an interesting observation considering lower accuracy of older adults compared to younger adults during 2-back and 3-back condition. However, it is important to remember that due to limitations of ERP methodology, which measures electrical activity from a large number of neural circuits and the precise location of these circuits is not determined, it should not be assumed that the lack of age differences in the latency and amplitude of P3 suggests similar cognitive processing in younger and older participants. This is specifically relevant for P3, which reflects activation of different neural structures across the brain (Johnson, 1993; Picton, 1992), and therefore the same end results may be observed despite differences in underlying processes. Also, the amplitude of P3 for older adults at the Pz electrode was smaller by 1 μV compared to younger adults (see Figure 10) and thus the lack of significant difference may be due to low power of the test when using a small sample size.

On the other hand, the lack of age differences in P3 latency is consistent with the lack of age differences in reaction time of the ERP sample. Older adults in the whole-study sample were slower than younger adults, however, and thus it is possible that age differences in P3 amplitude may be observed with a larger sample. In fact, the effect of age on the latency of P3 was close to being significant (p= .08), further supporting the idea that increasing the power of the test by having a larger sample may show a delayed P3 in older compared to younger adults. Importantly, however, the P3 seems to be affected in a similar manner by the task requirements in both younger and older adults (Friedman et al., 1997), which has been confirmed in the current study through the observation that increasing working memory load affected the latency and amplitude of

P3 similarly in both age groups.

No modality effect was observed for the P3 amplitude. There was a trend, however, for the P3 to occur earlier in AV compared to A-only condition, suggesting facilitation of cognitive processing during AV presentation. This is in agreement with the observed facilitation of reaction time during AV condition.

The trend toward an effect of load on the latency of P3 shown in the current study contradicts the previous findings of Watter and colleagues (2001), which showed decreased amplitude but constant latency of P3 across successive loads during the n-back task. This finding, however, needs to be interpreted with caution not only because it did not reach statistical significance and contradicted previous findings, but also because increasing load did not have a systematic effect on the latency of P3. In fact, the comparisons of means showed that only 1-back differed from 0-back and 2-back, while no other memory load conditions differed from each other. At this moment, the ERP data from only a half of the participants have been analyzed. Analyses derived from the entire sample will provide more information about reliability of this finding[4].

The significant interaction between memory load and electrode position observed in the current study is more difficult to interpret, as previous research examining the effect of increasing memory load on P3 during the n-back task concentrated more on amplitude than latency data (Segalowitz et al., 2001; Watter et al., 2001). Latency differences were observed only during 3-back condition, with P3 having longer latencies in frontal than parietal electrodes. Since P3 reflects the activation of distributed neural networks, it may be speculated that different cognitive processes are involved in the 3-back condition due to the higher level of difficulty and the use of different strategies during this task compared to less demanding ones. As a result, it may be speculated that frontal-posterior differences in P3 latencies reflect a longer engagement of frontal regions in highly demanding tasks. This speculation has to be considered with caution, however, as it is difficult to assign underlying neural sources to scalp-recorded ERPs.

It was hypothesized that the amplitude and latency of N1 during AV condition would be reduced compared to A-only and A+V condition, and that this reduction would be correlated with lower reaction time, higher accuracy, and higher weighted n-back score. Further, it was also predicted that a larger decrease in the amplitude of N1 when AV condition is compared to A-only and A+V condition, would be correlated with the larger amplitude and smaller latency of P3 in the AV condition. Thus, it was assumed that facilitation processes at the sensory level (as measured by the smaller and earlier auditory N1) would predict the better capacity of working memory (as measured by the larger amplitude and earlier latency of P3, and improved behavioural measures). In spite of clearly demonstrated behavioural benefits on working memory performance during AV presentation, this study did not identify a direct relationship between facilitation of sensory processing and enhancement of working memory. That is, no systematic correlations between a decrease in the N1 amplitude or latency during AV condition compared to A-only or A+V condition and behavioural data or the P3 data were identified. This may be partly due to the fact that no reduction in the amplitude of N1 during AV condition compared to A-only or A+V condition was observed in younger adults, and even in older adults the effects could likely have been enhanced by presenting the speech stimuli in background noise.

Limitations

The current study had several limitations that need to be considered and addressed in future research. First is a small sample size for the ERP data. At this moment, the ERP data from only a half of sample (12 people in each group) have been analyzed. This decreases the power of statistical tests and makes it more difficult to find significant correlations between ERP and behavioural data. In addition, the small sample size limits reliability of the data and decreases confidence with which the results can be generalized to the general population. Because of a small sample size and the exploratory nature of the current study, Bonferroni corrections were not applied in the analyses, increasing the possibility of Type I error in the reported findings. The findings, however, were predicted and consistent with the findings in the literature making the possibility of Type I error unlikely. Next, older adults who participated in the current study were well educated and high functioning individuals, and only those with normal (age appropriate) vision, hearing, and cognitive functioning were included in the study. This limits the generalizability of the findings to a broader population of older adults, especially those who may experience larger sensory decline or cognitive impairment. Finally, the study did not present stimuli under background noise, which possibly affected the absence of auditory N1 amplitude reduction during AV condition in younger adults. However, the behavioural improvements shown by younger participants provide clear evidence that AV presentation improved their working memory performance.

Implications and Future Directions

This study is the first that has examined how AV speech presentation influences the performance on working memory, using a well-established working memory task and concurrently recording ERPs. The study demonstrated that AV speech can help younger and older adults enhance their working memory performance. These findings have important theoretical and practical implications for learning as well as potentially helping to improve the quality of life not only of healthy younger and older adults but also older adults with cognitive impairment, such as those suffering from dementia.

The working memory benefits during AV speech demonstrated in the current study support previous findings that showed learning enhancement with AV presentation. For example, Jeung and Chandler (1997) showed that AV instructions (i.e., diagrams paired with voiced statements) led to superior learning of measurement geometry in primary school students from Australia, compared to equivalent visual-visual instructions (i.e., diagrams paired with written statements). The effect was, however, evident only for material that did not require a high level of visual search. The authors interpreted the results as suggesting that working memory capacity is increased when instructions are presented in the AV mode. However, when the material requires a great deal of visual search, working memory capacity becomes burdened by this task and the AV benefit disappears.

The AV presentation has also been found to promote the learning of a second language. Chung (2008) tested the learning of Chinese characters in English-speaking 7th Grade students who were learning Chinese as their second language. The characters were presented either in AV or V-only mode. In the AV mode, the characters were presented on a computer screen while the pronunciation and English translation of the word were provided aurally through earphones. In the V-only mode, the characters, their pinyin, and English translation were all presented on the computer screen. The results showed that students performed better on recall of word meanings when characters were presented in AV compared to V-only mode. In addition, 2 years later the same learners performed better on the recall of word meanings as well as pronunciation of words if the words were presented in AV compared to V-only mode. Thus, AV presentation seems to be a beneficial tool for learning a second language.

It is important to acknowledge that tasks used in the above presented studies differ in several dimensions from the working memory task used in this study. These learning tasks may be considered "higher order" memory tasks, possibly influenced by the strategy and dual encoding. Auditory and visual information presented in these studies were fundamentally different from each other and thus the improvement in AV condition may be more due to better encoding when both visual and auditory information is available. In contrast, working memory improvements in the current study seem to be more due to true sensory interaction, in that visual speech information influences auditory speech processing.

This study has shown that as the load on working memory increases, the benefits that older and younger adults derive from AV speech become even more prominent. Thus, the presentation of information in the AV modality could be especially beneficial when the information requires higher demands on working memory, for example when complex instructions are being explained. Face to face interactions and using video could be two methods how to promote learning by AV speech presentation.

Given the benefits that healthy older adults in the current study derived from AV presentation, it is likely that older adults with cognitive deficits, such as those with mild cognitive impairment (MCI) or Alzheimer's disease (AD), could benefit from AV speech to an even higher extent. AD usually starts with declines in episodic memory and then gradually progresses into global cognitive deficits (Collie & Maruff, 2000). Similarly, MCI, sometimes called the pre-dementia stage, is characterized by the impairment on one or more cognitive domains, but without meeting the criteria for AD (Petersen, 2004). Many people with MCI, 6-25% per year (Petersen et al., 2001), eventually progress to AD. Higher order functions, such as working memory (e.g., Baddeley, Baddeley, Bucks, & Wilcock, 2001; Lopez et. al., 2006) are impaired in both MCI and AD. Given their deficits in working memory, AV enhancement should have a pronounced effect in these patient groups.

The current study provided useful information about the working memory improvements derived from AV speech presentation. Two future directions are suggested from the present results. The first is to repeat the current study but with the inclusion of background noise. As previously mentioned, younger adults did not show sensory enhancement during AV speech presentation despite clearly shown behavioural improvements. Background noise would make speech comprehension more challenging, which could in turn magnify the AV effect on early sensory processing. The second possibility for direction of future research is to assess AV speech benefit in older adults with MCI or AD. This research would provide useful information with regard to the magnitude of the AV enhancement effect in older adults with cognitive deficits. In addition, the results could help determine compensatory strategies (e.g., facing individuals when talking to them or watching news on television rather then listening to the radio) that would help diminish their functional impairments and improve their quality of life.

References

Abel, S. M., Sass-Kortsak, A., & Naugler, J. J. (2000). The role of high frequency hearing in age-related speech understanding deficits. *Scandinavian Audiology, 29*, 131-138.

Baddeley, A. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A. D., Baddeley, H. A., Bucks, R. S., & Wilcock, G. K. (2001). Attentional control in Alzheimer's disease. *Brain, 124*, 1492-1508.

Baudouin, A., Clarys, D., Vanneste, S., & Isingrini, M. (2009). Executive functioning and processing speed in age-related differences in memory: Contribution of coding task. *Brain and Cognition, 71*, 240-245.

Bergenson, T. R., & Pisoni, D. B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes*. Cambridge, Mass: The MIT 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. *Journal of Neuroscience, 28*, 14301-14310.

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

Boersma & Weenink (2010). Praat: doing phonetics by computer (Version 5.1.30) [computer software]. Amsterdam, Netherlands.

Bopp, K. L., & Verhaeghen, P. (2009). Working memory and aging: Separating the effects of content and context. *Psychology and Aging, 24*, 968-980.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processing underlying perceptual enhancement by visual speech gestures. *Neuroreport, 14,* 2213-2218.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport, 10*, 2619-2623.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10,* 649-657.

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.

CHABA - Committee on Hearing and Bioacoustics. (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83*, 859-895.

Chung, K. K. H. (2008). What effect do mixed sensory mode instructional formats have on both novice and experienced learners of Chinese characters? *Learning and Instructions, 18*, 96-108.

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

Clavagnier S., Falchier, A, & Kennedy H. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 117-126.

Collie, A., & Maruff, P. (2000). The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neuroscience & Biobehavioral Reviews, 24*, 365-374.

Compumedics Neuroscan (2003). Scan (Version 4.3.1) [Computer software]. El Pasco, Texas.

Davis, C., & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition, 100,* B21-BB31.

Diamond, B. J., DeLuca, J., Rosenthal, D., Vlad, R., Davis, K., Lucas, et al. (2000). Information Processing in older versus younger adults: Accuracy versus speed. *International Journal of Rehabilitation and Health, 5,* 55-64.

Easton, R. D., & Basala, M. (1982). Perceptual dominance during lipreading. *Perception & Psychophysics, 32*, 562-570.

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: The novelty oddball. *Psychophysiology, 32*, 579-594.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*, 5749-5759.

Fitzgerald, P. G., & Picton T. W. (1984). The effects of probability and discriminability on the evoked potentials to unpredictable stimuli. *Annals of the New York Academy of Sciences, 425*, 199-203.

Friedman, D., Kazmerski, V., & Fabiani, M. (1997). An overview of age-related changes in the scalp distribution of P3b. *Electroencephalography and Clinical*

Neurophysiology, 104, 498-513.

Friedman, D., Simpson, G., & Hamberger, M. (1994). Age related changes in scalp topography to novel and target stimuli. *Psychophysiology, 30,* 383-396.

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.

Goolkasian, P., & Foos, P. W. (2005). Bimodal Format Effects in Working Memory. *The American Journal of Psychology, 118,* 61-78.

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.

Handy, T. C. (2005). *Event-related potentials: A method handbook.* Cambridge: A Bradford Book.

Haymes, S. A., Roberts, K. F., Cruess, A. F., Nicolela, M. T., LeBlanc, R. P., Ramsey, M. S., et al. (2006). The letter contrast sensitivity test: Clinical evaluation of a new design. *Investigative Ophthalmology & Visual Science, 47,* 2739-2745.

Hitch, J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology, 28,* 603-621.

Hockey, A., & Geffen G. (2004). The concurrent validity and test-retest reliability of a visuospatial working memory task. *Intelligence, 32,* 591-605.

Jeung, H. -J., & Chandler, P. (1997). The role of visual indicators in dues sensory mode instruction. *Educational Psychology, 17,* 329-343.

Johnson, R. Jr. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology, 30*, 90-97.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of compression: Individual differences in working memory. *Psychological Review, 99*, 122-149.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 615-622.

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

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*, 557-577.

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.

Lopez, O. L., Becker, J. T., Jagust, W. J., Fitzpatrick, A., Carlson, M. C., DeKosky, S. T., et al. (2006). Neuropsychological characteristics of mild cognitive impairment subgroups. *Journal of Neurology, Neurosurgery, and Psychiatry, 77*, 159-165.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass.: MIT Press.

Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., & Olivetti Belardinelli, M.

(2008). How the bimodal format of presentation affects working memory: An overview. *Cognitive Processing, 9*, 69-76.

McCabe, J., & Hartman, M. (2008). Working memory for item and temporal information in younger and older adults. *Aging, Neuropsychology, and Cognition, 15*, 574-600.

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

Millisecond Software (2008). Inquisit (Version 2.0) [Computer software]. Seattle, Washington.

Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., & D'Esposito M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging, 15*, 527-541.

Möttönen, R., Schürmann, M., & Sams M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: A magnetoencephalographic study. *Neuroscience Letters, 363*, 112-115.

Munhall, K. G., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by Eye II*. East Sussex, UK: Psychology Press.

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.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatric Society, 53,* 695-699.

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.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256,* 183-194.

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology, 56,* 1133-1142.

Phillips, N. A., Gagné, J-P., Basu M., Copeland, L., Gosselin, P., Winneke, A., et al. (2009). *The role of sensory and cognitive factors in audio-visual speech perception in younger and older adults.* Talk presented at the 2010 Cognitive Aging Conference, April 15-18th, Atlanta, GA.

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: Springer-Verlag.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42,* S59-S67.

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

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.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*, 456-479.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language and Hearing Research, 52*, 1073-1081.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*, 2128-2148.

Pisoni, D. B., Saldaña, H. M., & Sheffert, S. M. (1996). *Multi-modal encoding of speech in memory: A first report.* Unpublished manuscript, Indiana University, Bloomington.

Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive matrices.* Oxford, England: Oxford Psychologists Press.

Reale, R. A., Calvert, G. A., Thesen, T., Jenison, R. L., Kawasaki, H., Oya, H., et al. (2007). Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience, 145*, 162-184.

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

Salthouse, T. A. (1994). The aging and working memory. *Neuropsychology, 8*, 535-543.

Santangelo, V., Mastroberardino, S., Botta, F., Marucci, F. S., & Olivetti Belardinelli, M. (2006). On the influence of audio-visual interactions on working memory performance: A study with non-semantic stimuli. *Cognitive Processing, 7*, 187.

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age related changes in temporal processing: Implications for speech perception. *Seminars in Hearing, 22*, 227-239.

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 Associates.

Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition, 93*, B69-B78.

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.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Audio-visual speech perception examined by fMRI and PET. *Neuroscience Research, 47*, 277-287.

Singhal, A., & Fowler, B. (2005). The effects of memory scanning on the late Nd and P300: An interference study. *Psychophysiology, 42*, 142-150.

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

Stein, B. E., & Meredith, M. A. (1993). *Merging of 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.

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

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36*, 314-331.

Summerfield, Q. (1987). In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

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

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology, 48*, 380-396.

Tun, P. A., & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. *The Journal of Gerontology: Psychological Sciences, 54B*, 317-327.

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging, 24*, 761-766.

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition, 96,* B13-B22.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron, 43*, 271-282.

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

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*, 1181-1186.

Vaughan, L., Basak, C., Hartman, M., & Verhaeghen (2008). Aging and working memory inside and outside the focus of attention: Dissociation of availability and accessibility. *Aging, Neuropsychology, and Cognition, 15*, 703-724.

Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Ageing, 18,* 332-339.

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

Wang, Y., Song Y., Qu, Z., & Ding, Y. (2010). Task difficulty modulates electrophysiological correlates of perceptual learning. *International Journal of Psychophysiology, 75*, 234-240.

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.

Wechsler, D. (1981). *Manual for the Wechsler Intelligence Scale for Children, 3rd edn.* New York: Psychological Corporation/Harcourt Brace and Company.

Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constrains of memory. *Journal of the American Academy of Audiology, 7,* 175-182.

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.

Winneke, A. H., & Phillips, N. A. (2009). Visual speech cues make older ears younger: An investigation of age-related differences in audiovisual speech perception using event related potentials. Manuscript submitted for publication. Concordia University.

Wright, T. M., Pelphery, K., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interaction along lateral regions evoked by audiovisual speech. *Cerebral Cortex 13*, 1034-1043.

Footnotes

[1] The reaction time measured from the onset of the auditory trigger resulted in an underestimation of true reaction time in V-only trials since the onset of visual speech information preceded the auditory trigger. To measure reaction time in this manner was important, however, in order to compare different modality conditions.

[2] Files for OAs were downsampled to 512 Hz using Decimator (Biosemi, Amsterdam, Netherlands) in order to decrease the size of the files. Files for YAs could not be decimated because their Inquisit files were missing a necessary component for re-setting the triggers to 0, which created problems when files were supposed to be downsampled. This problem was discovered only after most of YAs had been tested and thus a decision to continue testing YAs with the same set up was made. The YAs' continuous EEG files were downsampled to 512 Hz in Scan 4.2 (Neuroscan, El Paso, TX, USA) after being divided into epochs.

[3] The outliers were replaced by the group means because it was noticed that these low scores were due to factors unrelated to the task (e.g., a person falling a sleep for a few seconds, or not pressing the buttons on the mouse hard enough in order for computer to detect the response). Leaving the outliers would skew the data and underestimate the accuracy in these conditions, making the comparisons of the means unfair.

[4] Previous ERP study, assessing multisensory processing used the sample of 17 younger and 17 older adults (Winneke & Phillips, submitted).

[5] The only exception was 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 2-back and a subsequent significant increase in 3-back condition.

Appendix A

History Questionnaire

We are interested in your personal history because it may help us to better understand the results of our study. Your answers to a few short questions will aid us in this effort. All answers will be kept strictly confidential. Thank you for your help.

Demographics:

1. Date of Birth (D/M/Y):______________ 2. Age:___________

3. Gender: *(circle response)* (1) Male (2) Female

4. Handedness: *(circle response)* (1) LEFT (2) RIGHT (3) BOTH

5. Present marital status: *(circle response)* (1) Single – never married
(2) Married
(3) Separated
(4) Divorced
(5) Widowed
(6) Cohabit

Language

6. Place of Birth:
____

7. Languages Spoken (in order of fluency):
____

8. Primary Language/Language of choice:
____

9. Language at home:_______________________
10. At Work: _______________________

11. Language of Education:___________________________________

12. At what age did you first learn English/French? _____________________

13. At what age did you become fluent in it?_____________________

14. How would you rate, from 1 to 5 , your level of proficiency in the languages you speak? What percentage of time do you speak it?
Language Rating (**L**istening, **R**eading, **S**peaking, **W**riting):

1. ___________________ L:____ R:____ S:____ W:____ %:____
2. ___________________ L:____ R:____ S:____ W:____ %:____
3. ___________________ L:____ R:____ S:____ W:____ %:____

4. ____________________ L:____ R:____ S:____ W:____ %:____

15. .How many years of education do you have at this time? (i.e., what is the highest level achieved?)

1 2 3 4 5 6	7 8 9 10 11	12 13	14 15 16	17 18 19 20	21 22 23 24 25
Elementary	Secondary	Cegep	Undergrad	Graduate	Professional

16. In what field did you complete your degree? ______________________________

17. Did you skip or repeat a grade?
   A) NO / YES
   B) Which one (s):____

18. Did you have any particular difficulty with any subject in school?
   A) NO/YES
   B) Which one (s):

____

19. What is or was your main occupation?

____

20. What was your longest held occupation?

____

21. When did you retire? ________________

22. How many hours per week do you engage in physical exercise? _______________

23. How many hours per week do you engage in a social activity (this can include interacting with members of your household)?_____________

Medical History

24. Do you have now, or have you had in the past -*(please circle your response)*

Visual problems:	A) Nearsighted / Farsighted B) Glasses / Contact lenses C) Cataract: Left / Right D) Colour blind: NO / YES
Trouble hearing:	E) NO / YES F) Hearing Aid: Left / Right

25. Have you ever been unconscious , had a head injury or had blackouts ?
   A) NO / YES

B) Cause:____
C) Duration:______________________________________
D) Treatment:_____________________________________
E) Outcome:______________________________________

26. Have you been seriously ill or hospitalized in the past 6 months?
A) NO / YES
B) Cause:____
C) Duration:____

Do you have now, or have you had in the past (conditions susceptible or influencing cognitive functions)...

27. a) A stroke? b) ^{s} Transient ischemic attack(mini-stoke[1])?	NO / YES NO / YES	
28^{s}. Bypass surgery?	NO / YES	
29^{S} Heart disease?	NO / YES	Nature (myocardial infarction [MI], angina, narrowing of arteries):
30 ^{S} High blood pressure?	NO / YES	Is it controlled? NO / YES What medication? ____
31 ^{S}. High cholesterol?	NO / YES	Is it controlled? NO / YES What medication? ____
32 ^{S}. a) Diabetes? b) Insulin dependent?	NO / YES	Type 1 / Type 2 Age of onset: ____ Treatment: ____
33. Other Surgery?	NO / YES	
34. Seizures?	NO / YES	Age Onset: ____ Frequency: ____ Cause: ____ Treatment: ____
35. Epilepsy?	NO / YES	
36. Thyroid disease?	NO / YES	
37. Frequent headaches?	NO / YES	Tension / migraine
38. Dizziness?	NO / YES	
39. Trouble walking Unsteadiness?	NO / YES NO / YES	
40. Arthritis?	NO / YES	

41. Any injuries to the lower limb? (e.g. hip, knee, ankle)	NO / YES NO / YES	
42. Serious illness (e.g. liver disease)?	NO / YES	
43. Neurological disorders[2]? (e.g. lupus, MS, Parkinson's)	NO / YES	
44. Exposure to toxic chemicals (that you know of)?	NO / YES	
45. Depression?	NO / YES	Did you seek assistance or feel the need to do so? ________ Is it controlled? ________
46. Anxiety?	NO / YES	Did you seek assistance or feel the need to do so? ________ Is it controlled? ________
47. Other psychological difficulties?	NO / YES	
48. Hormone replacement?	NO / YES	
49. Steroids?	NO / YES	

50. Medication: Please list the medication you are currently taking and any other medication that you have taken in the past year.

Type of medication		Reason for consumption	Duration of consumption and dose
A			
B			
C			
D			
E			
F			

51. Do you drink alcohol? a) YES, frequently.
b) YES, but infrequently.

[S] Risk factors for stroke. Exclusion criterion: More than one of those factors, if older participants.

[1] Mini-stroke: symptoms less than 24 hours.

[S] Risk factors for stroke. Exclusion criterion: More than one of those factors, if older participants.

[2] Automatic exclusion

52. Do you use non-prescription drugs such as homeopathic medications, vitamins, laxatives, syrups ?

NO / YES
If YES, which one (s): ______________________________________
How many times per week?
a) Occasionally b) 1 - 3 c) 4 - 6 d) more than 6

53. Do you use non-prescription drugs for recreational purposes?

NO / YES
If yes, do you use marijuana/hashish?
NO / YES
If YES, How many times per week?
a) Occasionally b) 1 - 3 c) 4 - 6 d) more than 6

Do you use any other non-prescription drugs for recreational purposes?
NO / YES
If YES, How many times per week?
a) Occasionally b) 1 - 3 c) 4 - 6 d) more than 6
If yes, which one (s): (*participant not obliged to answer*)
____

Ask participant to not use drugs prior to testing (~48hr)

54. Do you smoke?

NO / YES
If YES, How many packs a day (or average quantity)? ____________

55. Current problems: Are you currently troubled by any of the following ?

a) Concentration / Attention problems?
NO / YES
Nature: ____
b) Memory problems?
NO / YES
Nature: ____
c) Difficulties finding words?
NO / YES
Nature: ____

56) How would you rate your health? *(circle response)*

1) poor 2) fair 3) good 4) very good 5) excellent

<u>Participant contact information</u>:

Name: ____

Phone Number: ___________________________

Email: ___________________________

Address ***(remind participant that this section is optional)***:

____

____

____

Are you willing to be contacted by researchers in Dr. Phillips' lab for future studies? **NO / YES**

What year will you graduate? ______________________

Can we give your contact information to other Concordia researchers (name, tel. #, email address)?

NO / YES

Source: ____

Eligibility:

- You are not eligible for this study due to ________ reasons, but you may be eligible for other studies, so we'll keep your information on file
- I need to discuss some issues with my colleagues, and I will contact you to let you know if you are eligible to participate.
- If they ask why they are ineligible:
    - We are interested in cognitive processing and certain conditions, medications, and habits interfere with cognitive processing, therefore we cannot test people who meet those criteria.

Appendix B

Consent Form
Auditory Visual Speech Perception and Working Memory

Purpose of the Study:

I have been informed that the purpose of this research is to investigate the impact of a sensory modality (i.e., visual presentation, auditory presentation, or both) on performance on working memory task. Findings from this study will increase our present understanding of the processes involved in multi-sensory integration in the human brain.

Details of the Study:

The study will take place in the Cognitive Psychophysiology laboratory of the Department of Psychology at Concordia University. While I am performing a memory task, my electroencephalogram (EEG) will be recorded. This is a recording of electrical brain activity measured at the scalp (similar to an EKG recording of heart activity). To record the EEG, a nylon cap will be placed on my head and little sensors (electrodes) will be attached to the cap.

The study will be conducted in a small testing room and will last about two hours. I will be seated in a comfortable chair and 1) will see a video clip of a woman speaking digits on a computer monitor, 2) hear the woman speaking digits through headphones or 3) watch her speak the digits and hear the digits at the same time. In each condition I will be asked to decide whether the currently presented digit matches the one presented before. I will indicate my responses by pressing buttons on the computer mouse. I understand that the task will get increasingly more difficult as I will need to remember more and more digits and that I may make errors but the most important thing is that I will try to do my best.

I have been informed that I will be screened for hearing and vision function and that certain demographic information (age, sex, education, and health status) will be recorded. I understand that this test is for research purposes only and that it is not diagnostic, meaning that it will not yield any results about my health. I understand that my individual results will not be provided to me; however, I will be informed of the general findings of the study.

Disadvantages and Risks of Participating in the Study:

EEG testing is a painless and non-invasive (using no foreign substances like medications, tubes, or needle injections) procedure. It is possible that the working memory task will lead to fatigue and frustration because I may not be able to accurately respond at all times. However, I am asked to do the best that I can and I will be given frequent breaks whenever required to avoid this. I understand that in the unlikely event that any potentially significant abnormality in my EEG is observed, I will be encouraged to contact my family physician for appropriate follow-up.

Advantages to Participating in the Study:

The researchers hope to learn more about the different brain processes that are involved in processing information of two modalities at the same time. Although this will not benefit me directly, this research could add to our scientific understanding of multisensory processing, in general, and auditory visual speech perception in particular. In addition, I will gain knowledge about how psychological research is conducted.

Confidentiality:

I understand that my participation in this study is *confidential*, that is, the researcher will know but will not disclose my identity in any published report or scientific communication. My records will not be identified by name; instead a subject code will be used. If the present study is published, only group results will be mentioned, ensuring my confidentiality as a participant in this experiment.

Withdrawal from the Study:

I understand that my participation in this study is voluntary and, if I agree to participate, I may withdraw my consent and discontinue participation *at any time* without negative consequences.

Participant's Rights:

I have fully discussed and understood the purpose and procedure of this study and have had the opportunity to ask any questions. The following is the name, address, and telephone number of the researcher whom I may contact for answers to questions about the research or any injuries or adverse reactions which might occur: **Dr. Natalie Phillips, Department of Psychology, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec, H4B 1R6; tel: 848-2424 ext. 2218.** For any ethical concerns regarding this study I may contact **Adela Reid at the University Office of Research (adela.reid@concordia.ca or 848-2424 ext. 7481) or Dr.Virginia Penhune from the Psychology Department Ethics Committee (virginia.penhune@concordia.ca).**

Signature:

I have understood the contents of this consent form and have had the opportunity to ask questions. I agree to participate in this study.

____
Date

________________________	________________________
Signature of Subject	Print Name
________________________	________________________
Signature of Investigator	Print Name
________________________	________________________
Signature of Person explaining	Print Name

Appendix C

Debriefing Sheet
Auditory Visual Speech Perception and Working Memory

We are interested in how individuals process the information presented in different sensory modalities (i.e., visual and auditory) as well as how they process information presented through two modalities (i.e., auditory-visual) at the same time. Previous research showed that when we can both hear and see a person speaking we comprehend language better and we need fewer cognitive resources to process information than when we can only hear or see the person speaking. These observations have been confirmed at both behavioral level (i.e., people are both faster and more accurate) as well as at electrophysiological level (i.e., brain responses are lower). What is not known yet is whether the cognitive resources "saved" during auditory-visual presentation of language can be used for other cognitive processes such as working memory. Working memory plays a substantial role in processing language and previous research observed a trade-off between working memory and language comprehension. This means that the harder it is for us to comprehend language (such as when there is a lot of noise in the background), the harder it is to remember presented information.

The present research looks at whether there is a direct relationship between "saved" resources during auditory-visual presentation of language and working memory performance. It is hypothesized that when people can both hear and see a person speaking, they will be able to remember more information than when they can only hear or see the person speaking. It is also predicted that during auditory-visual presentation of language, people's brain responses will be lower than during the auditory or visual presentation of language and that the lower the brain response during auditory-visual presentation, the better the performance on working memory.

The research will help us to better understand the brain processes behind multi-sensory integration as well as further improve our understanding of cognitive processes involved in speech.

For further information or questions regarding this study please contact the experimenter, Jana Baranyaiova Frtusova, or the faculty supervisor, Dr. Natalie A. Phillips, at 848-2424 ext.7546. For ethical concerns regarding this study please contact Adela Reid at the University Office of Research (adela.reid@concordia.ca or 848-2424 ext. 7481) or Dr. Virginia Penhune from the Psychology Department Ethics Committee (virginia.penhune@concordia.ca).

Suggested readings:

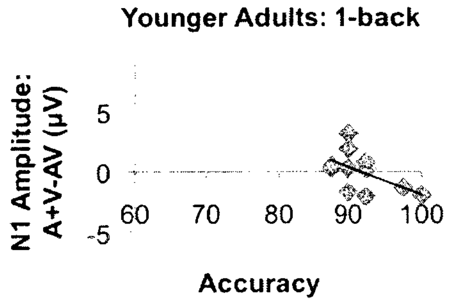
Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., & Olivetti Belardinelli, M. (2008). How the bimodal format of presentation affects working memory: An overview. *Cognitive Processing, 9*, 69-76.

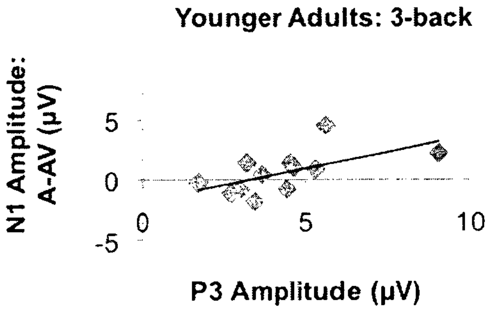
van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Science, 102*, 1181-1186.

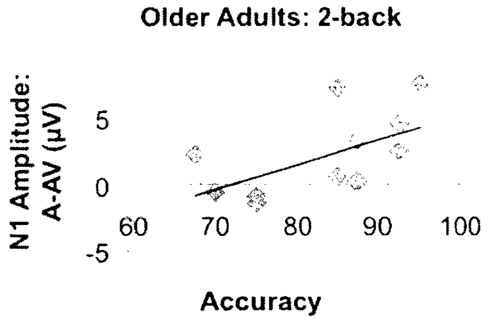
Appendix D

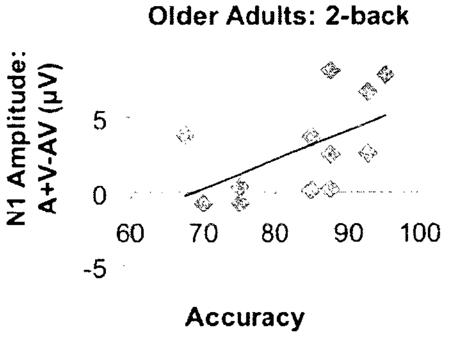
Digit	Lag between the visual and auditory trigger
1	268
2	434
3	340
4	287
5	529
6	283
8	422
9	516
10	479
Mean	395.33
SD	103.24

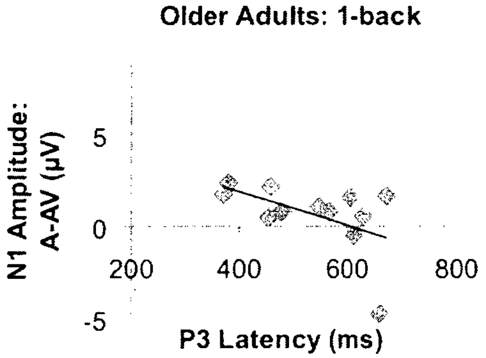
Appendix E

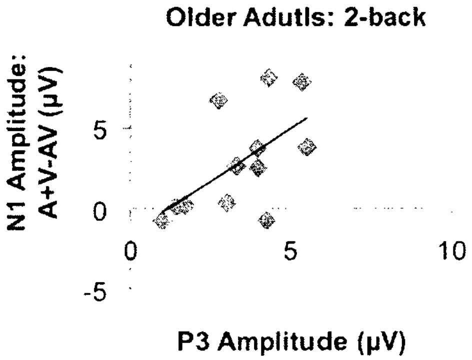

Younger Adults: 1-back
N1 Amplitude: A+V-AV (µV)
5
0
-5
60
70
80
90
100
Accuracy


Younger Adults: 3-back
N1 Amplitude: A-AV (µV)
5
0
-5
0
5
10
P3 Amplitude (µV)

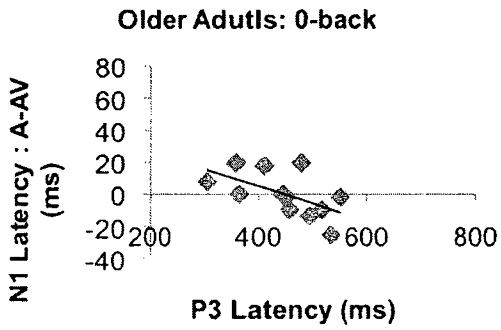

Older Adults: 2-back
N1 Amplitude: A-AV (µV)
5
0
-5
60
70
80
90
100
Accuracy

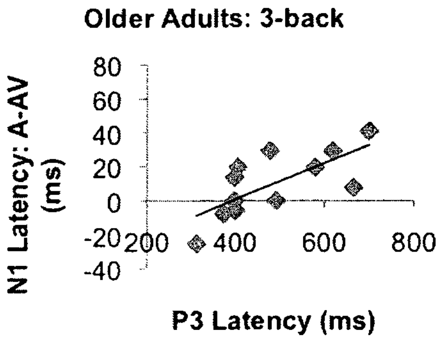

Older Adults: 2-back
N1 Amplitude: A+V-AV (µV)
5
0
-5
60
70
80
90
100
Accuracy

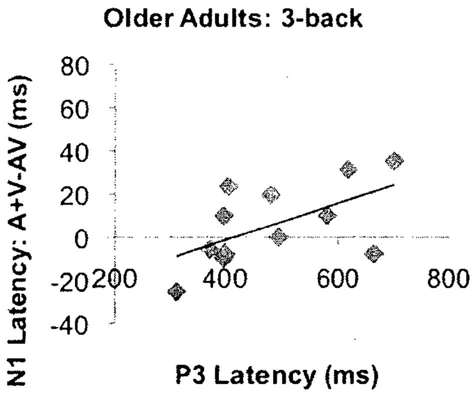

Older Adults: 1-back
N1 Amplitude: A-AV (µV)
5
0
-5
200
400
600
800
P3 Latency (ms)

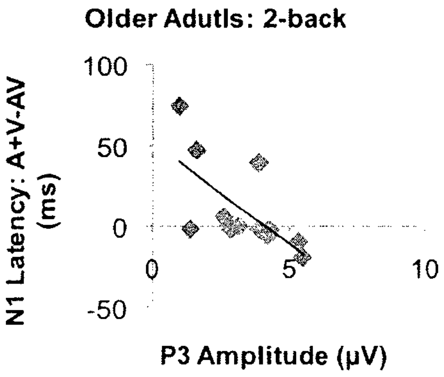

Older Adutls: 2-back
N1 Amplitude: A+V-AV (µV)
5
0
-5
0
5
10
P3 Amplitude (µV)

Appendix F


Older Adutls: 0-back


Older Adults: 3-back


Older Adults: 3-back


Older Adutls: 2-back