When eye meets ear:

An investigation of audiovisual speech and non-speech perception in younger and older adults

Axel H. Winneke

A Thesis in

The Department of Psychology

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Concordia University

Montreal, Quebec, Canada

November, 2009

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: 978-0-494-67365-2
Our file Notre référence
ISBN: 978-0-494-67365-2

ABSTRACT

When Eye Meets Ear: An Investigation of Audiovisual Speech and Non-Speech Perception and Age-Related Differences

Axel Winneke, Ph.D.
Concordia University, 2009

This dissertation addressed important questions regarding audiovisual (AV) perception. Study 1 revealed that AV speech perception modulated auditory processes, whereas AV non-speech perception affected visual processes. Interestingly, stimulus identification improved, yet fewer neural resources, as reflected in smaller event-related potentials, were recruited, indicating that AV perception led to multisensory efficiency. Also, AV interaction effects were observed at early and late stages, demonstrating that multisensory integration involved a neural network. Study 1 showed that multisensory efficiency is a common principle in AV speech and non-speech stimulus recognition, yet it is reflected in different modalities, possibly due to sensory dominance of a given task.

Study 2 extended our understanding of multisensory interaction by investigating electrophysiological processes of AV speech perception in noise and whether those differ between younger and older adults. Both groups revealed multisensory efficiency. Behavioural performance improved while the auditory N1 amplitude was reduced during AV relative to unisensory speech perception. This amplitude reduction could be due to visual speech cues providing complementary information, therefore reducing processing demands for the auditory system. AV speech stimuli also led to an N1 latency shift, suggesting that auditory processing was faster during AV than during unisensory trials.

This shift was more pronounced in older than in younger adults, indicating that older adults made more effective use of visual speech. Finally, auditory functioning predicted the degree of the N1 latency shift, which is consistent with the inverse effectiveness hypothesis which argues that the less effective the unisensory perception was, the larger was the benefit derived from AV speech cues. These results suggest that older adults were better "lip/speech" integrators than younger adults, possibly to compensate for age-related sensory deficiencies. Multisensory efficiency was evident in younger and older adults but it might be particularly relevant for older adults. If visual speech cues could alleviate sensory perceptual loads, the remaining neural resources could be allocated to higher level cognitive functions.

This dissertation adds further support to the notion of multisensory interaction modulating sensory-specific processes and it introduces the concept of multisensory efficiency as potential principle underlying AV speech and non-speech perception.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Natalie Phillips, for her tremendous support throughout the last couple of years. Her trust in my abilities helped me move forward and reinstate my confidence in my work when things were not going as planned. I could not have asked for a better supervisor.

I would also like to express my appreciation to all the lab members for helping me with countless ERP setups and for organizing much needed Sweets-days. In particular, I would like to thank the lab coordinators, Tsee-Leng, Sharon and Toby for making life easier. Special thanks go to Laura for agreeing to be the 'talking head' of the AV speech stimuli. The process was tedious but you did a fantastic job. Also, if it was not for you, I might have never played hallway-bowling.

I want to thank Dave Munro for his technical expertise and for helping us troubleshoot and for building fancy devices.

I am also thankful for the help I received from Shirley Black, Donna Craven and Pippa Ross. You helped me to find my way through the administrative jungle at Concordia.

I would also like to say thank you and Dankeschön to my parents and sisters for supporting me in countless ways. Without your support I would have never reached this point in my life.

Last but not least, I want to thank my wife, Annette. You have supported me all this time and allowed me to vent when times were tough and provided me with advice. Thank you for being so understanding and for believing in me. I am lucky to have you by my side and I could not have done this without you.

Table of Content:

List of Figures:

List of Tables:

List of Abbreviations

ANOVA=	analysis of variance
AE=	auditory enhancement
AEP=	auditory evoked potentials
A-only=	auditory-only
AV=	audiovisual
BOLD=	blood oxygenation level dependent
CDF=	cumulative distribution function
EEG=	electroencephalography/ electroencephalogram
EOG=	electrooculogram
ERP=	event-related brain potential
fMRI=	functional magnetic resonance imaging
HL=	hearing level
MEG=	magnetoencephalography/ magnetoencephalogram
OA=	older adults
PET=	positron emission tomography
PTA=	pure tone average
RSE=	redundant signal effect
RT=	response time
S/N=	signal-to-noise ratio
SPL=	sound pressure level
VE=	visual enhancement
VEP=	visual evoked potentials
V-only=	visual-only
WM=	working memory
YA=	younger adults

Authors' contributions

Study 1 and 2: Both studies were conceived by Axel Winneke together with Natalie Phillips. Axel Winneke created the stimuli, recruited participants, conducted the experiments, processed and analyzed the data. The tasks were completed under the supervision of Natalie Phillips. Axel Winneke and Natalie Phillips designed the experiments and interpreted the results. Axel Winneke wrote the complete first drafts of the manuscripts which were subsequently revised by Natalie Phillips.

Chapter 1: General Introduction

Traditionally, research on sensation and perception focuses on one modality in isolation. The amount of research that has been and is being conducted provides us with a rich understanding of the functioning of our sensory systems and creates a tremendous and invaluable insight into the underlying mechanisms of human sensation and perception. However, the world that surrounds us is full of sensory stimuli, stimuli that vary widely in their physical nature and that often stimulate more than one modality at a time. Imagine going to a farmer's market on a Saturday morning. You will SEE a colourful assortment of fruits. To make sure the produce you buy is ripe, you can TOUCH, SMELL, and TASTE it and by knocking on the outside of a watermelon you can even HEAR if it is ready to eat (Mierzejewski, 2009). This example illustrates the sensory diversity of our environment. The work in hand is not intended as guide on how to buy fruit but it discusses how modalities interact with each other.

Every object we perceive, be it a barking dog or a ripe banana, will stimulate more than one modality and even though the physical make-up of the sensory information that activates the highly specialized sensory-receptors is different, this information will be combined to form a coherent percept (Meredith, 2002). There is ample evidence in the multisensory literature illustrating that the brain not only combines information from multiple senses but that one modality can actually influence the processing in another. The findings reported in this dissertation will further elaborate on our understanding of the mechanisms involved in multisensory perception.

1.1 Organization of the Dissertation

The experiments presented here were designed to investigate processes of audiovisual (AV) stimulus identification. Two main questions guided this work, namely: 1) whether AV speech perception is processed differently than AV non-speech stimuli, and 2) whether processes responsible for AV speech perception change with age.

Chapter 2 will answer the first question and present results from two experiments investigating electrophysiological processing differences between AV stimulus recognition of speech and non-speech items. Findings from this study will provide a better understanding of basic mechanisms enabling sensory information from separate modalities to interact at the behavioural as well as neural level.

After addressing issues of a more basic nature, chapter 3 has a more applied character. Here, I will answer the second main question, namely whether the neural processes that underlie AV speech perception are the same in younger and older adults. I will also address the questions of whether the ability to combine auditory and visual speech cues remains intact in older adults, and to what extent older adults benefit from AV speech cues. The work presented here will go from the basic understanding of the processes involved in audiovisual object perception – AV speech and non-speech – to a more applied aspect of AV speech, and its implementation and relevance to communication in the aging adult. The overall organization of this work follows both a logical as well as a chronological order. However, before presenting the results regarding the experiments I conducted, I will review previous research to establish the necessary framework for my experiments. The review will briefly present well-known examples of how multisensory stimuli can influence human perception. Subsequently, behavioural effects associated with AV speech will be reviewed. This will be followed by an overview of results regarding the neural basis of AV speech and non-speech perception including findings from studies on animals and neurological patients, as well as functional neuroimaging research. The last section of the review will address potential mechanisms enabling the interaction of auditory and visual speech cues.

1.2 Framework

Jousmaki and Hari (1998) have shown that when participants were rubbing hands while an accompanying rubbing sound was played back but at an increased frequency, the participants' perception of roughness changed to the degree that it felt like rubbing a piece of parchment paper, the so-called parchment illusion. This shows how audition can influence somatosensation. In an experiment looking at the influence vision can have on olfaction, a group of wine experts were asked to describe the aroma of various wines (Morrot, Brochet, & Dubourdieu, 2001). What participants did not know was that the researchers had tinted white wine with odourless red dye. When asked to describe the aroma of the wine, participants described the wine with terms like pepper, chocolate, and plum, terms that are usually associated with red wines. Results clearly revealed that even the refined sense of smell of those experts was overruled by visual information and that vision interacts with olfaction. An example in the audiovisual domain is an experiment by Sekuler, Sekuler and Lau (1997). Participants watched two dots moving towards one another. On some of the trials a brief click tone was presented when both dots reached the center of the screen. The task of the participants was to describe the trajectory of the moving dots. Interestingly, when no tone was played, participants perceived the dots to cross and continue on their path. However, when the tone coincided with the time point when the dots arrived at the centre of the screen 60-70% of participants reported that dots were colliding and 'bounced off' of each other. These phenomena just described show that sensory modalities do not operate in isolation and that human perception can be influenced by multisensory interactions. Multisensory interaction effects have been reported in the domain of speech perception as well.

1.3 AV Speech Perception

Speech perception is a crucial factor for successful communication in social species like humans and is usually considered a process dominated by auditory processes rather than vision (Easton & Basala, 1982). Despite this auditory dominance, the well known MacDonald-McGurk effect reveals that the auditory speech system is not isolated from other sensory signals as visual speech cues have been shown to alter auditory perception (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). When participants in this study watched a video clip of someone saying /ga-ga/ but the audio track was dubbed with /ba-ba/, the vast majority of participants reported a fused perception namely that of /da-da/. However, when facing away from the monitor and only listening to the spoken syllables, participants identified the syllables correctly. It is a robust phenomenon that has been demonstrated frequently (Campbell, 2008), and it indicates that visual speech cues can influence auditory processes even in perfect listening conditions. It is important to note though that the MacDonald/ McGurk illusion only works for certain syllable combinations (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). One possible explanation is that /da/ is in terms of its visual properties closer to /ga/, but phonetically it is more similar to /ba/. In other words /da/ shares some commonalities with the other two (Summerfield, 1983). Despite the important implications of the MacDonald/ McGurk effect it is an illusion, rather artificial, and rarely encountered in everyday life (De Gelder & Bertelson, 2003).

However, there is longstanding evidence that audiovisual speech plays an important role in normal speech perception as well. Audiovisual speech refers to a situation during which you can HEAR your conversation partner but importantly you can also SEE her or his lips, face, and other non-verbal gestures. That is, in addition to auditory cues, visual speech cues are available, too. Given that auditory signals as well as visual speech signals from the mouth have the same origin there is a certain degree of information redundancy. However, there is only a partial overlap, which means that visual speech cues should not be considered as entirely redundant but rather as complementary (Campbell, 2006, 2008; Grant & Seitz, 2000b; Munhall & Vatikiotis-Bateson, 1998; Summerfield, 1979, 1987). More specifically, visual speech cues from the articulatory system including lips, tongue, and teeth deliver information regarding the place of articulation. These complementary signals can aid in the disambiguation of auditory signals. For example, a /v/ and /b/ are acoustically similar especially in a noisy environment, but seeing the upper and the lower lip close and touch as for the bilabial /b/ can help to distinguish it from a labio-dental /v/ where the low lip touches the upper teeth, and vice versa. Consequently, visual speech should lead to benefits especially when the auditory signal is distorted which is the case for individuals with hearing impairments or when listening to speech in noisy environments (Summerfield, 1987).

On the other hand, the visual system is not well equipped to pick up cues regarding manner of articulation such as voicing, which refers to the amount of vibration of the larynx. For example, /t/ and /d/ look similar in their visual properties, but the former is unvoiced and the latter is voiced. Also, the amount of nasality, which refers to the degree to which the oral and nasal cavities are coupled in order to produce a speech sound, is more clearly conveyed via the auditory system (Summerfield, 1983).

Even though the auditory modality is more dominant in speech perception than vision (Easton & Basala, 1982), visual speech cues or visemes can help to clarify ambiguous speech sounds or phonemes particularly in noisy environments or when hearing is impaired. For example, when talking to someone on the phone, phonemes /l/ and /r/, as in *'grass'* and *'glass'*, can be easily confused, but the difference becomes more obvious when visual cues or visemes are perceivable as well. A phoneme is the smallest unit of auditory speech that enables to distinguish between the meanings of spoken words, whereas a viseme is the counterpart for visual speech. Even though the availability of visual speech cues can improve speech perception, language comprehension based on visual speech cues alone, called lip- or speechreading, is very challenging. The reason for this difficulty stems from the fact that visemes can be ambiguous because one viseme does not correspond to just one phoneme, but instead several phonemes share the same viseme (Campbell, 2008; Erber, 1974; Summerfield, 1987). For example, the consonants /p/, /b/, and /m/, as in 'pan', 'ban' and 'man', share the same visual cues and are therefore grouped in the same visemic category (Summerfield, 1983). Therefore, it should be clearly stated that the role of vision in speech perception should be regarded as complementary to the more dominant auditory processes.

More than 50 years ago, Sumby and Pollack (1954) conducted an experiment on audiovisual (AV) speech perception in white noise. In one condition participants with normal hearing had to listen to spoken words (auditory (A) only) and identify them based on a list containing 8 to 256 words. In the AV condition they saw and heard the speaker. The trials were conducted in various signal/noise (S/N) ratios. The findings revealed a clear AV benefit as accuracy scores were higher for the AV condition than for A-only. Upon investigation of identification rates as a function of S/N, it was shown that the AV trials led to benefits equal to intensity increases of up to 10-15dB in the A-only condition. A recent reinvestigation of this study highlighted that this benefit is not equal at all S/N ratios (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). It was shown that the largest gain derived from the AV mode was at -12dB difference between signal and background noise. Other studies have also found AV speech benefits in noisy environments (e.g.: Callan et al., 2003; Erber, 1969; Helfer, 1997, 1998; Schwartz, Berthommier, & Savariaux, 2004; Sommers, Tye-Murray, & Spehar, 2005; Summerfield, 1979). In line with the findings of AV benefits in suboptimal listening environments, there is evidence that patients with hearing impairments benefit from the availability of visual speech cues (e.g.: Bergeson & Pisoni, 2004; Erber, 1974, 2002; Grant & Seitz, 1998; Grant, Walden, & Seitz, 1998; Hay-McCutcheon, Pisoni, & Kirk, 2005; Möbes et al., 2006; Rouger et al., 2007; Tillberg, Ronnberg, Svard, & Ahlner, 1996).

Importantly, there are also data suggesting that speech cues improve speech perception in healthy adults in optimal hearing environments in terms of faster reaction times (e.g.: Besle, Fort, Depuelch, & Giard, 2004; van Wassenhove, Grant, & Poeppel, 2005) or enhanced comprehension of a difficult subject matter (Arnold & Hill, 2001; Reisberg, McLean, Goldfield, Dodd, & Campbell, 1987). Visual cues are not only derived from the mouth but also from the eyes, forehead, and head movement and it has been shown that the availability of those cues influence speech perception as well (Davis & Kim, 2006; Munhall & Vatikiotis-Bateson, 1998). Rhythmic head movement has been linked to prosodic features of speech such as stress or emphasis of words (Hadar, Steiner, Grant, & Rose, 1983) and head motion seems to be correlated with the fundamental frequency and amplitude of the speaker's voice (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). According to Davis and Kim (2006) the advantage derived from the upper part of face could be due to segmentation cues, which help the listener to parse the continuous speech stream. The above mentioned findings suggest that visual speech cues interact with auditory speech signals and that this multisensory interaction improves speech perception particularly under impoverished listening conditions. The question that emerges is how the brain deals with multiple signals from different modalities and where in the brain those multisensory interactions take place.

1.4 Brain Areas of Multisensory Processing

The processing of sensory signals form the outside world involves different areas in the brain. Sensory-specific cortices are said to be specialized in processing a specific type of sensory signal. The auditory cortex in the superior section of the temporal lobes specializes in processing of auditory signals (Bushara et al., 1999; Goldstein, 2007; Rauschecker & Tian, 2000; Romanski et al., 1999), whereas the occipital cortex is designated to process visual signals (Gazzaniga, Ivry, & Mangun, 2002; Goldstein, 2007; Zeki, 1978). In contrast to sensory-specific or unimodal areas, multimodal association cortices receive information from multiple senses subsequent to sensory-specific processes. Audiovisual convergence zones have been found in the temporoparietal cortex, parietal cortex, but also in premotor and prefrontal regions (Bushara et al., 1999; Calvert & Lewis, 2004; Gazzaniga et al., 2002; Romanski et al., 1999).

1.4.1 Animal Studies

Benchmark studies by Stein and Meredith (1993) provided important results on multisensory integration in cats. Single-cell recordings from the cat's superior colliculus (SC) revealed the existence of multisensory neurons. Those multisensory neurons responded to somatosensory, auditory and visual stimuli presented individually, but importantly those cells also responded to stimuli from different modalities when presented simultaneously. For 45% of those neurons the firing rates to multisensory stimuli was larger than the sum of the responses to the unisensory stimuli (e.g., AV > A + V), which is called superadditivity. A subadditive response pattern (e.g., AV < A + V) was observed in 20% of multisensory cells. That is, the majority of multisensory cells in the SC of the cat indicated non-linear multisensory interaction (Meredith & Stein, 1986). An important aspect that one has to keep in mind is the fact that multisensory convergence (i.e., that several sensory-specific neurons project onto one multisensory neuron) is necessary but not sufficient to show multisensory neural integration (Calvert, 2001; Meredith, 2002). In addition to multisensory convergence, a neuron that responds to multiple modalities must also display a differential response during multisensory processing as compared to when processing unimodal information (Calvert, Campbell, & Brammer, 2000; Meredith, 2002), which is a sign of integration.

Further experiments led to the proposal of the principle of inverse effectiveness (Meredith & Stein, 1983; Stein & Meredith, 1993). This principle states that the less effective a unisensory stimulus is in eliciting a response, the more effective a combined multisensory stimulus will be. In a spatial orientation task neural as well as behavioural responses (i.e., percent accurate orientation responses) adhered to inverse effectiveness (Stein, Huneycutt, & Meredith, 1988). That is, the largest behavioural benefits were seen for AV stimulus combinations that consisted of auditory and visual stimuli that were not effective on their own. One possible explanation for this correspondence between neural activity and behavioural responses is the existence of direct connections between cells in the SC of cats to motor areas controlling movement involved in orientation (i.e., head, eyes, and ears) (Meredith & Stein, 1985). Subsequent studies replicated the presence of this principle in the superior colliculus of cats (e.g.: Stanford, Quessy, & Stein, 2005) and rhesus monkey (e.g.: Wallace, Wilkinson, & Stein, 1996) suggesting the existence of subcortical neurons capable of integrating sensory information from separate modalities. Intracranial cell recordings in rats have revealed multisensory activation patterns in secondary visual and auditory cortices in response to audiovisual stimuli consisting of clicks and strobe light (Barth, Goldberg, Brett, & Di, 1995).

To follow up on findings suggesting multisensory interaction at early sensory processing stages in humans (see below) Falchier and colleagues (2002) conducted tracer studies in macaque monkeys. They found direct connections between the belt, parabelt and the superior temporal plane of the auditory cortex and primary visual cortex (V1).

Such a finding is important as it indicates that multisensory modulations could occur at early sensory specific processing stages rather than at or in addition to hierarchically higher multimodal areas (Cappe & Barone, 2005; Ghazanfar & Schroeder, 2006). That is, audition could modulate vision, or vice versa, directly rather than indirectly. Support for this comes from neuronal response modulations in V1 of rhesus monkeys following AV stimulation (Wang, Celebrini, Trotter, & Barone, 2008) as well as in temporal auditory areas (Watanabe & Iwai, 1991). Interconnectivity of primary sensory cortices has also been found in other species such as the Mongolian gerbil (Budinger, Heil, Hess, & Scheich, 2006; Budinger, Laszcz, Lison, Scheich, & Ohl, 2008), ferrets (Bizley, Nodal, Bajo, Nelken, & King, 2007) and the Prairie Vole (Campi, Bales, Grunewald, & Krubitzer, 2009) adding further support to the notion that areas that have traditionally been regarded as sensory-specific, have multisensory capabilities.

As mentioned above, gestures are important for human communication. However, other animals like monkeys for example also possess an array of facial gestures. These facial gestures accompany certain communicative vocalizations uttered in a friendly atmosphere (i.e., 'coo'-calls) or hostile threatening context (i.e., threat-calls) (Ghazanfar & Logothetis, 2003). Tested with a preferential-looking paradigm, monkeys spent more time looking at auditory and visual parings of sound and gestures when the combinations matched as compared to mismatching AV pairings. According to the authors, this suggests that not only humans combine auditory and visual communication cues but that primates possess similar capabilities for AV integration (Ghazanfar & Logothetis, 2003). A follow-up study to investigate the neural basis of this AV interaction effect in primates revealed modulations of neural responses in primary and secondary auditory cortex when vocalizations were accompanied by facial gestures (Ghazanfar, Maier, Hoffman, & Logothetis, 2005). Despite the fact that animal studies provide important insights into neural structures and processes involved in multisensory processing, it is not necessarily the case that the human brain functions the same way. Particularly AV speech can only be directly examined in human participants given that a language system with a similar complexity is likely missing in primates and other animals. With the development of sophisticated imaging techniques, researchers are now able to visualize brain activity in healthy participants while performing a particular task of interest such as multisensory perception.

1.4.2 Human Studies

Results regarding multisensory interaction sites in human participants using non-invasive neuroimaging techniques vary. Calvert and colleagues (1999) used functional magnetic resonance imaging (fMRI) to measure the blood oxygenation level dependent (BOLD) response during a passive AV speech study. Stimuli consisted of spoken digits presented unimodally (A-only and V-only) or bimodally (AV). Multisensory interactions were evident in the visual area V5/ MT, relevant for motion perception, as well as in primary and secondary auditory cortex.

In another fMRI study on AV speech perception participants passively listened to and/ or watched someone narrate sections of George Orwell's novel *'1984'* (Calvert et al., 2000). The analyses identified the left superior temporal sulcus and again V5 as integration site of auditory and visual speech stimuli. The superior temporal sulcus as well as the superior temporal gyrus have been identified as site of multisensory integration by other fMRI and positron emission topography (PET) studies investigating the neural basis for AV speech processing (Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Callan et al., 2003; Callan et al., 2004; Calvert, 2001; Kang et al., 2006; Kawashima et al., 1999; Macaluso, 2006; Macaluso, George, Dolan, Spence, & Driver, 2004). Areas in the posterior and inferior parietal lobe are other sites where multisensory interaction effects have been observed (Macaluso et al., 2004; Saito et al., 2005). Studies investigating the neural structures involved in speechreading in hearing adults discovered that watching someone speak led to significant activations in the auditory cortex, including parts of Heschl's gyrus and superior temporal sulcus (Calvert et al., 1997; Calvert & Campbell, 2003; Pekkola et al., 2005). In other words, areas that are thought of as specific to auditory processing were recruited during silent lipreading. The superior temporal sulcus was also activated in a different study that measured BOLD responses to gestures of British Sign Language (MacSweeney et al., 2004). The findings regarding the neural basis of speechreading and sign language demonstrate that auditory areas do not only respond to auditory signals but are also tuned to signals that are relevant for visual communication cues such as lipreading and sign-language.

Data from patients with acquired brain damage have further contributed to our understanding of which cortical regions play a role in AV speech processing. A patient with damage to the left temporo-occipital area revealed intact face recognition and speech comprehension, but had marked deficiencies in speechreading (Campbell, Landis, & Regard, 1986). When presented with a MacDonald/ McGurk illusion the patient appeared completely immune to the illusion and consistently perceived the auditory syllable.

Another patient who had a similar damage but in the right hemisphere suffered from prosopagnosia, but despite the inability to recognize faces or identify facial expressions this patient was susceptible to the McGurk illusion. This double dissociation suggests that the left temporo-occipital region plays an important role in AV speech perception and that it is independent of mechanisms relevant for face recognition (Campbell, 1992; Campbell et al., 1986). Another area that has been shown to be critical for intact lipreading ability is V5 in the occipito-temporal cortex. A patient with damage to V5 showed deficits in motion perception and was not able to extract speech information from dynamically moving lips (Campbell, Zihl, Massaro, Munhall, & Cohen, 1997). When exposed to incongruent AV speech the patient reliably repeated the auditory syllable suggesting that visual speech cues did not interact with auditory speech perception. Findings from neurological patients show that the temporo-occipital cortex is important for speechreading and that lesions can disrupt interactions between visual and auditory speech cues.

The fact that various areas have been proposed to be engaged in multisensory processing indicates that integration of signals from distinct modalities is likely achieved via a network comprised of sensory specific as well as higher level association and multimodal regions in frontal and parietal cortex.

The neural basis of AV integration has also been investigated outside the domain of speech and language. For example, Laurienti and colleagues (2002) presented visual checkerboards and auditory sound bursts of white noise separately and simultaneously while measuring the BOLD response. The presentation of auditory stimuli led to a reduction of activity in visual cortical regions and the presentation of visual stimuli led to similar 'deactivations' of auditory cortical areas. Responses to AV trials revealed multisensory interaction in visual and auditory areas as responses were larger than the sum of unisensory activation. These results suggest that sensory specific cortices were modulated by stimuli presented in another modality. However, given the experimental design an alternative explanation could be attentional shifts away from the modality that was not stimulated during unisensory trials.

A study looking into the neural effects underlying sound-induced change of visual motion perception (i.e., similar to Sekuler and colleagues (1997)) found BOLD response reductions in temporal auditory and occipital visual cortical areas during audiovisual trials (Bushara et al., 2003). The opposite response pattern, namely activity increase, was found for multimodal areas in the frontal lobe and parietal cortex. This distributed pattern of activity in response to audiovisual stimuli suggests that multisensory processing involves a complex network of cortical areas, including those traditionally regarded as sensory-specific in nature. Presenting checkerboards and sound bursts individually or together, Calvert and colleagues (2001) also found a wide network of areas revealing multisensory interaction in response to AV stimuli. Interaction sites were again found in frontal areas as well as superior temporal sulcus but not in the occipital cortex. The largest interaction effects were found in the superior colliculus in the form of superadditivity. The fact that the superior colliculus was identified would suggest that auditory and visual signals already interact before they reach the cortex. As mentioned above, these early, subcortical interaction effects have also been shown in animals (Stein & Meredith, 1993; Wallace et al., 1996). The superior colliculus in cats has been shown to receive projections from sensory cortical areas (Jiang, Wallace, Jiang, Vaughan, &

Stein, 2001; Wallace, Meredith, & Stein, 1993), suggesting that the multisensory interaction effect observed by Calvert and colleagues (2001) could possibly be due to late cortical feedback modulations rather than early upstream interaction effects.

To investigate multisensory integration of auditory and visual information belonging to the same real-world object (e.g., tools and animals), the BOLD response was measured during an object identification task (Beauchamp, Lee, Argall, & Martin, 2004). The posterior superior temporal sulcus and the middle temporal sulcus were the only sites identified as multisensory interaction as, as those areas were more active during AV trials compared to the unisensory responses. According to the authors these areas are good candidates for multisensory feature integration, given that they border the sensory specific areas of visual and auditory cortices.

Taken together, the results obtained from fMRI and PET studies vary in terms of which areas were identified as multisensory. The reason for that could be that studies differed in their stimuli, the task and task demands, but also the fact that criteria for multisensory integration were not homogeneous across studies contributes to the variance. Even though findings are not consistent, it is obvious that multisensory processing involves a network of areas including sensory specific and multimodal cortical regions.

Given the poor temporal resolution of the BOLD response, fMRI is less able to establish the sequence of activation to see whether interactions in unisensory areas occur before interaction effects in hierarchically higher multimodal areas or after. This information is useful in order to speculate on whether connections are feedback or feed-forward. Other techniques, such as the recordings of electroencephalograms (EEG) or magnetoencephalograms (MEG), have an excellent temporal resolution and can therefore provide information regarding the timing of neural processes.

These non-invasive electrophysiological techniques have also been used to study the neural brain processes underlying multisensory perception. Unlike fMRI, which is a relatively recent technique, the electroencephalogram (EEG) (i.e., recordings of ongoing electrical activity of the brain) was developed in the 1920s by Hans Berger (Jung & Berger, 1979; Zifkin & Avanzini, 2009). Event-related brain potentials (ERPs) are extracted from the EEG and, given the importance of this technique for this dissertation, it merits a brief description.

ERPs are derived from an EEG by averaging the electrophysiological responses to the same stimulus or class of stimuli during a specified time window surrounding the stimulus of interest. By presenting the same (type of) stimulus many times, random activity (i.e., activity unrelated to the event of interest) will cancel each other out and what is left is activity related to the stimulus of interest – the ERP response. A typical ERP response is visualized as a waveform that shows a series of peaks and troughs that indicate voltage changes. These deflections can be assessed in terms of their electrical amplitude or voltage as well as their latency. In addition to their latency and amplitude, deflections of an ERP waveform can also be described in terms of the topographical scalp distribution of electrical activity they are associated with. Topographical distributions can differ depending on the underlying neural generators. These voltage deflections are sometimes referred to as ERP components but the term component is not clearly defined. Voltage deflections are often due to simultaneous activity of more than one generator, which could be considered as the actual ERP components. This would be a more neurophysiological definition. Alternatively, a component can be defined in terms of its function, which in turn depends on the experimental design (Coles & Rugg, 1995). With respect to the latter definition a voltage deflection in the ERP waveform can be defined as a component if the underlying neural generators are related to the same cognitive or sensory processes.

An EEG picks up electrical activity of large clusters of neurons that have the same orientation in terms of their polarity. This so-called open-field configuration enables electrodes on the surface of the scalp to pick up the summed electrical activity of neurons (Luck, 2005). The recorded neuronal activity does not stem from action potentials but rather from the sum of post-synaptic potentials. However, in order to elicit a measurable voltage level at the scalp, large populations must be active at the same time, neurons must be spatially aligned and they must receive the same input (i.e., all excitatory or all inhibitory neurotransmitters). Pyramidal cells in the cortex contribute largely to the EEG signal given that their orientation is perpendicular to the cortical surface (Luck, 2005). Other structures like the thalamus for example do not have an open-field configuration, which means that these structures do not contribute to the EEG signal (Coles & Rugg, 1995).

The fact that the electrodes are at a distance relative to the source of activity, called a dipole, poses a problem for the technique of EEG/ ERP. The head and brain are a conductive medium, but neural tissue, the skull, and the scalp present obstacles to the free flow of electrical current. Given that the head has conductive properties and because electricity follows the path of least resistance, activity picked up at one electrode does not necessarily imply that the source of that activity is in near proximity as well (Coles &

Rugg, 1995). This issue leads to the main problem studies using EEG/ERP have to deal with, namely the inverse problem (Luck, 2005). The inverse problem refers to the problem of trying to infer the location of the dipole(s) based on the observed topographical voltage distribution. Given the number of unknown variables (e.g., orientation of the dipole, strength of the dipole and number of dipoles) makes the localization of the relevant dipole(s) mathematically challenging. To be precise, there are an infinite number of dipole configurations that can produce any voltage distribution (Luck, 2005). On a positive note, increased understanding of neurophysiology, improved mathematical models and combining fMRI and EEG data has made source localisation of ERP dipoles more reliable, as additional information can place constraints on the number of plausible dipole locations. Nevertheless, fMRI is superior to ERPs in terms of spatial resolution. However, the advantage of ERPs is their extraordinary temporal resolution in the range of milliseconds. ERPs therefore enable to measure the time point at which differences between conditions occur which, if the experiment is carefully designed, allows detecting the onset of a particular sensory, motor or cognitive process.

The current work focused on auditory and visual evoked potentials. Early auditory ERPs consist of a series of components, namely the P1, N1 and P2 (Vaughan & Ritter, 1970). Components are labelled according to their polarity, with P referring to a positive and N to a negative amplitude, and their sequential order. Sometimes components are also labelled according to their peak latency. The N1 is sometimes called the N100 as it tends to peak at around 100ms after stimulus onset. The auditory P1 peaks around 30-100 ms after stimulus onset and is followed by the N1 which peaks between 90 to 150 ms after stimulus onset and both are largest at central electrode sites around the vertex

(Eggermont & Ponton, 2002; Picton et al., 1999; Yvert, Fischer, Bertrand, & Pernier, 2005). Exact locations of the dipoles for P1 and N1 vary but a likely source for P1 is at the border of Heschl's gyrus whereas the N1 is suggested to have its dipole source in the planum temporale of the superior temporal gyrus in or near primary auditory cortex (Eggermont & Ponton, 2002; Hyde, 1997; Näätänen & Picton, 1987; Picton et al., 1999; Yvert et al., 2005). The P1 and N1 are automatic brain responses elicited by the onset of a sound and their amplitudes are modulated by attention, with attended stimuli eliciting larger amplitudes than unattended stimuli. The P1-N1 complex has been shown to be involved in feature analysis and it increases in amplitude with increasing stimulus intensities (Näätänen & Picton, 1987). Interestingly, The N1 has been shown to be sensitive to stimulus predictability, with the auditory N1 amplitude decreasing if the upcoming auditory stimulus is predictable (Näätänen & Picton, 1987). The auditory P2 peaks at around 200 ms after stimulus onset and is largest in amplitude at central electrode sites. It is assumed to be involved in higher order feature analysis as it is modulated by sound complexity (Shahin, Roberts, Pantev, Trainor, & Ross, 2005) and its neural source is said to be in secondary auditory cortex (Eggermont & Ponton, 2002; Picton et al., 1999).

Visual evoked potentials consist of the same component sequence as auditory evoked potentials, namely a P1-N1-P2 complex. The visual P1 peaks between 60-130 ms followed by the N1 peaking at around 140-160 ms after stimulus onset. Visual P2 reaches its maximum amplitude around 200ms. All three components are largest at occipital electrode sites with a slight right hemispheric dominance (Hillyard, Mangun, Luck, & Heinze, 1990; Mangun, Hillyard, & Luck, 1993). The visual P1 has been shown to be involved in stimulus detection (Luck, Heinze, Mangun, & Hillyard, 1990; Luck & Hillyard, 1994; Luck & Hillyard, 1995) and the role of the visual N1 is related to early feature discrimination such as colour and form (Hopf, Vogel, Woodman, Heinze, & Luck, 2002; Lobaugh, Chevalier, Batty, & Taylor, 2005; Murray et al., 2002; Vogel & Luck, 2000) and possibly even object categorization (Eimer, 2000; Rossion, Joyce, Cottrell, & Tarr, 2003; Thorpe, Fize, & Marlot, 1996). The visual P2 is also involved in stimulus analysis and related to feature or, more generally, target detection (Luck & Hillyard, 1994). In other words, as for the early auditory evoked potentials, visual evoked potentials reflect early visual processing. Visual evoked potentials, just like auditory ones, are larger to attended than unattended stimuli which has been interpreted as 'sensory gain' (Hillyard & Anllo-Vento, 1998; Hillyard et al., 1990; Mangun, 1995; Mangun et al., 1993). The P1 has been localized in lateral occipital, extrastriatal regions (Clark, Fan, & Hillyard, 1994; Di Russo, Martinez, & Hillyard, 2003; Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2001; Mangun et al., 1993) and the N1 in ventral occipito-temporal regions (Di Russo et al., 2003; Di Russo et al., 2001; Hopf et al., 2002; Murray et al., 2002). The source of the P2 is suggested to lie in the posterior parietal cortex or dorsal anterior occipital cortex (Clark et al., 1994).

A number of studies have measured electrophysiological responses in order to study the processes of audiovisual interaction during audiovisual speech and audiovisual non-speech object perception. Early studies on the electrophysiological effects of simultaneously processing stimuli from separate modalities (i.e., auditory sound, visual flash, electric shock) revealed a multisensory amplitude reduction of the N1 and P2 (Davis, Osterhammel, Wier, & Gjerdingen, 1972; Hay & Davis, 1971). Interest in electrophysiological processes of AV interaction emerged again about two decades ago (Sams et al., 1991) with a strong focus on multisensory spatial attention (Alho, Woods, & Algazi, 1994; for review see: Eimer & Driver, 2001; Eimer & Schröger, 1998). Research using EEG/MEG to investigate processes of multisensory stimulus identification, including AV speech perception, intensified only over the last 10 years and, because studies vary in choice of stimuli and task demands, findings are not homogeneous.

In an AV object recognition study, Giard and Peronnet (1999) presented ellipses that differed in their shape, as well as tones of various frequencies. Participants were instructed to learn two different objects defined by different shape-tone combinations and in the experimental task participants were asked to recognize the learned objects. Responses were more accurate and faster during AV trials compared to unisensory trials and the authors found small amplitude enhancements for the auditory N1. However, this latter interaction pattern was only evident for a subset of participants whose weaker or less dominant modality, as defined by unisensory reaction times, was audition. A more robust finding was an amplitude reduction for the visual N1 for AV trials evident in all participants which was interpreted as an indication of a decrease in energy demand for visual processes.

The results by Giard and Perronet (1999) were extended by replicating their findings of a visual N1 reduction alongside behavioural benefits associated with AV stimuli in an object recognition task including non-redundant auditory and visual stimulus combinations (Fort, Delpuech, Pernier, & Giard, 2002). No multisensory modulation of the auditory N1 was reported. A visual N1 amplitude reduction together with reaction time benefits in response to audiovisual stimuli was also reported by

Molholm and colleagues (2002). An additional multisensory interaction effect was found at parieto-central sites around 120ms after stimulus onset, but topographical analyses and comparisons to the ERP response to A-only stimuli ruled out that this interaction reflected a modulation of the auditory N1. Even though their study required a simple stimulus detection task rather than object identification, it revealed that concurrent auditory information can modulate visual processes.

To further investigate multisensory object recognition using ecologically more valid stimuli than arbitrary sound and shape combinations (Fort et al., 2002; Giard & Peronnet, 1999), Molholm and colleagues (2004) presented animal pictures and animal vocalizations. Their target detection task revealed multisensory benefits in terms of reaction times and early multisensory modulations of the visual N1 at right-posterior electrode sites. Subsequent dipole source analyses located the source of the interaction effect in right lateral occipital complex of the ventral visual stream, which has been shown to be relevant for object identification tasks (Ishai, Ungerleider, Martin, & Haxby, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Ungerleider & Haxby, 1994). No multisensory interaction effects were evident for auditory evoked potentials.

A more recent study looking at AV non-speech processes found audiovisual modulations of the auditory N1 component (Stekelenburg & Vroomen, 2007). Participants watched video clips of an actor clapping his hands or tapping a spoon against a cup. To ensure participants were attending the video clips, they had to detect infrequently presented irrelevant targets presented on the screen. However, there was no task that would have allowed for assessment of multisensory interaction at the behavioural level. The ERP results revealed that AV trials lead to an amplitude reduction of the auditory N1 relative to summed responses of unisensory trials (i.e., just watching someone clap (V-only) plus just hearing someone clap (A-only)). In addition to the amplitude reduction, a speeding of the auditory N1 peak latency was evident. The study's goal was to investigate whether AV non-speech processes differ from those underlying syllables presented audiovisually (i.e. AV speech). The findings revealed that, as was the case for AV non-speech stimuli, the auditory N1 following AV speech trials was reduced and peaked earlier relative to the summed responses of syllables presented unimodally. The lack of differences led the authors to conclude that multisensory processes involved in AV speech perception matched those for non-speech processing suggesting that AV speech is not special (Stekelenburg & Vroomen, 2007).

Other ERP studies on audiovisual (AV) speech presenting individually spoken syllables found amplitude reductions of auditory ERP components in response to AV syllables contrasted with responses to auditory only (A-only) or with the summed response of unisensory conditions (A + V) (Besle et al., 2008; Besle et al., 2004; Pilling, 2009; Reale et al., 2007; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). In addition to a reduced N1 amplitude in response to AV stimuli, a subset of those studies also found earlier N1 peak latencies, indicating faster auditory processing (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). One explanation for the amplitude reduction and latency shift in AV speech studies has been explained in terms of the predictability of the auditory stimulus due the preceding visual cues provided by the onset of lip movements (Besle, Bertrand, & Giard, 2009; Besle et al., 2008; Hertrich, Mathiak, Lutzenberger, & Ackermann, 2009; Hertrich, Mathiak, Lutzenberger, Menning,

& Ackermann, 2007; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). This is outlined in more detail further below.

Another neuroimaging technique that has been employed to study multisensory interactions is magnoencephalography (MEG). This technique is similar to EEG but it measures the magnetic fields perpendicular to the flow of electric current. It has the same excellent temporal resolution as EEG, but MEG is superior to EEG/ ERP in its ability to localize dipoles because the magnetic signal is not obstructed by neural tissue or the skull. Studies using MEG found AV related modulations of auditory processes and results indicated the auditory cortex to be the origin of the auditory N1 and of the AV interaction effects (Hertrich et al., 2009; Möttönen, Schürmann, & Sams, 2004). Furthermore, in addition to modulations of early auditory areas, multisensory interaction effects were also found at a later stage (around 300ms after stimulus onset) (Möttönen et al., 2004). The location of this latter interaction was in the superior temporal sulcus. This could suggest that multisensory interaction effects spread from early primary sensory areas to later hierarchically higher areas (Möttönen et al., 2004). The conclusion of the auditory cortex as the site of multisensory interaction during AV speech perception is in agreement with conclusions from other studies (Besle et al., 2009; Besle et al., 2008; Besle et al., 2004; Hertrich et al., 2009; Hertrich et al., 2007; Pilling, 2009; Reale et al., 2007).

Additional support for the auditory cortex being the location of audiovisual interactions comes from studies using mismatch negativity (MMN) paradigms. The MMN is an auditory ERP component that is elicited when an auditory stimulus deviates from another frequently presented auditory stimulus. The MMN is an indicator of preattentive detection of change in the auditory landscape and its source is presumably the auditory cortex (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). For example, Colin and colleagues (2002) have used the illusory McGurk-McDonald effect to investigate the role of visual speech in altering auditory perception. They found that during the AV condition in which participants watched and heard a speaker utter the same syllable repeatedly, changing the visible speech (e.g., from /bi/ to /gi/) elicited a MMN even though the auditory stimulus was identical (i.e., /bi/). A visual only condition did not yield a MMN suggesting that the effect found in the AV condition was not due to a simple change in the visual stimulus. Instead it demonstrates that the visual speech cue modulated the phonetic perception and that the MMN was elicited by an illusory, not a real, auditory change (Colin et al., 2002).

This is in line with other findings measuring the MMN in AV speech contexts (Besle, Fort, & Giard, 2005; Hertrich et al., 2007; Kaiser, Hertrich, Ackermann, Mathiak, & Lutzenberger, 2005; Kislyuk, Mottonen, & Sams, 2008; Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007; Sams et al., 1991). The results of ERP studies on AV speech suggest that multisensory interactions take place in the auditory cortex. However, as mentioned before, EEG/ ERP and MEG have to deal with the inverse problem. Support for the accuracy of those dipole localizations based on scalp recordings comes from intracranial recordings in human patient populations.

Intracranial recordings have obvious benefits in their spatial resolution but, given the invasiveness of this procedure, these studies are rare. Reale and colleagues (2007) recruited a group of epilepsy patients with permanently implanted electrodes covering auditory responsive sections of the superior temporal gyrus. The stimuli consisted of syllables that were presented in an A-only, V-only or AV mode. AV syllables were either congruent (i.e., the sound matched the lips) or incongruent. Lipreading did not yield significant activation in the auditory cortical regions, as would have been consistent with the BOLD studies above (Calvert et al., 1997; Calvert & Campbell, 2003; Pekkola et al., 2005), but auditory N1 responses were reduced during AV conditions relative to A-only, a finding that is consistent with scalp recordings of ERP studies on AV speech (Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). No early processing differences were found between responses to congruent and incongruent AV syllables. Another AV speech study recording intracranial ERPs from electrodes permanently implanted in the temporal lobe of epilepsy patients provided valuable information regarding the sequence of activations during AV speech (Besle et al., 2008). The results indicated that visual speech cues activated motion sensitive area V5 first and then secondary auditory cortex. During AV speech trials, auditory evoked responses in secondary auditory cortex were reduced. These interaction effects started early and modulated an ERP component at around 120ms after sound onset. This component likely corresponds to the auditory N1 which has been shown to decrease in amplitude during AV speech studies (Besle et al., 2008; Pilling, 2009; van Wassenhove et al., 2005). Results from the two studies recording from intracranial electrodes in epilepsy patients are very valuable as they bridge the gap between cell recordings in animals and neuroimaging data from humans. Furthermore, they confirm conclusions drawn from EEG and MEG studies that visual speech cues during AV speech perception modulate auditory responses in the auditory cortex in form of an amplitude reduction.

The previous section reviewed results regarding the neural underpinnings of multisensory processing in both animals and humans. Even though the human literature has readily adopted multisensory integration principles derived from single cell recordings in cats and monkeys, this transfer has to be regarded with care (Laurienti, Perrault, Stanford, Wallace, & Stein, 2005; Schroeder et al., 2003). The main concern is that neuroimaging techniques (fMRI, PET; MEG, EEG/ERP) are fairly crude in that they record activity from large groups of neurons. Logothethis and colleagues (2001) in a landmark study has shown, for example, that the BOLD response corresponds best to local field potentials which arise from the pooled activity of multiple cells. Single cell recordings can pinpoint individual neurons that respond to unisensory as well as multisensory stimuli. However, when activity comes from a relatively large area with millions of individual neurons it is not certain whether multisensory integration has occurred at the level of individual cells (Laurienti et al., 2005; Schroeder et al., 2003). The response to a multisensory stimulus might differ from the unisensory responses because 1) the same neurons are activated but in a different manner (e.g., multisensory integration) or 2) new neurons are recruited in addition to or instead of neurons activated by unisensory trials. The second case would not necessarily reflect neural integration at the level of individual cells. However, because it is the pooled activation of a large population of neurons that is being recorded, a particular area might still meet the criterion for multisensory interaction if the multisensory response (e.g., AV) differs from the additive model of the unisensory responses (e.g., A+V).

Findings from single cell recordings and anatomical tracer studies in animals, combined with modern neuroimaging data in humans provide compelling support that one modality can modulate early sensory processes in another. This seems to be the case for non-speech as well as speech perception and particularly within the realm of AV speech perception different models have been proposed regarding the mechanisms that lead to multisensory interaction effects.

1.5 Mechanisms of AV Speech Perception

Based on findings from an ERP study on AV speech, van Wassenhove and colleagues (2005) proposed an analysis-by-synthesis model. In this study participants were asked to identify three syllables (/pa/, /ta/, & /ka/) that were presented as visual-only (V-only; i.e., lipreading), auditory-only (A-only), congruent AV combinations and incongruent McGurk combinations. The analyses were restricted to the auditory N1-P2 complex and revealed AV induced amplitude reductions, for both incongruent and congruent AV trials, relative to the sum of A + V. Interestingly though, the ERP components peaked earlier following congruent AV speech trials. However, those latency shifts were dependent on the particular syllable. The three syllables differed in their ease of identification in the V-only condition. It is typical of AV speech that the visual speech cues, like lip movements, precede the first auditory signal by up to several hundreds of milliseconds (e.g.: van Wassenhove et al., 2005). When the authors looked at the relation between degree of predictability and N1-P2 latency shift, they noticed a positive association. That is, the more informative the visual speech cue was of the upcoming auditory speech sound the shorter the N1-P2 latency was. The degree of how informative visual speech cues were was determined by the percentage accurate identification in the V-only condition. The authors proposed that the reported AV interaction took place in auditory cortices, and suggested that the predicting visual speech cue (i.e., viseme) narrowed the range of possible frequency ranges to be processed by the auditory system (van Wassenhove et al., 2005). According to this analysis-by-synthesis model, the visual speech cue provides an internal representation which is used as a template to evaluate the incoming auditory speech sound. Perceptual processes in AV speech are therefore modulated by visual speech constraints. The electrophysiological indices (i.e., N1 latency shift) are seen as measures of differences between visual template and auditory signal. The more they match, the faster the processes (van Wassenhove et al., 2005). Related to the role of visual speech cues in constraining the auditory signal, Summerfield (1987) proposed that visemes could fulfill the role of frequency filters or tuners. This is particularly important in background noise where many sounds in addition to the speech signal enter the auditory system (Summerfield, 1983). If, however, visual speech cues can provide a constraint on the range of frequencies that are about to be produced by the speaker, the auditory system of the listener can use this information and 'tune in' to those frequencies and filter out unrelated ones. Support for this theory that speech cues could potentially provide frequency information was supplied by Grant and Seitz (2000b) who reported significant correlations between area of lip opening and second and third formants. Formants are high energy frequency bands characteristic for vowels. That is, the size of the mouth opening is related to particular frequencies of the auditory speech signal and therefore indicates which vowel is likely being uttered. To briefly sum up, visual speech cues could provide information regarding auditory frequencies which could help to extract relevant speech cues to improve speech perception (Grant & Seitz, 2000b; Summerfield, 1983, 1987).

Grant, Walden and Seitz (1998) developed a basic framework of AV speech perception that incorporates not only sensory related processes but also considers higher order factors such as linguistic abilities (e.g. semantic and syntactic knowledge) and effective use of context. In this model these latter factors exert top-down modulations on the AV integration processes which follow signal processing in the auditory and visual modalities. The model was originally developed for research on individuals with hearing impairment but cognitive factors are likely to play a role for normal hearing individuals as well making it an interesting model for AV speech perception in general. What makes this model so appealing is the fact that it allows the investigation of possible causes for individual differences in AV integration skills. These causes can be at sensory and/or at top-down levels. At higher order levels linguistic factors play an important role in speechreading (Boothroyd, 1988). It is assumed that an individual's lexicon or semantic knowledge is important for the process of decoding the (visual) speech signal (Grant et al., 1998). Using the available knowledge of the context provided by the sentence or a topic in general can further improve speechreading. The visual speech signal is ambiguous and speechreading of individual words is quite difficult (i.e., around 5-20% correct identification). However, information extraction from the visual speech signal improves dramatically (i.e., 40-50% correct) when constrained by sentence or topic context (Boothroyd, 1988; Grant & Seitz, 2000a; Rönnberg & Lyxell, 1998).

Grant and colleagues (1998) designed their model specifically for AV speech integration and given its focus on linguistic factors it does not readily transfer to multisensory processing outside the domain of speech. Consequently, the question that arises is whether multisensory integration mechanisms for AV speech perception can be generalized to AV non-speech domains. According to findings by Stekelenburg and Vroomen (2007) the (sensory) processes underlying AV speech are not special to speech but actually apply to non-speech processes as well. Their conclusions were based on the fact that the auditory N1 peaked earlier and was reduced in amplitude for AV speech stimuli (i.e., syllables) as well as AV non-speech stimuli, observing manual actions. Based on the fact that there was no stimulus specific task required, it is as arguable whether the underlying processes tapped into mechanisms designated for object recognition. Nevertheless, given their choice of non-speech stimuli, it is possible that similar processes were recruited in both instances; processes related to perception of motor actions. Even though it is unlikely that the observation of motor actions was directly related to the observed N1, there are some findings in the domain of AV speech that suggest that mechanisms relevant for motor actions might play an important role in AV speech perception.

An interesting finding in the primate literature was the discovery of so-called mirror-neurons in the primate brain area F5. It was shown that this area is activated when a particular action is performed but also when observing the same action being performed by someone else. The human area corresponding to monkey F5 has been said to be Broca's area and premotor cortex. It is not certain whether humans possess actual 'mirror-neurons' but Broca's area is suggested to house a 'mirror-system' that behaves similarly to response patterns of primate 'mirror-neurons' (Rizzolatti & Craighero, 2004). The implications for the existence of neurons with those response characteristics range from motor learning to speech perception (Arbib, 2005; Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004). The 'mirror-system' has also emerged in the context of

AV speech perception. Skipper, Nussbaum and Small (2005) used fMRI to measure brain activity while participants were either listening to a story (A-only), just watching the storyteller but not hearing him (V-only) or watching the storyteller and hearing him tell the story (AV). Relative to the unisensory condition, areas in posterior superior temporal sulcus as well as premotor cortex were more active during the AV condition. According to the authors, activity in premotor areas suggests that during AV speech perception visual cues provide information on how phonemes are produced. This in turn can be matched with the actual auditory speech sound perceived (Skipper et al., 2005). The premotor cortex was also active during the lipreading (V-only) condition but to a lesser extent, which is similar to a studies that report activity of Broca's area (Callan et al., 2003; Calvert & Campbell, 2003; Capek et al., 2004; Ojanen et al., 2005; Santi, Servos, Vatikiotis-Bateson, Kuratate, & Munhall, 2003). The finding that an area that is involved in speech production is also activated during AV speech perception is in agreement with a mirror-system. It also provides empirical support for the motor theory of speech perception that states that speech perception involves articulatory mechanisms required for speech production (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). That is, a perceived phoneme activates the same processes that are used when producing that phoneme.

The fact that Stekelenburg and Vroomen (2007) used action based, non-speech stimuli could imply that the neural processes they were observing were more related to action-observation mechanisms rather than to non-speech object recognition. Speech perception could be considered an object or more generally a stimulus recognition task with the words being the objects. Even though findings by Stekelenburg and Vroomen

(2007) are interesting, it is arguable whether their findings shed light onto the mechanisms of multisensory stimulus recognition especially because the chosen AV stimulus combinations in the non-speech condition identified an action and not an object. Therefore, the question of whether the mechanisms involved in stimulus identification of AV non-speech items are different than those for AV speech items is still a matter of debate. This question motivated the first study of this dissertation.

Chapter 2: Manuscript 1: AV speech vs. AV non-speech

Seeing a cow say "Moo" is different than watching a human say /kaʊ/:

Behavioural and electrophysiological differences between audiovisual speech and non-speech processing.

2.1 Abstract

The importance of speech perception for communication merits an investigation of whether humans process audiovisual (AV) non-speech (Experiment 1) differently than AV speech (Experiment 2). In Experiment 1 participants were asked to categorize animal pictures and animal vocalizations. Experiment 2 employed the same task and stimulus tokens in form of spoken animal names. Both experiments presented stimuli as auditory-only, visual-only and audiovisually (AV). AV trials yielded more accurate (Experiment 1) and faster responses (Experiment 2) than unisensory trials. ERP results indicated multisensory interactions reflected in reduced amplitudes of the visual N1 (Experiment 1) and the auditory N1 (Experiment 2) to AV stimuli compared to the summed unisensory responses, suggesting that AV modulations of sensory-specific processing depends on the modality dominant for the task. Results are also discussed in terms of multisensory efficiency.

2.2 Introduction

We are surrounded by a world rich in information kindling all of our senses. When visiting a concert hall to enjoy a Mozart symphony performed by an orchestra, we not only hear the music but also watch the musicians hit the timpani or bow the strings of their violins. The different sensory channels that are stimulated provide complementary pieces of information that are very different in their physical characteristics. Although the brain is able to effortlessly combine and integrate this multisensory information and form one coherent percept of the particular object to which we attend, the underlying mechanisms of this ability are not fully understood. The auditory and visual modalities have received the most attention in the multisensory research literature. In the context of audiovisual (AV) integration, the majority of research has been devoted to AV speech processing (i.e., hearing speech and reading lips simultaneously). AV speech, however, is only one instance where auditory and visual information is combined. Other non-speech stimuli in our environment, such as a barking dog or a bouncing ball, provide auditory and visual information as well; information that needs to be integrated by the observer's brain. Relative to AV speech, less is known about object identification of AV non-speech stimuli and it is interesting to know whether the underlying mechanisms involved in multisensory integration differ for the two classes of stimuli. To shed more light on this question, the current study investigated electrical brain responses associated with processing auditory and visual information from environmental, non-speech stimuli as well as AV speech stimuli. First, we will briefly review electrophysiological findings concerning AV non-speech processing, then studies looking into AV speech processes, and finally studies regarding differences between AV speech and AV non-speech processing.

2.2.1 Electrophysiology of AV Non-Speech.

In order to investigate neural processes underlying AV perception the current study recorded event-related brain potentials (ERPs). The two main ERP components of interest were the visual and auditory N1. Both are related to early visual and auditory object processing, respectively, and are said to be elicited in sensory cortices (Di Russo et al., 2001; Hopf et al., 2002; Näätänen & Picton, 1987). Both components are sensitive to change of a stimulus and hence 'detect' the onset and offset of a stimulus. Their amplitude and latency are influenced by stimulus parameters such as intensity, and frequency; however, attention can also modulate the neural response underlying the visual and auditory N1. According to Näätänen and Picton (1987), the auditory N1 consists of several subcomponents that differ in their latency and topographical distribution. One subcomponent is sensitive to stimulus properties and peaks maximally at frontocentral sites at around 100ms after sound onset. Its source is believed to be in the supratemporal plane of the auditory cortex and its amplitude increases with stimulus intensity.

The visual N1, which peaks around 150 ms after stimulus onset, reflects processes underlying visual stimulus discrimination, suggesting that this component is involved in basic stimulus identification and feature analysis (Mangun et al., 1993; Vogel & Luck, 2000). The precise neural source of the visual N1 has been difficult to determine but it is most prominent over occipital electrode sites and is likely to be part of the visual processing stream in inferior occipito-parietal regions (Di Russo et al., 2001; Hopf et al., 2002; Mangun et al., 1993).

To our knowledge only a few studies have used event-related brain potentials (ERPs) to look into AV object recognition. In one study (Giard & Peronnet, 1999), simple auditory sounds (e.g., 540 Hz tone), visual shapes (e.g., ellipse) and the simultaneous AV presentation of both were presented to participants who had to learn and remember these arbitrary sound and shape associations in order to perform a stimulus recognition task. The authors found an amplitude reduction of the visual N1 component and an enhancement of the auditory N1, indicating modulations of sensory specific ERP components during multisensory conditions relative to unisensory trials (i.e., auditory alone and visual alone).

Similarly, Molholm and colleagues (2002) found a reduced visual N1 and an enhanced auditory N1 in response to AV trials relative to unisensory trials. The stimuli consisted of a 1000 Hz tone and a red disc presented on a computer screen; participants were simply asked to respond as soon as they detected a stimulus. These results provide further support to the notion of multisensory interaction occurring in what have been traditionally thought of as exclusively unisensory brain areas. In a subsequent study, Molholm and colleagues (2004) used more ecologically valid stimuli, namely animal sounds (auditory only; A), animal drawings (visual only; V), and congruent as well as incongruent AV stimulus pairings, in contrast to the abstract and arbitrary auditory and visual stimulus pairs used in their previous research. Ecological validity could be an important issue because the multisensory mechanisms involved in combining existing representations that share natural associations and have been learned over time and through experience may differ from those recruited to combine stimuli that have been arbitrarily learned over a brief training period in order to meet task demands posed by laboratory tasks (De Gelder & Bertelson, 2003). In order to determine how attention affects brain responses during the multisensory perception of natural stimulus pairs, participants were asked to detect and respond to both the sound and picture of a target (e.g., one of eight animals) that was specified before the beginning of each block. Contrary to previous findings that showed AV-related amplitude reductions of the visual N1 (Giard & Peronnet, 1999; Molholm et al., 2002), Molholm and colleagues (2004) found that the simultaneous presentation of congruent AV information enhanced the amplitude of the visual N1 at 150 ms after stimulus onset, relative to the summed response to the unisensory target trials (A+V). No multisensory modulation was found for the auditory N1. Furthermore, the visual N1 amplitude was sensitive to the congruency of the AV information such that it was larger for congruent AV trials (e.g., a picture of a cow and the corresponding 'moo' sound) relative to incongruent AV trials. These early electrophysiological processing effects were accompanied by behavioural benefits in form of faster reaction times (RTs) to congruent AV trials relative to unisensory trials.

2.2.2 Electrophysiology of AV Speech Perception

The term AV speech refers to situations when the perceiver can both hear and see the other person speak. That is, the information coming in is multisensory, as the auditory speech signal enters the auditory system and the visual information (i.e., lip and articulator movements, gestures, facial expressions) enters the visual processing stream. Since the 1950s it has been known that speech perception improves when visual speech information is also available, particularly in noisy environments (Ross et al., 2007; Sumby & Pollack, 1954). Furthermore, the well-known McDonald-McGurk effect suggests that the visual and auditory modalities do not operate independently of one another but interact, both contributing to what is being perceived (McGurk & MacDonald, 1976). Despite the fact that there is longstanding evidence of the behavioural benefits associated with AV speech perception over and above just listening to someone speak, less is known about the underlying mechanisms that enable this seemingly effortless ability to integrate information from two modalities.

There are few studies that have compared the electrophysiological underpinnings of AV speech to those of unimodal speech processing (i.e., A-only and V-only). A study by Besle and colleagues (2004) required participants to detect a predefined target syllable. RT data revealed an AV facilitation effect over A-only and V-only performance alongside an AV-related reduction of the auditory N1 whose source was localized in auditory cortical areas. A magnetoencephalogram study on passive AV syllable perception also found AV reduction effects in primary auditory cortices (Heschl's gyrus Brodman area BA 41/42) as well as in the superior temporal sulcus, but interestingly, AV interaction effects occurred earlier in the former than the latter (Möttönen et al., 2004). Van Wassenhove and colleagues (2005) recorded ERPs during a syllable identification task. In line with previous studies, the authors observed an N1 amplitude reduction and a shortening of the peak latency in response to syllables presented audiovisually compared to unisensory responses.

Taken together, research on electrophysiological processes related to AV speech perception consistently found multisensory reductions of the auditory N1. On the other hand, findings for the modulation of ERPs by AV non-speech stimuli are not so consistent, with some studies finding a reduction in the visual N1 and an enhancement of the auditory N1 (Giard & Peronnet, 1999; Molholm et al., 2002) and others finding enhancement of the visual N1 (Molholm et al., 2004). The discrepancies of results are potentially due to differences in task design, nature of the stimuli, and/or attentional demands.

2.2.3 Audiovisual Speech versus Non-Speech

The importance of speech perception for human communication begs the question as to whether verbal face-to-face conversations have led to the development of specialised mechanisms devoted to AV speech processing. To our knowledge there is only one study that has used electrophysiological recordings to address this question directly (Stekelenburg & Vroomen, 2007). In that study, participants perceived spoken syllables (/bi/ and /fu/) and ecologically valid non-speech stimuli (e.g., clapping hands, a spoon tapping a cup). All were presented in unisensory (A-only, V-only) and AV formats. To ensure participants were attending to the video clips, they were required to detect dots that occurred on the screen on fewer than ten percent of the trials. Participants passively observed syllables being spoken and actions being performed but task instructions did not explicitly require conscious object identification; this absence of a behavioural task specific to the object or syllable stimuli precluded an assessment of multisensory interaction at the behavioural level. Nevertheless, the ERP results indicated multisensory interaction at the level of the auditory N1 in form of an amplitude reduction and a reduction in its latency. Interestingly, the ERP response pattern was the same for syllables (speech) and observed actions (non-speech). In a second experiment, the authors included the factor of congruency but found no difference between congruent and incongruent stimulus pairs until the P2 component, with congruent stimuli yielding larger P2 amplitudes than incongruent stimuli. Experiment Three presented a set of actions (e.g., sawing wood, tearing paper) with sound and movement starting at the same time. Unlike the previous two experiment, a clear auditory N1 modulation was absent during the AV trials relative to unisensory conditions. According to the authors, the auditory N1 amplitude reduction and latency shift in the first two experiments was due to the visual cue preceding and predicting the onset of the auditory signal (Stekelenburg & Vroomen, 2007). Based on these results, the authors concluded that the mechanisms for processing AV non-speech stimuli are similar to those involved when perceiving AV speech syllables.

However, due to the nature of the task, participants were not required to identify a given object. Rather, participants were passively observing someone perform a certain action. Therefore it can be argued whether participants in the AV non-speech condition were accessing the same processes as in an object recognition task. There is no question that the study by Stekelenburg and Vroomen (2007) required AV processing but it is not certain whether the observed processes reflected object identification. Speech comprehension, whether it is in an auditory-only or audiovisual mode, requires active object (i.e., spoken words) recognition and so in order to better understand the differences between perception of AV speech and AV non-speech stimuli choosing the proper object identification task is important.

The current study required participants to perform an identical object recognition task for both speech and non-speech stimuli. The primary objective of speech perception is to understand WHAT the partner is saying, making speech perception essentially an object recognition task with the spoken words being the objects to be identified. In most conversations the goal is to understand every single word being uttered and not just focus on one particular word, making speech perception more than a mere target detection task. Therefore, in designing the current study we opted for a categorization task that required identification of and a response to every stimulus. This stands in contrast to previous studies that required detection of just one predefined target at a time (Besle et al., 2004; Molholm et al., 2004) and to studies that did not require a task directed at the stimuli of interest (Möttönen et al., 2004; Stekelenburg & Vroomen, 2007). In this report, the AV non-speech (Experiment 1) and AV speech (Experiment 2) experiment employed stimuli that were semantically identical enabling us to use the same task for both experiments. The stimuli in Experiment 1 consisted of ecologically valid non-speech stimuli similar to those used by (Molholm et al., 2004), namely animal vocalisations and animal pictures. Experiment 2 presented the names of the same animals as in Experiment 1 as individually spoken words. Since the meaning of the stimuli was matched between both experiments, any differences between speech and non-speech processing could not be attributed to language-specific factors such as semantics nor to differences in the processing task. The ERP studies on AV speech mentioned above restricted their stimuli to a few single syllables. The current study investigated ERP processes of AV speech tokens that were ecologically valid, namely whole words. Also, we attempted to balance perceptual load between the two experiments. In AV speech perception, the auditory modality is more dominant than the visual modality (Easton & Basala, 1982). Visual speechreading is challenging because a given visual speech cue can map onto groups of acoustic speech sounds, making visual speech information more ambiguous than the auditory speech signal (Campbell, 2008). In order to match visual perceptual load for both experiments, we blurred the visual stimuli (i.e., animal pictures) of the AV non-speech study (Experiment 1) in an attempt to make them as perceptually difficult as the visual speech tokens in Experiment 2.

Given that participants were required to respond to each stimulus, accuracy and RT data were collected concurrently with ERP recordings, enabling the assessment of multisensory interaction at the neural and behavioural levels. We also attempted to match auditory attentional resource allocation during the AV speech and non-speech experiments. That is, because speech perception is an inherently auditory dominant task, task instructions for the AV non-speech trials also directed attention to the auditory stimulus (i.e., animal vocalisations).

The main question we tried to address with these two studies was whether AV speech perception is special. Support for the assumption that speech sounds are processed differently than non-verbal, environmental sounds comes from functional neuroimaging studies (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Lewis et al., 2004; Scott & Johnsrude, 2003) and aphasia patients (Clarke, Bellmann, de Ribaupierre, & Assal, 1996). Regarding electrophysiological processes of speech and non-speech sounds, there is evidence that auditory N1s evoked by speech are more distributed over the left hemisphere than those by non-speech (Pérez, Meyer, & Harrison, 2008). Even though recent research has been focused on whether speech perception is special, little is known about how speech and non-speech sounds interact with concurrent visual speech and non-speech cues, respectively. This study investigates the differences between AV interactions for speech and non-speech stimuli in an object recognition task.

Based on previous ERP studies investigating mechanisms of AV non-speech object perception, we predicted modulations of visual evoked potentials, particularly the visual N1. According to Giard and Peronnet (1999), multisensory perception reduces processing demands in the dominant modality. Based on this, we expected reduced processing demands for the visual system during AV non-speech object recognition, reflected in a visual N1 amplitude reduction. Consistent with previous findings on ERPs and AV speech processing, for Experiment 2 we hypothesized sensory specific multisensory interaction at the level of the auditory N1. Specifically, this modulation would be reflected in an amplitude reduction indicating a reduction in sensory processing demands during AV trials relative to unisensory trials. Moreover, this auditory as well as visual N1 amplitude reduction should be accompanied by behavioural benefits (i.e., greater accuracy and faster RTs).

2.3 Methods (Experiment 1: AV non-speech)

2.3.1 Participants

Fifteen young adults were tested but two were excluded due to poor behavioural performance, namely RTs that exceeded the group mean by more than two standard deviations. Thus, the final sample consisted of 13 individuals (7 men), between the ages of 18 and 33 (*M* = 25.1 years, *SD* = 3.8), recruited from a participation pool in the Department of Psychology, at Concordia University. All participants reported intact hearing and normal or corrected-to-normal vision. This study was approved by the Concordia University research ethics board.

2.3.2 Stimuli.

Twelve photographs of animals and their corresponding vocalisation sounds comprised the experimental stimuli. The twelve animals were divided into six large (i.e., elephant, cow, horse, lion, sheep, and wolf) and six small animals (rooster, cat, duck, cricket, bird, and frog), whereby small was defined as being small enough to fit underneath the chair the participant was sitting on. The visual stimuli were photographs taken from various online picture archives. Using Adobe Photoshop v. 6.0 we converted the pictures into grey scale images and degraded them by applying a Gaussian blur. The pictures of large and small animals did not differ in their basic visual properties as shown by an independent sample t-tests on mean pixel luminance (*t*(10)= 1.7, *p*= .12) and R.M.S. contrast (*t*(10)= -2.1, *p*= .7). The visual stimuli occupied a visual angle of 8.3 x 8.3° and were presented on a 16.1'' CRT monitor.

The auditory animal vocalisation samples were selected from various online sound effect libraries and edited using CoolEdit2000 software (Syntrillium Software Corporation, Seattle, WA, USA). Sound clips were cropped at the end to a duration of 600ms which did not alter the characteristics of the animal vocalisations. Stimuli (11025 Hz, 8 bit) were presented binaurally at 75dB SPL using EARLINK tube ear inserts (Neuroscan, El Paso, TX, USA). Using the software PRAAT (Boersma & Weenink, 2006) we verified that the frequencies associated with the highest dB level (i.e., the most auditory energy) were the same for vocalisation sounds of large (*m*= 1018 Hz, *SD*= 1239) and small (*m*= 1796 Hz, *SD*= 1368) animals (*t*(10)= 1.03; *p*= .33). We also compared the fundamental frequencies of large (*m*= 2140 Hz, *SD*= 1121) and small (*m*= 3064 Hz, *SD*= 1567) animal sounds to verify that they did not differ (*t*(10)= 1.17; *p*= .27).

2.3.3 Procedure

The experiment consisted of three blocks. The two unisensory blocks or conditions, A-only and V-only, consisted of 180 trials each in which each animal exemplar was presented 15 times in a random order. The multisensory AV block consisted of 360 randomised trials in which the sound and picture of each animal appeared 30 times. In the Av condition the image and the vocalization sound were presented simultaneously and half of the AV block trials were comprised of congruent stimulus pairs (i.e., the picture and sound belonged to the same animal; AV_{match} condition) and the other half were incongruent pairs ($AV_{mismatch}$ condition). In the AV_{match} condition 50% of the animals were small and 50% were large. In the $AV_{mismatch}$ condition, half of the trials included the sound of a large animal and the other half included the sound of a small animal. For the mismatching AV trials we counterbalanced the number of pairs composed of different animals from the same size category (e.g., a picture of a bird and a sound of a frog, both from the small category) and those composed of animals from the opposing size category (e.g., a picture of a bird and a sound of a cow). The sequence of the unisensory blocks was counterbalanced; however, the AV block always came last in order to prevent learning effects from the perceptually easier AV condition from confounding behavioural and electrophysiological responses to the perceptually more challenging unisensory stimuli. The stimuli were presented with Gentask software (NeuroScan, v. 2.4.18). Trials were separated by a stimulus onset asynchrony of 2.5s and each stimulus was presented for 600ms with auditory and visual stimuli starting simultaneously in the AV block. The response time window was set to 2s starting with stimulus onset.

Participants were asked to categorize the animals as either large or small by pressing the left or right button on a response box (Neuroscan, El Paso, TX, USA) with response assignments counterbalanced across participants. In the A-only condition, participants were instructed to base their response on the vocalisation sound, in the V-only condition responses were based on the visual stimulus, and in the AV condition, participants were asked to attend to both stimuli but base their response on the auditory cue. The selection of one modality was necessary in light of the $AV_{mismatch}$ trials. Participants were seated in a comfortable chair and informed consent was obtained before the testing session.

2.3.4 Data Acquisition

A continuous electroencephalogram (EEG) was recorded from 32 tin electrodes mounted in an elastic nylon cap (Electro-Cap International, Inc., Eaton, OH, USA) and arranged according to the International 10/20 system using a cephalic (forehead) location as ground and the left ear as online reference. All EEG data were re-referenced offline to linked ear lobes. The EEG was amplified using NeuroScan Synamps (Neuroscan, El Paso, TX, USA) and was recorded at a sampling rate of 500 Hz in a DC to 100 Hz bandwidth with electrical impedances kept below 5 kΩ. Horizontal and vertical electrooculograms (EOGs) monitored eye movements and trials with EOG activity exceeding +/- 75 μV were rejected. For a participant to be included in the analysis, a minimum of 90 accepted trials per condition had to be retained. The continuous EEG was divided into 900 ms epochs including a 100 ms pre-stimulus baseline interval and filtered offline for frequencies between 1-30 Hz.

2.3.5 ERP Components

The electrophysiological response to an auditory or visual stimulus typically consists of a series of early, sensory-driven and automatic ERPs. As seen in Figure 2, the first one is the P1, which constitutes the first peak with a positive-going amplitude at around 50 to 120 ms after stimulus onset. The P1 is followed by the first negative peak called N1 which reaches its maximum at around 90 to 170 ms after stimulus onset. The last component is the P2, which is the second positive peak and it is largest at around 200-250 ms after stimulus onset. This series of ERP components is also referred to as P1-N1-P2 complex (Eggermont & Ponton, 2002).

The amplitude of the visual and auditory N1 was calculated by computing the absolute peak-to-peak difference between P1 and N1. The amplitude of the visual and auditory P2 was calculated by computing the absolute peak-to-peak difference between N1 and P2. Component latencies were recorded at the components' peaks.

2.4 Results (Experiment 1)

Since all behavioural and ERP dependent variables involved repeated measures, all ANOVAs reported here were adjusted with the Greenhouse-Geisser non-sphericity correction (Greenhouse & Geisser, 1959) and, according to convention, we report uncorrected degrees of freedom, the Greenhouse-Geisser epsilon (ε), mean square error (*MSE*) and adjusted *p*-value. Significant main effects and interactions were followed by analyses of simple effects and, unless stated otherwise, the differences reported below are significant at $\alpha = .05$.

2.4.1 Behavioural Results

Two separate one-way analyses of variance (ANOVAs) were conducted on the mean RTs and percent correct responses (accuracy) with the factor Condition consisting of four levels (A-only, V-only, AV_{match} and $AV_{mismatch}$) using statistical software SPSS v. 11.5. The one-way ANOVA on accuracy revealed a significant effect ($F(3,36)= 7.4$; $MSE= 41.75$; $p= .007$; $\varepsilon= .52$), indicating that performance in the A-only condition was significantly more accurate than in the V-only condition. The AV_{match} condition yielded the highest percentage of correct responses relative to the two unisensory conditions. Accuracy in the $AV_{mismatch}$ condition was significantly lower than performance in the AV_{match} and A-only conditions (Figure 1a).

Results of the ANOVA on RT showed an effect of Condition ($F(3,36)= 20.6$; $MSE= 9615.4$; $p< .001$; $\varepsilon= .68$). Responses to matching auditory and visual stimuli (AV_{match}) were faster than those in A-only and $AV_{mismatch}$ trials. However, the V-only trials yielded the fastest RTs relative to all other conditions (Figure 1b).

2.4.2 Electrophysiological Results

As shown in Figure 2, the auditory condition elicited a negative-going peak between 100-130 ms (the auditory N1) followed by a positive-going peak at approximately 200 ms (P2); both were most prominent at the fronto-central electrodes. These same components were elicited in the AV conditions and were, in fact, enhanced in amplitude.

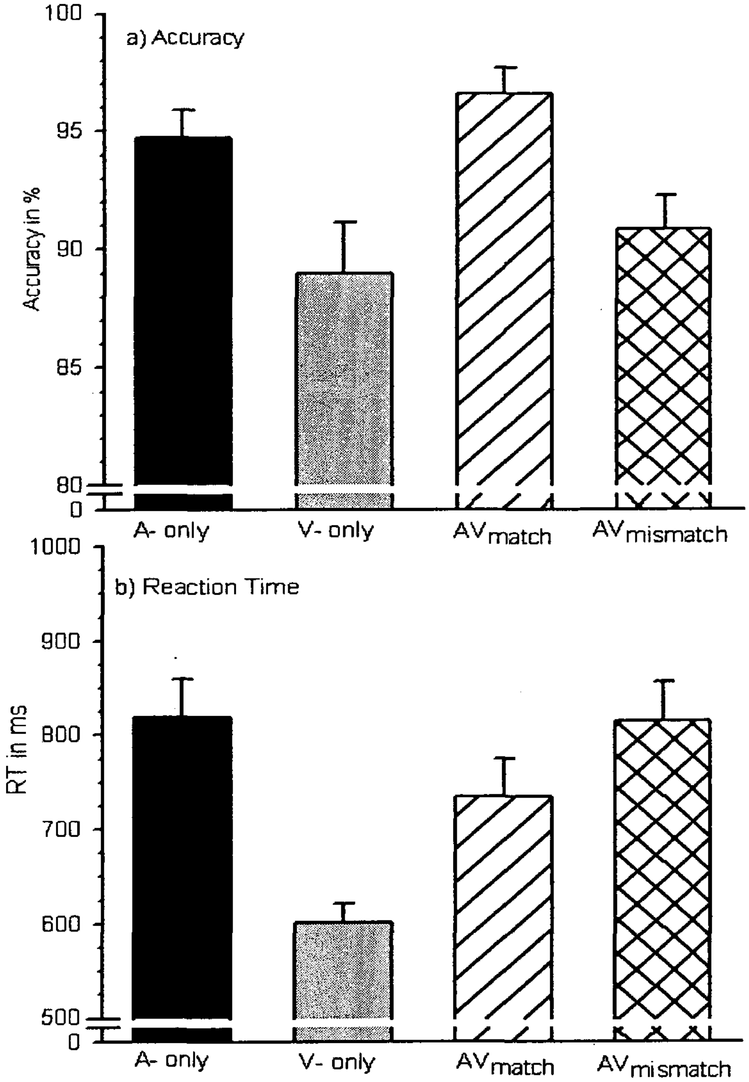

Figure 1: Behavioural results of Experiment 1: a) Mean percent correct responses and standard error bars; b) Mean reaction time and standard error bars.

Note that, although there is a negative deflection in the V-only condition at the same time as the auditory N1, this frontal negativity is not the same as the auditory N1 nor is it the visual N1 (visible at occipital sites; Figure 3). Instead, it appears to be the visual P1 which shows a polarity inversion at more anterior sites (Figures 2 & 3).

The visual condition was characterized by a clear positivity at approximately 100 ms (P1) followed by the visual N1 (at approximately 145 ms) and the visual P2 (at approximately 200 ms). These components were most prominent at occipital leads (O1 and O2), somewhat less so at Pz, and were virtually absent at more anterior electrode locations. The visual N1 was present in the AV conditions but was smaller than in the unimodal V-only condition (Figure 3).

To assess multisensory modulations of the ERP components of interest, we compared the electrophysiological response to the multisensory stimulus (AV) to that of the summed response to the unisensory trials. To obtain the latter (A+V), we added the waveforms of the unisensory conditions of each individual participant and extracted peak score values (i.e., latency and amplitude) of the ERPs of interest. Multisensory interaction can manifest itself as 1) multisensory signal enhancement, or superadditivity, which is present when the electrophysiological response to the multisensory stimulus exceeds that of the summed response to the unisensory trials (i.e., AV> A+V); or 2) as multisensory response reduction or subadditivity, which refers to an amplitude reduction for multisensory stimuli (i.e., AV< A+V) (Calvert et al., 2001).

ERP analyses. Given the differences in timing and topography of the visual and the auditory N1 components, data of the two unisensory conditions A-only and V-only were never included in any of the ERP analyses together. Initial analyses assessed hemispheric differences in the visual N1 and P2 and the auditory N1 and P2.

An initial ANOVA with factors Hemisphere (left & right), Condition (A-only, A+V, and AV), and Anteriority (six sites from frontal to centro-parietal regions (left: F3, FC3, C3, CP3; P3, O1; right: F4, FC4, C4, CP4, P4, O2) was conducted, followed by an ANOVA with factors Condition and Site (Fz, FCz, Cz, CPz, Pz, O1, O2). Because the lateral sites did not yield any additional information (all *F* values < 2.4 and all *p*-values > .11), only electrophysiological results from midline and occipital sites are reported below. For each of the ERPs of interest potentials (i.e., N1, P2) a separate ANOVA was conducted for peak latency and peak amplitude with factors Condition and Site. Analyses of auditory evoked potentials were restricted to the electrode sites that showed the clearest auditory components, namely Fz, FCz, Cz, and CPz. Analyses of visual evoked potentials were restricted to the electrode sites that showed the clearest visual components, namely Pz, O1 and O2. For the sake of brevity, the effect of site is only reported below when the main effect or interaction is significant.

Auditory N1 responses. The two multisensory conditions AV_{match} and $AV_{mismatch}$ elicited larger auditory N1 responses than the unisensory A- and V-only conditions at midline sites (Figure 2). To assess multisensory interaction, we compared the auditory N1 amplitude of the two AV conditions to that of A-only and to that of the sum of A- and V-only conditions (A+V).

The auditory N1 amplitude was analyzed with a 4 X 4 repeated measures ANOVA with the factors Condition (A-only, AV_{match}, $AV_{mismatch}$, A+V) and Site (Fz, FCz, Cz, CPz). There was a main effect of Site (*F*(3,36)= 10.1; *MSE*= 9.02; *p*= .002; ε=

.54) with N1 amplitudes being largest at fronto-central electrode sites. There was a main effect of Condition (*F*(3,36)= 8.3; *MSE*= 20.7; *p*= .002; ε= .63) such that A-only trials elicited the smallest N1 amplitudes and the other three conditions did not differ from each other. Thus, the N1 amplitude to the multisensory conditions (AV) did not meet the criterion of superadditivity (i.e., AV > A+V); instead, it was additive in that AV was equal to A+V.

Another approach to assess multisensory interaction is to calculate the difference between responses to AV_{match} and A+V and to conduct paired t-tests at each time point for the first 300ms after stimulus onset (i.e., every two ms at a 500 Hz sampling rate). A minimum of 12 consecutive t-tests had to exceed the critical t-value of 2.14 in order to be significant (Guthrie & Buchwald, 1991), but no significant difference in amplitude between AV_{match} and A+V was found at midline electrodes.

Analysis of the auditory N1 peak latency did not reveal a main effect of Condition (*F*(3,36)= .42; *MSE*= 399.2; *p*= .67; ε= .7) or Site.

Visual N1 responses._A robust visual N1 was evident in the V-only and AV conditions but absent from the A-only condition. Figure 3 clearly shows that the visual N1 of the AV conditions was reduced in amplitude relative to the V-only condition. Given that visual evoked potentials at Pz were relatively small compared to those elicited at O1 and O2, Figure 3 displays ERP waveforms for occipital sites only. To test for multisensory interaction, we compared the amplitude of the visual N1 of the two AV conditions to that of the V-only trials and to that of the summed response of the unisensory conditions (A+V) at the electrodes Pz, O1, and O2 in a 4 (Condition) X 3 (Site) repeated measure ANOVA.

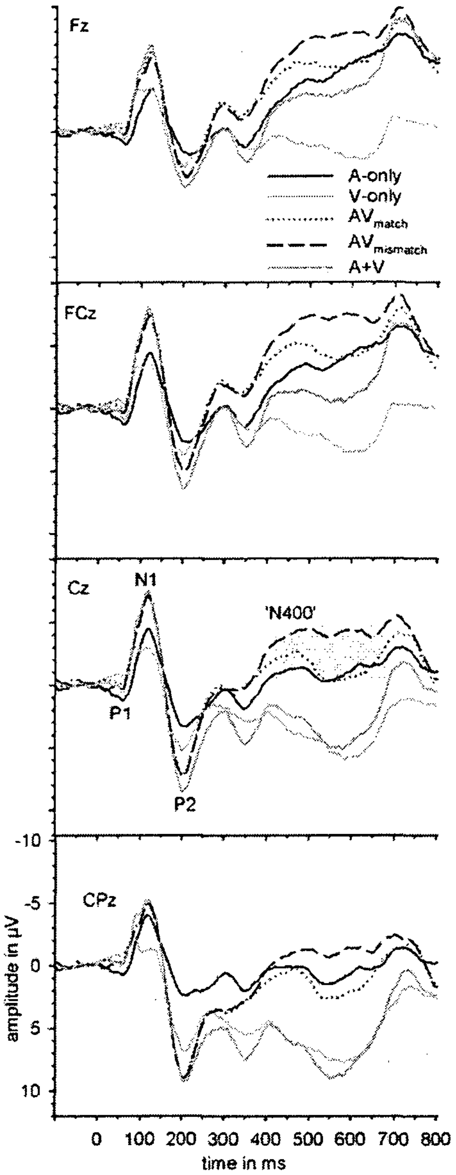

Figure 2: ERP waveforms of 5 conditions for Experiment 1 showing auditory evoked potentials and an N400-like component at midline sites Fz, FCz, Cz, and CPz. Solid black line = Auditory-only, solid light grey line = Visual-only, solid dark grey line = summed unisensory response (A+V), dashed black line = $AV_{mismatch}$, dotted black line = AV_{match}. Note the enhancement of the auditory N1 in the AV conditions versus the unisensory conditions.

This yielded a main effect of Condition (*F*(3,36)= 11.3; *MSE*= 10.3; *p*= .002; *ε*= .44) and a Condition by Site interaction (*F*(6,72)= 4.7; *MSE*= 5.8; *p*= .02; *ε*= .35). The latter indicated a right hemispheric dominance for the visual N1 with amplitudes at O2 significantly larger than at O1 and Pz for all three conditions. AV_{match} and $AV_{mismatch}$ conditions did not differ from each other at any of the three sites but were significantly smaller than the V-only and A+V waveforms at the two occipital sites. Consistent with the analysis of the visual N1 peak amplitude, assessment of the AV_{match} minus (A+V) difference waveform using consecutive t-tests revealed significant differences from 144 to 220 ms after stimulus onset at electrode sites O1 and from 140 to 300 ms at O2.

Even though participants were instructed to attend to both signals in the AV condition, they were asked to respond based on the auditory signal. To confirm that the observed visual N1 reduction was not due to less visual attention during AV trials, we also analyzed the visual P1 peak amplitude which has been shown to be modulated by attention the same way as the subsequent visual N1 (Di Russo et al., 2003; Di Russo et al., 2001; Hillyard & Anllo-Vento, 1998). The 4 (Condition) X 3 (Site) repeated measure ANOVA revealed a main effect of Site (*F*(2,24)= 35.1; *MSE*= 34.02; *p*< .001; *ε*= .63) with visual P1 amplitudes being smallest at Pz but no main effect of Condition (Figure 3), indicating that the visual P1 was not reduced in the AV conditions.

Analysis of visual N1 peak latencies did not yield a main effect of condition (*F*(3,36)= 3.1; *MSE*= 184.9; *p*= .063; *ε*= .65) but a main effect of Site (*F*(2,24)= 3.1; *MSE*= 726.9; *p*= .001; *ε*= .59) with the visual N1 latency peaking earlier at Pz relative to O1 and O2.

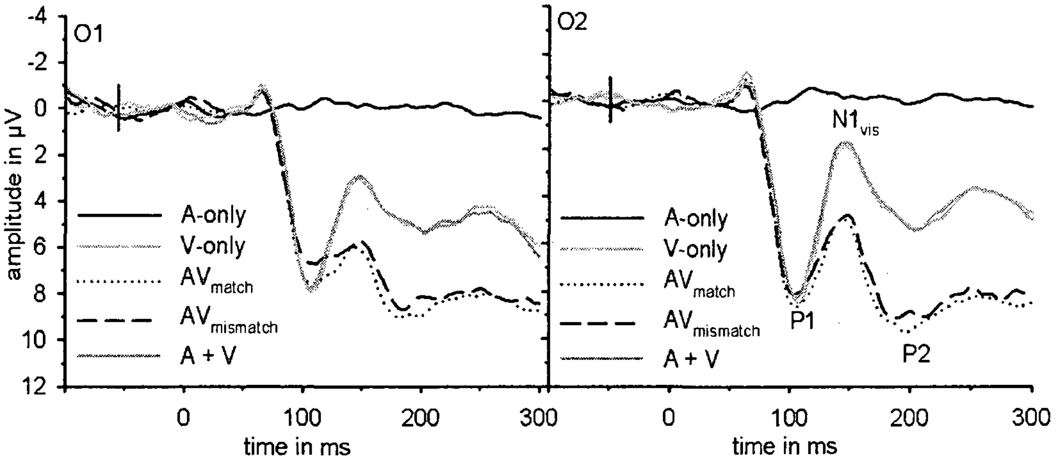

Figure 3: ERP waveforms of 5 conditions for Experiment 1 showing visual evoked potentials at occipital sites O1 and O2. Solid black line = Auditory-only, solid light grey line = Visual-only, solid dark grey line = summed unisensory response (A+V), dashed black line = AV$_{mismatch}$, dotted black line = AV$_{match}$. Note the reduction in the visual N1 in the AV conditions relative to the Visual-only condition at O1 and O2.

Auditory & visual P2. At around 200 ms after stimulus onset, both auditory and visual unimodal conditions elicited a P2 component at all midline sites (which was also visible at occipital sites in the V-only condition). This P2 appeared to be larger for the AV conditions relative to the unisensory conditions (Figure 2). Therefore, to assess multisensory interaction, the two AV conditions (i.e., AV_{match} & $AV_{mismatch}$) were compared against the summed response of the A-and V-only conditions at the five midline and the two occipital sites. A 3 (Condition) X 7 (Site) ANOVA revealed a main effect of Site ($F(6,72)= 37.0$; $MSE= 63.9$; $p< .001$; $\varepsilon= .35$) indicating that the P2 amplitude was largest at Cz in all three conditions. There was also a Condition by Site interaction effect ($F(12,144)= 4.2$; $MSE= 6.3$; $p= .02$; $\varepsilon= .21$). Subsequent tests of simple effects showed that the two AV conditions did not differ from each other, but $AV_{mismatch}$ was smaller than A+V at Fz and FCz. An additional investigation of P2 peak latencies did not reveal an effect of Condition ($F(2,24)= 1.5$; $MSE= 206.6$; $p= .24$; $\varepsilon= .67$) or Site.

Late ERP effects. Visual inspection of the ERP responses revealed that, over the 400 to 800 ms time window, $AV_{mismatch}$ waveform became more negative in amplitude than the AV_{match} condition (Fig. 2). Given the time range and its sensitivity to the mismatching condition, we considered this similar to an N400 effect (Kutas & Hillyard, 1980; Kutas, Van Petten, & Kluender, 2006). To analyse this 'N400'-effect, we calculated the mean waveform amplitude in four 100 ms windows (i.e., 400-500 ms, 500-600 ms . . .) and submitted these to a 4 (Time Interval) X 2 (Congruency; AV_{match} and $AV_{mismatch}$) X 7 (Site; 5 midlines and O1 & O2) repeated measures ANOVA. These revealed a significant main effect of Congruency ($F(1,12)= 8.1$; $MSE= 33.0$; $p= .015$) and Site ($F(6,72)= 39.3$; $MSE= 296.3$; $p< .001$; $\varepsilon= .25$). In addition there was a significant

Congruency by Time interaction (*F*(3,36)= 17.8; *MSE*= 4.9; *p*< .001; *ε*= .53) and Congruency by Site interaction (*F*(6,72)= 7.7; *MSE*= 11.1; *p*= .006; *ε*= .26). The latter two interactions indicated that the $AV_{mismatch}$ amplitude was more negative than AV_{match} at sites Fz, FCz, Cz, and CPz between 400 and 700 ms and at Pz between 500 and 700 ms.

2.5 Discussion (Experiment 1)

Results of Experiment 1 revealed multisensory interaction effects both at the behavioural and neural levels. With respect to the behavioural data, congruent auditory and visual cues (i.e., AV_{match}) led to more correct responses than the unisensory stimuli while incongruent trials ($AV_{mismatch}$) led to multisensory interference as the number of accurate responses dropped below that in the A-only and AV_{match} conditions. Similarly, adding a congruent picture to a non-impoverished animal sound facilitated RTs relative to the unisensory A-only trials. Although the V-only trials unexpectedly yielded the fastest RTs relative to all other conditions, we interpret this as a speed-accuracy trade-off given that responses to visual stimuli were relatively fast but inaccurate.

To assess multisensory interaction in terms of the underlying electrophysiological processes we analysed auditory and visual evoked potentials separately, due to their different topographical distributions and different neural generators. For the auditory N1, responses to AV_{match} and $AV_{mismatch}$ trials did not differ from the arithmetic sum of the unisensory A-only and V-only responses. That is, no multisensory interaction effects were found for the auditory N1. Given that the auditory N1 has been localized in the auditory cortex (Eggermont & Ponton, 2002; Giard & Peronnet, 1999; Näätänen &

Picton, 1987), these results suggest that the AV non-speech stimuli were not integrated at the sensory specific level of the auditory cortex. Analyses of the visual N1 amplitude pointed to multisensory interaction, as responses to AV trials (AV_{match} and $AV_{mismatch}$) were reduced in amplitude relative to V-only and A+V. Subsequent analyses of the difference wave between AV_{match} and A+V confirmed the presence of multisensory interaction for early visual processes.

The key finding from this study was that the visual N1 amplitude was reduced during AV trials relative to V-only and A+V. This could indicate multisensory interaction in form of response reduction at the level of visual processing areas. No such modulation was evident for the auditory N1; thus, AV non-speech stimuli modulated visual but not auditory processing.

This is consistent with findings by Giard and Perronet (1999) who also reported a reduction of the sensory driven posterior N1 during AV trials and who interpreted this attenuation as an index of a reduced requirement for visual processing effort during multisensory perception of non-speech stimuli. Furthermore, the fact that fewer neural resources were recruited during AV relative to V-only perception while accuracy increased indicates that multisensory processing is more efficient than unisensory perception.

Even though previous research has also found that AV non-speech stimuli led to a reduction of visual N1 amplitude (Giard & Peronnet, 1999; Molholm et al., 2002), an alternative explanation to the N1 reduction observed in the current study being due to multisensory interaction could be differences in attention between AV blocks and V-only blocks. For V-only trials, attention was focused on the visual stimuli, but in the AV condition task instructions required participants to respond to the sound and not the visual stimulus. This might have biased participants to focus more on the auditory stimuli and less on the images. Previous ERP work has shown that auditory and visual N1 amplitudes decrease when stimuli are not attended to (Hillyard & Anllo-Vento, 1998; Näätänen & Picton, 1987). However, there are two aspects of our data that mitigate against this argument and show that participants did not ignore the visual stimuli, namely the behavioural results and the P1/P2 ERP results.

Based on the behavioural data, the participants clearly attended to both visual and auditory sources of information, because AV_{match} responses were faster and more accurate compared to the A-only condition, indicating that participants were processing the visual information as well as the auditory information. If they had ignored the additional visual information, RTs and accuracy for A-only and AV_{match} trials should not have differed. Furthermore, $AV_{mismatch}$ trials produced behavioural interference effects due to the mismatching auditory and visual information in the stimulus pairs, again indicating that both modalities were attended. With respect to the visual P2, we did not find amplitude differences between the V-only and AV conditions. The visual P2 has been shown to be involved in visual object processing and, like the N1, is modulated by visual attention (Luck & Hillyard, 1994). Similarly, there is evidence that lack of visual attention leads the same reduction in P1 as in N1 (Di Russo et al., 2003; Di Russo et al., 2001; Hillyard & Anllo-Vento, 1998). If participants in our study were not attending to the visual information, then the P1 and P2 component should have shown similar attenuation effects as seen in the N1 response, but they did not.

In terms of semantic congruency, we found that neither visual nor auditory N1 responses to AV_{match} and $AV_{mismatch}$ stimuli differed from each other, suggesting that semantic congruency did not affect early sensory processing. However, electrophysiological responses to AV_{match} and $AV_{mismatch}$ conditions differed from each other for later time intervals starting at around 400ms after stimulus onset. The finding that incongruent stimulus pairs, relative to congruent pairs, elicited a significantly larger negativity compared to congruent stimuli at central and centro-parietal sites could be interpreted as an indication of difficulties integrating incongruent semantic information. This is consistent with a robust ERP component used in language processing called the N400, which is elicited by semantically incongruent sentences (Connolly & Phillips, 1994; Kutas & Hillyard, 1980; Kutas et al., 2006) and word-pairs (Anderson & Holcomb, 1995; Kutas & Van Petten, 1994). This late AV incongruency effect is in agreement with findings by Molholm and colleagues (2004). Other studies reported similar negative deflections to real world objects that were incongruent with the context they were presented in (McPherson & Holcomb, 1999; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Sitnikova, Kuperberg, & Holcomb, 2003).

To summarize, Experiment 1 revealed behavioural benefits associated with AV_{match} trials as well as a multisensory reduction of the visual N1 amplitude. To answer the question whether multisensory processing is different for AV non-speech versus AV speech stimuli a second experiment was conducted. To compare results from the AV non-speech study (Experiment 1) to those of the AV speech study (Experiment 2), the latter used the same animal tokens and same task as in Experiment 1. The crucial difference is that in Experiment 2 auditory and visual stimuli were spoken animal names rather than the animals' vocalisations and their pictures. Thus, even though the stimuli were physically different, the semantic concept that each stimulus referred to was identical.

2.6 Methods (Experiment 2: AV speech)

2.6.1 Participants

Fourteen young adults (10 female) between the ages of 18 and 34 (*M* = 21 years, *SD* = 4.2), recruited from a participation pool in the Department of Psychology at Concordia University participated in the study. None of them participated in Experiment 1. All participants were right-handed, had English as their first language, reported good health, intact hearing, and normal or corrected-to-normal vision. Participants gave informed consent and this study was approved by the Concordia University research ethics board.

2.6.2 Stimuli

The same 12 animal tokens as for Experiment 1 were used, divided into six large (elephant, cow, horse, lion, sheep, and wolf) and six small animals (rooster, cat, duck, cricket, bird, and frog). We videotaped a female speaker uttering the animal names and subsequently edited the videos using Adobe Premiere to reveal only the head, face, and neck of the speaker. Each video consisted of the utterance of a single word. Unlike Experiment 1 where we blurred the visual stimuli, we did not alter the visual signal for Experiment 2 as lipreading is very difficult to begin (recall that we blurred the visual stimuli in Experiment 1 to balance the degree of difficulty for processing visual information in both experiments). During the experiment each video subtended a visual angle of 8.3° x 8.3° and was presented on a 16.1'' CRT monitor. The sound files were digitized at 48 kHz and matched on sound intensity using Adobe Audition and PRAAT (Boersma & Weenink, 2006). The average duration of each spoken word was 556 ms (*SD*= 107 ms; range: 407 to 763 ms). Auditory stimuli were presented binaurally at 65dB SPL using EARLINK tube ear inserts (Neuroscan, El Paso, TX, USA).

As before, Experiment 2 included three different presentation conditions, namely auditory-only (A-only), visual-only (V-only), and AV. In the AV condition participants watched and heard a video-clip of the woman speaking. The AV stimuli served as basis to create the stimuli for the other two conditions. The presentation of the V-only condition included the same stimuli as the AV trials but with the audio track turned off. Likewise, the A-only trials were the same as AV trials but without the video track. In order to record visual and auditory ERPs, we marked the onset of the lip movement and the onset of the sound, respectively, with transistor-transistor logic (TTL) triggers. Given that the AV stimuli served as the basis for all conditions both triggers were present in all three conditions even if a given modality was not perceptible. That is, the V-only condition included a trigger to mark the onset of the speech sound even though the sound was not audible to the participant. This was particularly important because it allowed us to compute the sum of responses to A-only and V-only (A+V) aligned to the same point in time. This careful alignment of time points allowed us to accurately assess any non-linear interaction effects present in the AV trials.

2.6.3 Procedure

The experimental procedures of Experiment 2 were identical to those of Experiment 1 with the exception that the AV condition contained no incongruent AV speech stimuli (i.e., all AV trials were congruent). The congruency factor was dropped because there was no evidence from Experiment 1 that it had any impact on early ERPs. The A, V, and AV conditions were presented in blocks, each consisting of 204 trials with each animal name presented 17 times in a random order. As in Experiment 1, the sequence of unisensory blocks was counterbalanced across participants whereas the AV block always came last. For all three conditions participants were instructed to categorize the presented animal names as either large or small by pressing one of two keys on a standard keyboard (i.e., *'S'* and *'L'* keys) with response assignments counterbalanced across participants.

At the beginning of each trial a fixation dot was presented in the centre of the monitor for 450ms (Figure 4). Trials involving visual information (i.e., V-only and AV) replaced the fixation dot with a sequence of 18 still frames (600ms) of the speaker's face as a lead-in to avoid an abrupt onset. Following this time period of still frames, the speaker's lips started to move. In the AV condition the lip movement preceded the first auditory speech cue on average by 216 ms (*SD*= 140 ms, range 0 to 431 ms). In the V-only trials, no auditory speech was presented and only the image of the person speaking was visible. After the speaker had finished saying the word a series of still frames were added (2.7 s) followed by a 450ms inter-stimulus interval to give participants a sufficiently long response time window. The stimulus onset asynchrony between the onsets of the first video frame was 4.5 seconds for each trial. The software program Inquisit 2.0 (2006) was used for stimulus presentation.

2.6.4 Data acquisition

EEG recording parameters were the same as for Experiment 1 with the exception of the epoch length.

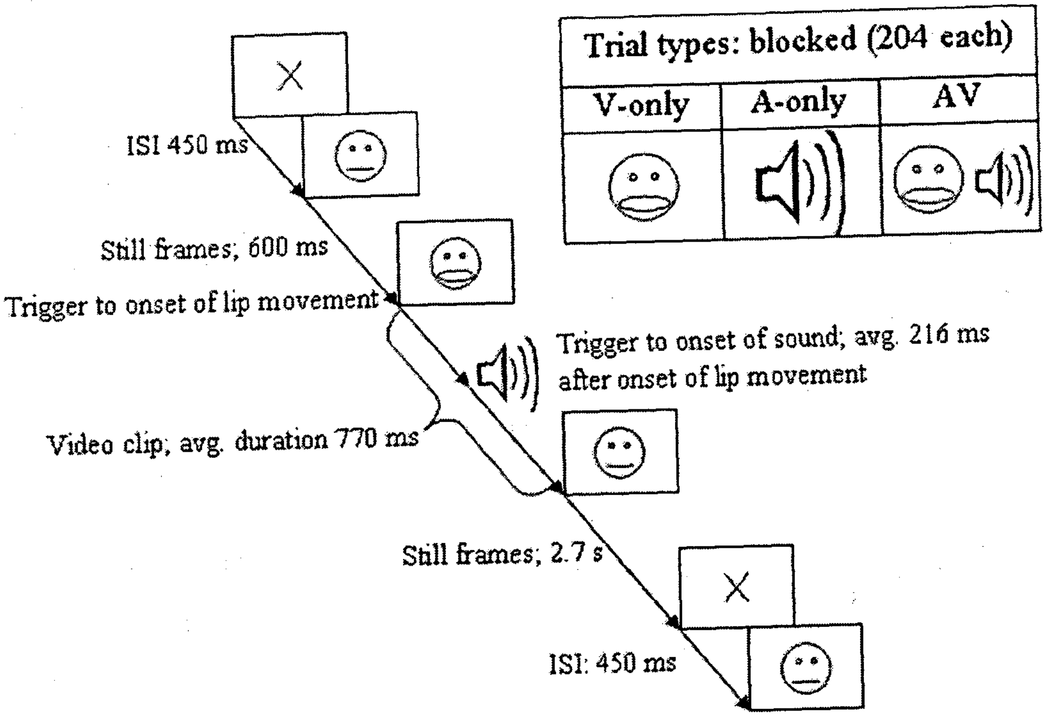
Trial types: blocked (204 each)
V-only
A-only
AV
ISI 450 ms
Still frames; 600 ms
Trigger to onset of lip movement
Trigger to onset of sound; avg. 216 ms after onset of lip movement
Video clip; avg. duration 770 ms
Still frames; 2.7 s
ISI: 450 ms

Figure 4: Schematic representation of a trial sequence. ISI= Inter Stimulus Interval.

The continuous EEG was divided into 700 ms epochs with a 100 ms pre-stimulus baseline interval.As was the case for Experiment 1, the amplitudes of the N1 and P2 ERP deflections were calculated by computing the absolute peak-to-peak difference between P1 and N1 and between N1 and P2, respectively. Component latencies were recorded at the components' peaks relative to the 0 ms stimulus onset.

2.7 Results (Experiment 2)

2.7.1 Behavioural Analyses

A separate one-way ANOVA with the factor Condition (3 levels: A-only, V-only, and AV) was conducted for the mean RT and percent accuracy data. The analysis of accuracy results revealed a main effect of Condition ($F(2,26)= 84.4$; $MSE= 57.6$; $p< .001$; $\varepsilon= .53$). As displayed in Figure 5a, performance was worse in the V-only condition relative to the other two conditions which did not differ from one another. The ANOVA on the RT data also yielded a main effect of condition ($F(2,26)= 164.8$; $MSE= 6095.3$, $p< .001$; $\varepsilon= .92$). Figure 5b shows that responses during AV trials were faster than those to A-only which were faster in turn than V-only trials.

Another way to determine multisensory interaction is to analyze RTs with respect to the race model (Miller, 1982). To some extent the AV speech signal is redundant as the information that is provided by the visual signal (i.e., lip movement) overlaps largely - but not entirely - with the spoken auditory input (Campbell, 2008). According to the independent race model, two sensory information channels are independent from one another and it is the faster of the two channels that will be successful in eliciting a response (Miller, 1982).

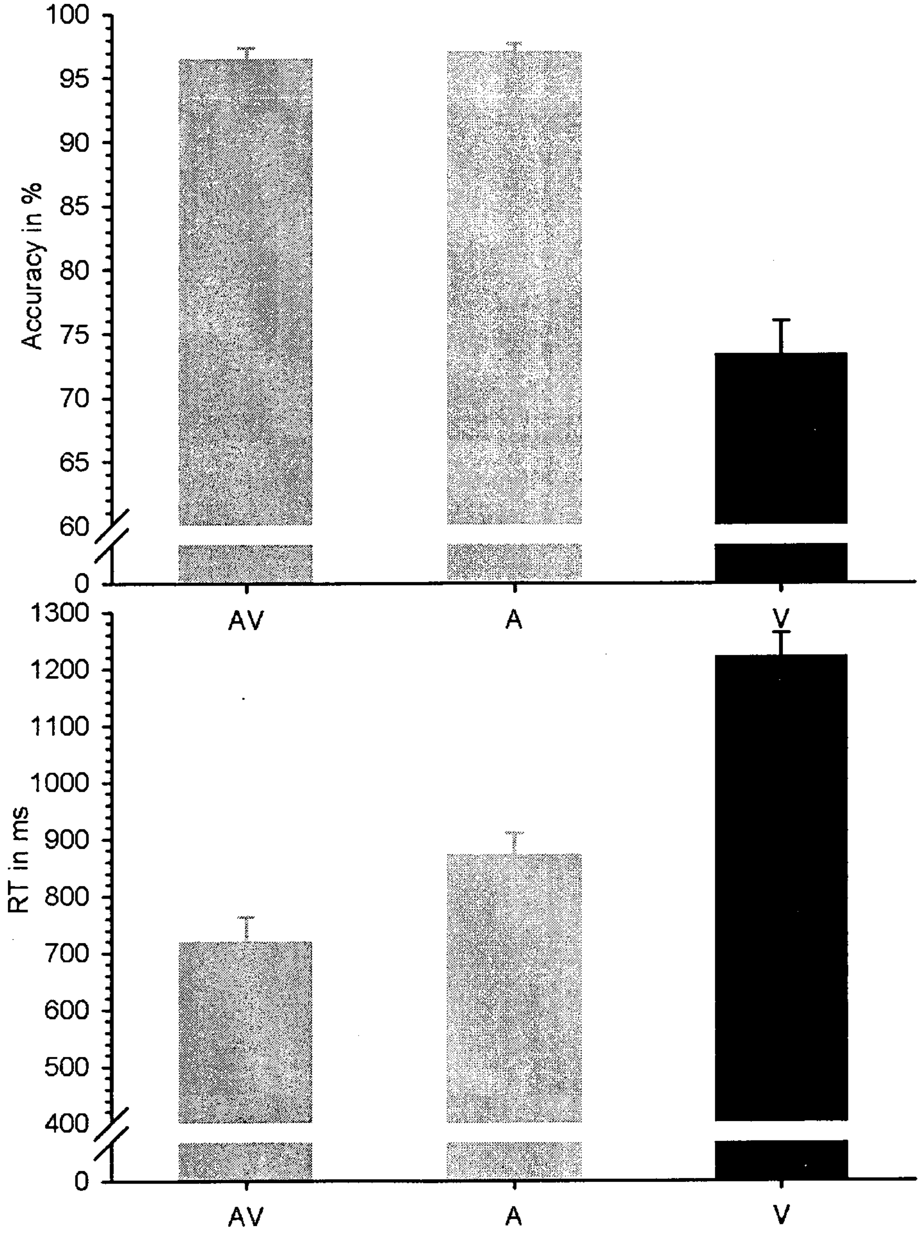

Figure 5: a) Mean accuracy scores in % with standard error bars. b) Mean reaction time in ms with standard error bars.

In order to assess the validity of the race model, individual trial RT data from each condition were transformed into cumulative distribution functions (CDFs) and the multisensory condition (i.e., AV) was compared to the joint probability of the unisensory responses ((A+V) - (AxV)). The independent race model is said to be violated when, for a given RT, the probability of the AV condition exceeds that of what is predicted by the combined probability of the unisensory responses (i.e., p(AV) > (p(A+V) - p(AxV)) (Miller, 1982). (Note that a race model analysis was not conducted for Experiment 1 because the mean RT of the AVmatch condition was not faster than RTs of the unisensory V-only condition, preventing a violation of race model predictions).

A violation of the race model supports the co-activation model which states that two information channels interact allowing for the possibility of neural integration (Laurienti, Burdette, Maldjian, & Wallace, 2006; Miller, 1982). To perform this analysis, the response time window from 300 to 1600 ms after stimulus onset was divided into 10 ms bins and, at each time bin, the cumulative probability of a response occurring at that time point or faster was computed. CDFs were calculated for each condition and each individual. Figure 6 displays the group averaged CDFs for each condition. Multisensory CDFs from the AV condition were compared to the combined CDFs from the unisensory conditions (i.e., (A+V) - (AxV)) with a two-tailed t-test conducted at each time bin. The analysis revealed that the probability of the AV response exceeded that of the combined unisensory probability in the 320 to 960 ms response window. That is, the prediction of the race-model that two sensory channels operate independently of one another was violated.

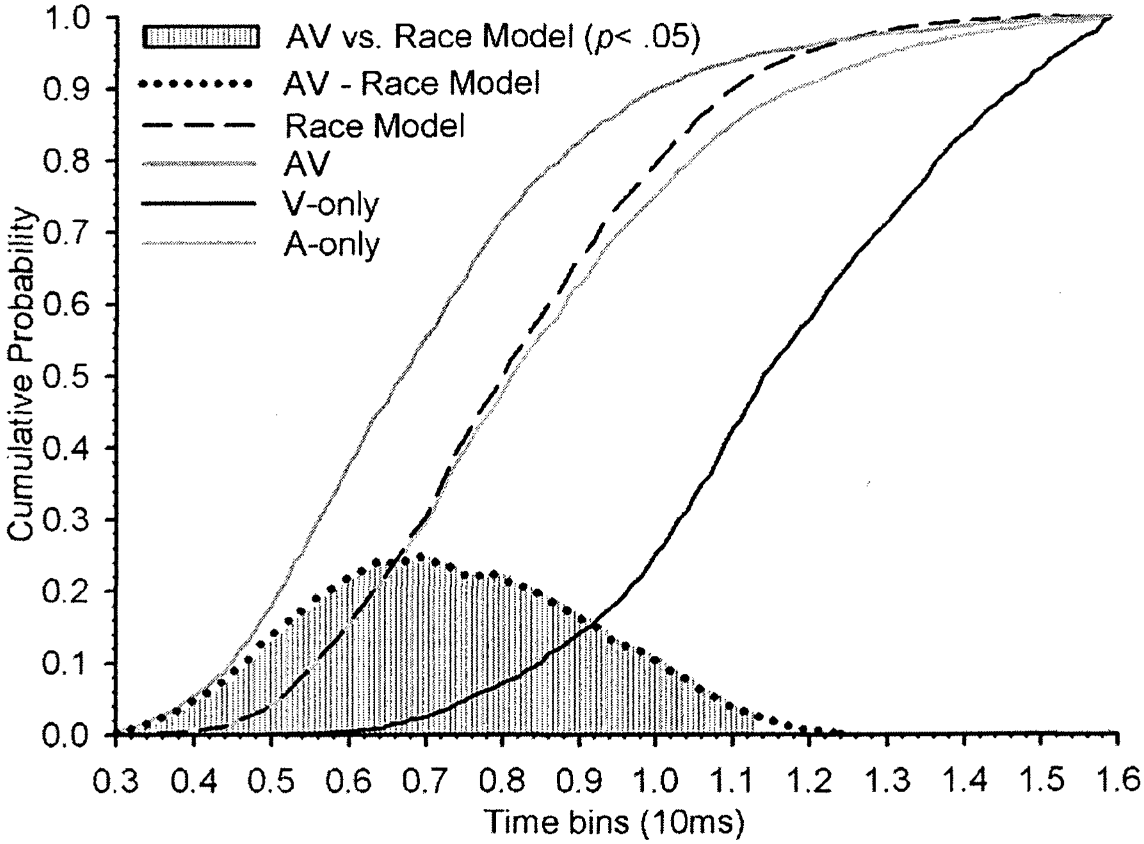

Figure 6: Cumulative distribution functions for reaction times of the conditions A-only (light grey, solid), V-only (black, solid), AV (dark grey, solid) and Race Model predictions ((A+V)-(AxV)) (black, dashed). The black doted line displays the difference between AV – ((A+V)-(AxV)) and the light grey shaded area indicates for which time bins this difference reached significance at α= .05. Each time bin (X-axis) indicates the probability of a response (Y-axis) occurring at that time point or faster.

2.7.2 Electrophysiological Analyses

Inspection of the ERP waveforms in response to A-only and AV trials revealed an absence of clear auditory evoked potentials at occipo-parietal sites. Auditory evoked potentials tend to be largest at fronto-central sites and typically decrease at more posterior sites (Figure 7). As expected V-only trials did not elicit auditory evoked potentials and are therefore not depicted in the figures. Given that the current experiment was designed to investigate multisensory effects at the level of early auditory processing, subsequent analyses were restricted to frontal and central sites at which the ERPs were maximal. To test for multisensory interaction the responses to the AV condition were compared to the combined response of A- and V-only (i.e., A+V).

An initial ANOVA with factors Hemisphere (left & right), Condition (A-only, A+V, and AV), and Anteriority (4 sites from frontal to centro-parietal regions (left: F3, FC3, C3, CP3; right: F4, FC4, C4, CP4) was conducted, followed by an ANOVA with factors Condition and Site (4 midline sites: Fz, FCz, Cz, CPz). Analyses were restricted to electrode sites showing clear auditory evoked potentials. Because the lateral sites did not yield any additional information (all *F* values < .31 and all *p*-values > .59), only electrophysiological results from midline sites are reported below. For each of the auditory evoked potentials (i.e., N1, P2) a separate ANOVA was conducted for peak latency and peak amplitude with factors Condition and Site.

Auditory N1 response. The results of the N1 amplitude analysis revealed a main effect of Condition ($F(2,26)= 15.4$, $MSE= 3.4$, $\varepsilon= .77$, $p< .001$) with N1 amplitudes in response to AV stimuli being smaller than those in response to A-only trials and to the sum of the unisensory conditions (A+V; Figure 7).

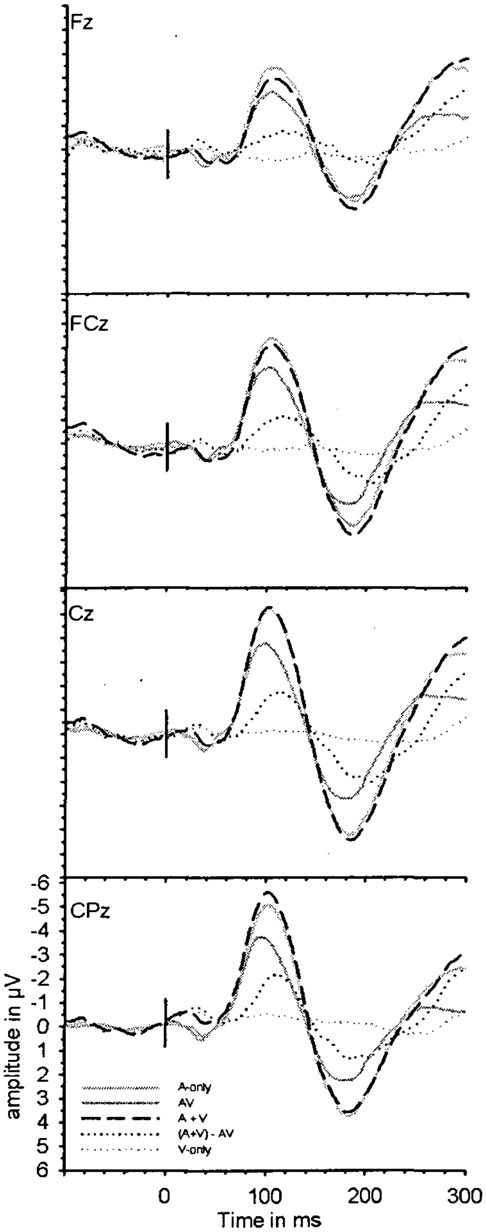

Figure 7: Grand average waveforms of auditory evoked potentials (P1, N1, & P2) of AV speech (dark grey, solid), auditory speech only (light grey, solid), visual speech only (light grey, dotted) the sum of the unisensory conditions (A+V; black, dashed), and the difference waveform between AV and (A+V) (black, dotted) at midline sites Fz, FCz, Cz, and CPz. The vertical line at 0 ms indicates stimulus onset.

The A-only response did not differ from the summed unisensory response at any of the four sites. A main effect of Site (F(3,39)= 7.0, MSE= 2.4, ε= .55, p< .01) revealed that N1 amplitudes were largest at the vertex and decreased at frontal and more posterior sites. Analyses of N1 latency did not reveal a main effect of Condition, a main effect of Site or a Condition by Site interaction (all F values < 3.4 and all p-values > .07).

Auditory P2 response. Analysis of P2 amplitude yielded a main effect of Condition (F(2,26)= 7.95, MSE= 23.1, ε= .61, p= .01) and a main effect of Site (F(3,39)= 27.9, MSE= 5.98, ε= .48, p< .001). The main effects were modulated by a Condition by Site interaction (F(6,78)= 9.3, MSE= 1.1, ε= .3, p= .001). As can be seen in Figure 7, the P2 amplitude was reduced during AV trials relative to A-only and to the sum of the unisensory trials at all sites, with the effect being reliable only at FCz, Cz, and CPz. These differences increased from frontal to central sites and decreased at more posterior areas. The P2 amplitude of A-only trials did not differ from the sum of the unisensory responses. Similarly, no main effect of Condition, Site or Condition by Site interaction were found for auditory P2 latency (all F values < .6 and all p-values > .55).

ERP analysis of items without preceding visual speech cues. One important difference between the stimuli in Experiment 1 versus Experiment 2 was whether or not the onset of the auditory and visual signals was simultaneous. In Experiment 1 the onset of the animal picture coincided with the onset of the animal vocalization. However, during AV speech as used for Experiment 2, the lips tended to start moving before a speech sound is produced (mean lag = 216 ms). However, for two of the words presented in Experiment 2 (i.e., lion and duck), the onset of the first lip movement coincided with the onset of the first speech sound. Analysis of those two items alone allowed us to address whether the reason we did not find an auditory N1 amplitude reduction in Experiment 1 was due to a lack of a visual cue preceding the auditory sound. Therefore, new ERP waveforms based on responses to those two words only were computed. Analysis of the N1 amplitude in response to the two spoken words 'duck' and 'lion' revealed a main effect of Condition ($F(2,26)= 4.12$, $MSE= 15.3$, $\varepsilon= .88$, $p= .03$), a main effect of Site ($F(3,39)= 16.1$, $MSE= 2.4$, $\varepsilon= .67$, $p< .001$) and a Site by Condition interaction ($F(6,78)= 3.4$, $MSE= 2.1$, $\varepsilon= .36$, $p= .04$). As was the case for the analysis including all stimuli, N1 amplitudes were largest at central sites. More importantly, during AV trials N1 amplitudes were reduced relative to the sum of the unisensory responses (A+V), indicating a significant multisensory effect. No main or interaction effects were found for N1 latencies.

2.8 Discussion (Experiment 2)

As for Experiment 1, results of Experiment 2 revealed multisensory interaction effects for behavioural was well as ERP responses. In Experiment 2, behavioural benefits were evident in the RT data but not in accuracy. This lack of multisensory benefit for accuracy data was likely due to auditory stimuli not being degraded; thus performance during A-only was already at ceiling, leaving no room for improvement during AV trials. However, AV speech trials in Experiment 2 were associated with faster RTs than unisensory trials indicating an AV speech benefit. The advantage of AV speech over A-only speech received further support from the race-model analysis. The analysis revealed that the prediction of the race-model that two sensory channels operate independently of one another was violated in the 320 to 960 ms response window, supporting the co-activation model (Miller, 1982). This indicates audiovisual interaction and reflects neural integration of the two unisensory information streams (Laurienti et al., 2006).

Even though multisensory interaction was not obvious for accuracy data, reaction time findings clearly demonstrated that speech perception in a face-to-face context led to better performance compared to when only listening to someone speak. One explanation for this AV benefit could be complementary information derived from the visual speech cues. Visual speech cues provide information regarding place of articulation (Campbell, 2008) which can help clarify auditory signals that might be ambiguous. As the current data showed, this is true not only when the auditory speech signal is distorted or when hearing is impaired; in the present study, the presence of visual speech cues also benefited those individuals with intact hearing in optimal environments. It has been shown that mouth movements correlate with second and third formant frequencies and possibly this information is used by the auditory system to process the auditory speech signal (Grant & Seitz, 2000b). These visual speech cues could help process sound more efficiently. Support for this increased efficiency comes from the electrophysiological data.

The ERP results showed that both the auditory N1 and the subsequent P2 were reduced in amplitude during AV trials relative to A-only and the summed A+V trials. This reduction provides support for multisensory interaction in form of a response reduction which has been shown by AV speech studies recording ERPs in response to syllables (Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). An AV induced amplitude reduction of the auditory N1 and P2 could indicate reduced sensory processing demands during multisensory as compared to unisensory processing (Giard & Perronet, 1999). Taking our behavioural and ERP findings together, Experiment 2 revealed that superior performance during AV speech perception relative to A-only trials was achieved with fewer neural resources being expended, which supports the notion of AV speech being processed more efficiently than auditory speech alone.

To address the possibility that the observed auditory N1 amplitude reduction was due to visual speech preceding the auditory speech cues, we analyzed a subset of stimuli for which the onset of the visual speech cues (i.e., mouth movement) coincided with the onset of the speech sound. Again, the auditory N1 amplitude was reduced relative to A+V trials. This finding speaks against the explanation that the absence of an auditory N1 amplitude reduction for AV non-speech trials in Experiment 1 was due to the simultaneous onset of visual and auditory stimuli (i.e., lack of preceding visual cues).

2.9 General Discussion

The objective of the current study was to examine the brain processes underlying multisensory perception when identifying non-speech stimuli like animal sounds and photos (Experiment 1) and whether they differ from those of AV speech stimuli (Experiment 2). Given that speech perception, including audiovisual face-to-face conversations, is crucial for human communication the hypothesis that the human brain developed specialised mechanisms devoted to AV speech stimuli is interesting to investigate. There were three important experimental design elements that enabled us to compare multisensory processes underlying AV speech and AV non-speech perception. First, in Experiment 1, we impoverished the visual signal by blurring the stimuli to make the processes more comparable to AV speech, where visual perception (i.e., lipreading) is more challenging than auditory speech perception. The behavioural data confirmed that V-only performance was the least accurate in both experiments. Second, by matching the stimuli of both experiments in their semantic properties, potential differences between AV speech and non-speech could not be due to a higher level mechanism such as semantic integration. Third, the task was identical in both experiments. The task did not require selective attention to a specific target; rather, participants were required to process and respond to every stimulus ensuring that attentional load was comparable across trials.

2.9.1 Behavioural Findings

As expected, both experiments revealed behavioural benefits associated with AV stimuli. For Experiment 1, accuracy scores significantly improved during AV trials compared to unisensory performance and reaction time was facilitated for the congruent AV condition relative to the unimodal A-only condition. Similarly, the reaction time data for Experiment 2 revealed significantly faster responses during AV speech trials compared to unisensory conditions (i.e., just listening or just watching someone speaking). Further analyses of the RT data of Experiment 2 revealed violations of the predictions made by the race model, indicating that the auditory and visual speech cues interacted (Miller, 1982). Results of both experiments, therefore, indicated that the availability of congruent visual and auditory signals led to behavioural benefits relative to when information was presented in only one modality. In addition to behavioural effects, multisensory interaction effects were also observed for electrophysiological measures but with important differences between the experiments.

2.9.2 Electrophysiological Findings

The findings of Experiment 1 revealed a reduction of the visual N1 amplitude at occipital areas during AV trials compared to when only a picture was presented. Previous studies using AV non-speech stimuli have found similar multisensory interaction effects for early visual processes (Giard & Peronnet, 1999; Molholm et al., 2002). Interestingly, Experiment 1 did not indicate multisensory interactions in auditory ERP components during AV non-speech trials which stands in contrast to the modulation of the auditory N1 during AV speech perception in Experiment 2. There, participants showed clear multisensory interaction effects reflected in auditory N1 amplitude reductions compared to when an auditory-alone stimulus was presented, a finding that is in line with previous research (Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). Given the large body of research pointing to the source of the N1 in the auditory cortex (Näätänen & Picton, 1987), this suggests that the presence of visual speech cues modulated and interacted with auditory processes at an early sensory-specific stage supporting previous findings (Besle et al., 2009; Besle et al., 2004; Möttönen et al., 2004).

Taken together, the findings suggest that AV non-speech stimuli as well as AV speech stimuli led to behavioural benefits compared to when perceiving information in the separate modalities on their own. However, the processes that were involved in achieving this interaction seemed to differ. One explanation for those differences could be related to task dependent sensory dominance. During AV speech perception, multisensory interaction effects were observed at the stage of early auditory processing, possibly because the dominant modality during speech perception is audition (Easton &

Basala, 1982), rather than vision. A similar argument can be made for the findings of the modulation of visual responses in Experiment 1. In that experiment, AV non-speech stimuli modulated early visual processing stages, possibly due to human perceptual preference. Our dominant modality is vision when it comes to object information processing (Colavita, 1974; Koppen, Alsius, & Spence, 2008; Posner, Nissen, & Klein, 1976). Consequently, the reason why we observed a multisensory interaction effect on the visual components may be due to an innate tendency to rely more on our eyes than on our ears when it comes to object recognition outside the domain of speech.

However, the question that remains is how this AV modulation of sensory specific processes is achieved. One explanation for the observed amplitude reductions during AV speech is that visual speech cues tend to precede the auditory signal and that those cues are used predictively by the auditory system to make processing more effective (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005).

Our findings stand in contrast to those of Stekelenburg and Vroomen (2007) who compared AV speech and non-speech perception. Their data suggested that there was no difference between ERP responses to their AV speech and AV non-speech stimuli as long as a visual cue preceded the onset of the auditory signal. Those results would suggest that the commonly observed auditory N1 amplitude reductions in AV speech studies could simply be due to the visual lip speech cues preceding the auditory speech signal and that there is nothing special about the nature or the neural processes of AV speech perception. With this finding in mind, it could be argued that the reason we observed an auditory N1 amplitude reduction during the AV speech experiment but not during the AV non-speech experiment was simply because in the latter, the onsets of visual and auditory cues were simultaneous. However, we ruled out this possibility by analysing data restricted to two of the AV speech stimuli that had simultaneous onsets of visual and auditory cues (i.e., the first lip movement coincided with the first audible utterance). Separate analyses of the responses to those two stimuli clearly revealed a significant multisensory effect in the form of an auditory N1 amplitude reduction, similar to that observed for all the stimuli averaged together. This finding speaks against the argument that the lack of auditory N1 amplitude reductions during AV non-speech object identification (Experiment 1) were only due to the lack of visual stimuli preceding and predicting the onset of the auditory signal.

However, the question remains as to what caused a reduction in N1 amplitude during AV. It is important to keep in mind that the visual information from the lips is not entirely redundant. Visual speech cues from the lips, tongue, and teeth provide information regarding place of articulation and could therefore resolve potential ambiguities in the auditory modality (Campbell, 2008). In other words, seeing lips not only helps to cue the onset of the utterance but provides visual cues that complement the auditory signal. We speculate that it is the combination of timing cues as well as complementary speech information provided by visual cues that aids auditory processing and leads to reduced processing demands. Similarly, in the AV non-speech study, auditory information supplemented the degraded visual information resulting in lower perceptual demands (i.e., a visual N1 reduction) as well as a behavioural benefit.

Both experiments revealed behavioural benefits together with electrophysiological amplitude reductions during AV perception relative to unisensory perception. During AV trials fewer neural resources were recruited at the level of early signal processing – visual processing for AV non-speech and auditory processing for AV speech – yet behavioural output was superior. This output-input relation suggests that relevant signal processing was more efficient during multisensory perception. This in turn could have important implications for higher level processing. Previous studies have shown functional relations between sensory-perceptual load and cognitive performance. More specifically, improvement of speech perception under poor hearing conditions (i.e., noisy environment or hearing impairment) can lead to better cognitive performance (Pichora-Fuller, 1996; Pichora-Fuller, Schneider, & Daneman, 1995; Tun, McCoy, & Wingfield, 2009). If visual speech cues lead to fewer resources being recruited for sensory signal encoding, more resources could be used for higher level cognitive tasks. Support for this idea comes from studies that have shown better memory performance for items presented in an AV format (Mastroberardino, Santangelo, Botta, Marucci, & Belardinelli, 2008). However, whether this benefit is due to more efficient sensory processing during multisensory perception is at this point speculative. Current research is underway to address this hypothesis.

2.10 Conclusion

The current study investigated electrophysiological processing differences between audiovisual non-speech and speech perception. The data indicated that the former modulated visual processing in visual cortex whereas the latter modulated auditory processes in the auditory cortex. The electrophysiological modulations took the form of a multisensory response reduction of the visual N1 for AV non-speech stimuli in Experiment 1 and of the auditory N1 for AV speech stimuli in Experiment 2. This dissociation could be due to differences in sensory dominance for speech and non-speech perception indicating that processes in the more dominant modality were modulated by processes in the less dominant modality. Importantly, multisensory benefits for behavioural measurements were evident in both instances. This suggests that the underlying mechanisms might be analogous but expressed differently depending on the dominance of the modality involved in the task. Furthermore, when the perceptual systems processed congruent auditory and visual stimuli, fewer neural resources were recruited at the sensory signal processing stage but the behavioural response system was nevertheless able to achieve equal or even superior performance, providing compelling support for the notion of increased multisensory efficiency.

2.11 Prelude to Study 2

Chapter 2 of the current work looked at the fundamentals of AV object recognition including AV speech. The goal was to investigate whether humans developed mechanisms specialized on combining visual and auditory speech cues to improve communication. The results from study 1 suggest that AV speech and non-speech object recognition operate with the same mechanism but apply it to the more dominant modality, with dominance being determined by the nature of the task.

Chapter 3 continued the investigation of mechanisms of AV speech perception, but approached the issue from a more applied direction. Chapter 3 addressed the question whether older adults can use auditory and visual speech cues to make speech perception and communication more effective.

Communication takes place in a social context and requires social interaction. Humans are a social species and effective communication is key for a society to function well. Communication is necessary to express needs and desires but also to convey warnings of threats. In other words, the ability to communicate is important but in order for communication to be effective there must also be the ability to understand a message. If, for example, hearing is impaired, communication can be affected as well. One factor that has been shown to influence the ability to perceive speech is an individual's age (Erber, 2002).

Aging is accompanied by cognitive and sensory changes which can influence communication (Hummert & Nussbaum, 2001). Due to demographic changes, especially in the developed world, the issue of aging increases in its importance. Advances in technology and medical knowledge and changes in lifestyle have allowed people to grow older than a few decades ago and, consequently, the proportion of older adults increases steadily. In Canada for example, it is estimated that by the year 2031 close to 25% of the population will be 65 years or older as compared to about 12% in 2001 (Government of Canada, 2002; Statistics Canada, 2005). This merits a better understanding of the aging process and age-related changes in order to ensure a high quality of living. Age-related changes at the cognitive, sensory, and motor levels can influence the ability to communicate and interact in a social context (Li & Lindenberger, 2002; Schneider & Pichora-Fuller, 2000). Regarding cognitive aging it has been shown that memory, attention, and working memory decline with age (Craik & Salthouse, 2000; McDowd & Shaw, 2000; Park et al., 2002; Zacks, Hasher, & Li, 2000).

Not all abilities deteriorate with age, though. Linguistic skills for example seem to remain intact and to some degree increase in older adults (Park et al., 2002). The intact linguistic skills have been shown to be help compensate for sensory deficits like decreased auditory functioning (Pichora-Fuller et al., 1995). Pichora-Fuller and colleagues (1995) demonstrated that older adults were better in using the context of a sentence than younger adults and that this benefit helped older adults to understand auditory speech in a noisy environment. Furthermore, the improved perceptual functioning led to an improvement in memory measured by word recall. This indicates that sensory strain taxes cognitive functioning and if sensory functioning is improved, higher level cognition can improve as well (Pichora-Fuller et al., 1995; Tun et al., 2009). That is, speech perception under adverse conditions is challenging and increased efforts to encode the spoken signal seem to come at the expense of other cognitive functions. However, if perception can be improved (i.e., increase S/N ratio) without increasing sensory processing effort (e.g., by using sentence context), higher level functioning is less affected by noise. In other words, the intact linguistic knowledge in older adults compensates for perceptual deficits (Pichora-Fuller et al., 1995). Support for this effortfulness hypothesis comes from studies that showed that an increased effort to encode sensory signals comes at the expense of cognitive performance (Cervera, Soler, Dasi, & Ruiz, 2009; McCoy et al., 2005; Rabbitt, 1968; Yampolsky, Waters, Caplan, Matthies, & Chiu, 2002). This effortfulness hypothesis is derived from the limited resources hypothesis that states that all mental processes, including perceptual, share resources, and if one of the processes requires more, less resources will be left for other processes (Just & Carpenter, 1992; Kahneman, 1973; Rabbitt, 1968). One of the factors that can lead to a higher demand of resources is changes in sensory functioning.

Aging can be accompanied by changes in the sensory systems (Schieber, 2006; Schneider & Pichora-Fuller, 2000). Visual functioning declines because of the hardening of the lens which influences the refractive power and leads to a reduction in visual acuity (Bergman & Rosenhall, 2001; Erber, 2002). Another phenomenon that is commonly observed is yellowing of the lens. Also, the eye becomes less opaque and pupil size decreases with age allowing less light to hit the retina. The last aspect leads to significant decreases in contrast sensitivity which has been shown to be an important variable in predicting daily functioning, more so than visual acuity (Schieber, 2006; Schneider & Pichora-Fuller, 2000).

The ear or the auditory system is not spared from age-related changes. These include changes of the outer ear, the middle ear and the inner ear. Particularly changes to the inner ear can lead to functional deficits. Older adults exhibit particular deficits in perceiving high pitch tones (i.e., 4 kHz and above) called presbycusis (Abel, Sass-Kortsak, & Naugler, 2000; Divenyi, Stark, & Haupt, 2005; Erber, 2002; Schneider & Pichora-Fuller, 2000; Wingfield & Tun, 2001; Wingfield, Tun, & McCoy, 2005). These deficits stems from the fact that aging is accompanied by inner hair cell loss close to the base of the cochlea which is the area where high frequencies are encoded. Functionally, this change reveals itself in impaired speech perception as a large number of speech sounds are around frequencies of 4 kHz (Erber, 2002; Schneider & Pichora-Fuller, 2000). Loss of neural temporal synchrony of auditory nerve cells has also been shown to contribute to deficits in speech perception (Pichora-Fuller, Schneider, Macdonald, Pass, & Brown, 2007). Perceptual deficits increase in noisy environments even in older adults with clinically normal audiometric functioning (CHABA - Committee on Hearing and Bioacoustics, 1988; Kim, Frisina, Mapes, Hickman, & Frisina, 2006). Kim and colleagues (2006) showed, that in order for older adults with clinically normal hearing to perform as well as younger adults in a speech recognition task, older adults required a S/N ratio that was about 40% more favourable than that of the younger adults. In quiet conditions older adults required a signal increase of about 20% to achieve the same level of sentence comprehension than younger adults.

Deficits in speech perception have clear negative implications for communication. In extreme cases this can lead to avoidance of social interactions, social gatherings and can lead to social isolation (Bergman & Rosenhall, 2001; Hummert & Nussbaum, 2001; Jagger, Spiers, & Arthur, 2005). Improving the ability to perceive speech can in turn improve quality of life (Erber, 2002). As outlined earlier there is evidence for young adults showing that AV speech is associated with better speech perception than when just listening to someone speak. One idea is therefore that AV speech could potentially be useful to offset hearing deficits such as presbycusis. This prediction is based on AV speech related benefits seen in patients with hearing impairments (e.g.: Bergeson & Pisoni, 2004; Grant & Seitz, 1998; Grant et al., 1998; Hay-McCutcheon et al., 2005; Möbes et al., 2006; Rouger et al., 2007; Tillberg et al., 1996).

Findings of AV speech perception in older adults have shown that the ability to integrate auditory and visual speech cues remains intact in older adults. Helfer (1998) presented non-sense sentences to older adults in background noise and participants were asked to repeat as many words as they could after each sentence. Sentences were presented either as A-only or in an AV mode. Results showed that older adults performed significantly better during AV than A-only indicating that additional visual speech improved speech perception independent of semantic knowledge. The magnitude of the benefit was the same as achieved by younger adults reported in an earlier study (Helfer, 1997). A study by Tye-Murray and colleagues (2007) involving older adults found perceptual benefits associated with AV speech for consonants, words and sentences presented in background noise. Their study, however, did not compare performance to that of a younger control group.

A study by Sommers and colleagues (2005) investigated AV speech perception in noise in younger and older adults with clinically normal sensory functions. The results showed that older adults performed more poorly in the lipreading condition but importantly, AV speech integration in older adults was intact. Younger and older adults benefitted equally from additional visual speech cues. Research from our lab provides further support for the notion that older adults are able to use visual speech cues as their performance improved significantly from A-only to AV speech perception in noise. (N. A. Phillips et al., 2009). The conclusion that older adults are able to integrate auditory and visual speech was also drawn in a study by Cienkowski and Carney (2002). Younger adults did better than older adults for lipreading but both age groups were equally susceptible to the McGurk-MacDonald illusion showing that fusion or integration of auditory and visual speech cues occurred for older adults as well. Furthermore, for trials where no fusion occurred, older adults revealed a more pronounced reliance on visual speech cues whereas younger adults relied more on auditory cues. Other studies support the finding that during AV speech processing older adults rely more on visual speech cues than younger adults is supported by other studies (Thompson, 1995; Thompson & Malloy, 2004). Thompson and Malloy (2004) superimposed infrequent dots at various spots on the speaker's face while participants were looking at the screen. Compared to younger adults, older adults detected more dots that appeared close to the mouth. The results demonstrated that older adults paid more attention to mouth regions whereas young adults distributed their attention across the whole face.

It should also be noted that research on multisensory integration in older adults is not restricted to the AV speech domain. Using a simple target detection paradigm, reaction time benefits in audiovisual conditions relative to unisensory conditions were equal for older and younger adults (Bucur, Allen, Sanders, Ruthruff, & Murphy, 2005; Bucur, Madden, & Allen, 2005).

Taken together there is evidence that older adults make use of visual speech cues to improve speech perception. Some behavioural findings even suggest that older adults benefit more from AV speech than younger adults (Hugenschmidt, Mozolic, & Laurienti, 2009; Laurienti et al., 2006). With respect to the inverse hypothesis (Stein & Meredith, 1993) this increased benefit for older adults might be due to reduced effectiveness of the sensory modalities on their own (Laurienti et al., 2006). That is, because auditory and visual functions decline with age, the benefit when both are combined should be even larger. Despite the fact that there are behavioural data on AV speech and aging, it is unknown whether AV speech processes involved in AV speech perception are the same or differ in older and younger adults. This is the question that will be addressed in Chapter 3 which entails an experiment investigating the differences – behavioural and electrophysiological – in AV speech perception in younger and older adults.

Chapter 3: Manuscript 2: AV speech and Aging

Visual Speech Cues Make Older Ears Hear 'younger':

An Investigation of Age-Related Differences in Audiovisual Speech Perception Using Event-Related Potentials.

3.1 Abstract

The current study addressed the question whether audiovisual (AV) speech can improve speech perception in older and younger adults in a noisy environment. Event-related potentials (ERPs) were recorded to investigate age-related differences in the processes underlying AV speech perception. Participants performed an object categorization task in four conditions; namely auditory-only, visual-only, AV_{speech}, and AV_{photo}. In the AV_{photo} condition participants saw a still picture of the speaker while listening to spoken words to see whether dynamic visual speech cues are required to achieve an AV benefit. Only younger adults showed a modest benefit associated with AV_{photo} trials, indicating the importance of dynamic visual speech cues, particularly for older adults. However, both age groups revealed an AV_{speech} behavioural benefit over unisensory trials. Older adults benefitted more from AV cues than younger adults as was seen in larger auditory enhancement scores. ERP analyses revealed an AV_{speech} related auditory N1 amplitude reduction relative to the summed unisensory response in both age groups. This amplitude reduction is interpreted as an indication for multisensory efficiency as fewer neural resources were recruited to achieve better performance. Younger and older adults also showed an earlier auditory N1 in AV_{speech} relative to A-only trials. In older adults this latency shift was larger and its size was predicted by basic

auditory functioning. Together, the results show that AV speech processing is intact in older adults and that they seem to benefit more from additional visual speech cues than younger adults possibly to compensate for sensory aging.

3.2 Introduction

Thanks to medical and technical advancements, better nutrition, and healthier lifestyles, the life expectancy and hence the proportion of senior citizens is increasing (Government of Canada, 2002; Statistics Canada, 2005). It has been shown that normal, healthy aging can lead to changes in sensory-perceptual abilities as well as higher-order cognitive functions (Schneider & Pichora-Fuller, 2000). Despite growing interest in age-related changes of sensory and cognitive functioning, many aspects remain poorly understood. One of these areas is the relation between aging and changes in audiovisual (AV) speech perception. The ability to integrate both sources of sensory information is especially important when information in one or both of the sensory channels is unclear or ambiguous (e.g., when having a conversation at a cocktail party with a lot of background noise). There is clear evidence, dating back to the 1950s, that the availability of visual speech input in a noisy acoustic environment is perceptually equivalent to boosting the volume of the auditory speech by 10-15 dB (Ross et al., 2007; Sumby & Pollack, 1954). This finding highlights the potential of AV speech to improve communication even in individuals who do not have a hearing impairment and are not trained in lipreading.

It is well known that there is an inverse relation between increasing age and the functioning of our sensory systems. With respect to auditory function, many older adults experience an age-related hearing loss (presbycusis) which affects the perception of high frequency sounds and can lead to difficulties in speech comprehension (Erber, 2002; Schneider & Pichora-Fuller, 2000; Wingfield et al., 2005). Even older adults with age-appropriate normal hearing reveal speech perception deficits in quiet listening conditions and this deficit is exacerbated in suboptimal, noisy environments where the auditory speech signal is ambiguous or degraded (CHABA - Committee on Hearing and Bioacoustics, 1988; Kim et al., 2006).

Similarly, visual abilities decline with age, including visual acuity and contrast sensitivity (Erber, 2002; Schneider & Pichora-Fuller, 2000). Even though there is ample evidence of sensory decline in older adults in each separate modality, less is known about their interactions in older adults. With both unisensory information channels compromised one could expect that audiovisual perception including AV speech would also decline with aging. Alternatively, age-related decline of sensory functioning makes the issue of multisensory interaction particularly interesting in light of the inverse effectiveness hypothesis. This hypothesis states that the gain derived from a multisensory stimulus is larger the less effective the unisensory channels are on their own (Stein & Meredith, 1993). Consequently, due to the decline in unisensory abilities, older adults could benefit more from the combination of audiovisual stimuli than younger adults whose sensory channels are intact.

The Effect of Age on AV Processing

Previous research on age-related changes in AV speech perception has led to a variety of findings. One explanation for this variance could be related to differences in stimulus materials (e.g., syllables, words, or sentences) and screening measures for participation (e.g., visual acuity, hearing level, cognitive functioning). The following section provides an overview of previous studies including brief descriptions of the participants, tasks and stimuli, and the general unisensory and multisensory findings.

Cienkowski and Carney (2002) investigated AV speech perception in a group of healthy younger and older participants who listened to consonant-vowel syllables in a quiet environment. A third group consisted of young controls who listened to syllables in a noisy background to match hearing thresholds to that of the older adults. All participants demonstrated normal values on tests of visual acuity, visual contrast sensitivity, and age-appropriate auditory hearing levels (with the exception that the older adults showed mild hearing loss for higher frequencies). The task was to name the syllable they perceived and syllables were presented auditory-only (A-only) and audiovisually (AV) to measure the extent to which participants showed the McGurk effect (McGurk & MacDonald, 1976) in the AV condition. In a classic McGurk paradigm an auditory syllable is dubbed onto a video of a speaker saying a different syllable (e.g., an auditory /ba/ combined with a visual /ga/, leading to the perception of /ga/). The McGurk effect refers to a perceptual phenomenon in which participants report the perception of a syllable that was neither presented auditorily nor visually, suggesting that auditory and visual speech cues were integrated. Cienkowski and Carney (2002) showed that all groups integrated syllables equally well. However, when integration failed the older adults and young controls with auditory background noise tended to choose the visual rather than the auditory alternative more often than younger adults with intact hearing (i.e., no noise). That is, older adults showed a larger visual bias than younger adults, suggesting that they relied on visual speech cues when auditory information was ambiguous possibly due to sensory decline.

Sommers, Tye-Murray, and Spehar (2005) showed poorer speechreading abilities for older adults compared to younger adults. Younger and older adults, screened for normal visual acuity, visual contrast sensitivity and pure-tone hearing thresholds, had to identify syllables, words and sentences presented in V-only, A-only and AV format. To measure the extent to which additional visual speech cues enhanced performance relative to A-alone trials (i.e., visual enhancement), error rate in the A-alone condition was equated in each group to 50% by titrating the intensity of a 20-talker background babble noise track. The same signal/noise (S/N) ratio was used for the AV condition. Older adults performed more poorly than younger adults in the V-only and AV conditions. However, after factoring out V-only performance, both age groups showed the same degree of visual enhancement indicating that younger and older adults were equally successful in integrating visual speech cues.

Even though previous studies have shown that the AV performance of older adults was generally poorer than younger adults, which may be explained by poorer speechreading abilities in the older adults, the ability to integrate auditory and visual speech cues remained intact (Cienkowski & Carney, 2002; Sommers et al., 2005; Tye-Murray et al., 2008). This conclusion has also been made in a bimodal target detection task (Bucur, Allen et al., 2005). In this study older and younger adults responded faster to AV targets than to unimodal targets. The analyses revealed that this facilitation was due to interaction of the two sensory channels allowing for the integration of multisensory information. Interestingly, older adults appeared to use the visual speech cues more than younger adults, possibly to compensate for sensory decline (Cienkowski & Carney, 2002; Thompson, 1995) or for limited attentional resources (Thompson & Malloy, 2004).

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

To briefly summarize, behavioural findings have consistently shown that the ability to integrate bimodal, audiovisual information was preserved in older adults (Bucur, Allen et al., 2005; Cienkowski & Carney, 2002; Helfer, 1998; Hugenschmidt et al., 2009; Laurienti et al., 2006; Sommers et al., 2005; Tye-Murray et al., 2007) and that older adults demonstrated either an equivalent multisensory benefit (Bucur, Allen et al., 2005; Cienkowski & Carney, 2002; Sommers et al., 2005) or even larger benefit (Hugenschmidt et al., 2009; Laurienti et al., 2006; Thompson, 1995) relative to young adults. Despite these findings, there is relatively little information about the neural mechanisms underlying AV speech perception in older adults. To date there have been a few studies investigating neural processes of AV speech perception and these have been restricted to young adults and stimuli usually comprised syllables rather than words or sentences.

Previous AV speech studies investigating the electrophysiological processes of AV speech mainly looked at early auditory event-related brain potentials (ERPs). Early auditory ERPs consist of a series of positive and negative voltage deflections which peak between 50 to 250 ms after stimulus onset. This sequence of obligatory brain responses is also referred to as the P1-N1-P2 complex. They are elicited by the presence of an auditory signal and their neural source has been suggested to lie in the auditory cortex (Eggermont & Ponton, 2002; Näätänen & Picton, 1987). Their functional role is related to discriminatory processes and stimulus detection.

AV speech studies recording ERPs elicited by syllables have showed that 1) the amplitude of the auditory N1 during the AV speech condition was reduced relative to the summed ERP responses of the A and V conditions (Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005) and 2) that the auditory brain processes were sped up relative to auditory-alone trials (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). Due to the visual speech cues preceding the first auditory speech cue by up to 300 ms, van Wassenhove and colleagues (2005) proposed that auditory processing during AV speech benefits from the visual cue which predicts what the auditory system can expect. Interestingly, the authors showed that the latency shift of the N1 increased with increasing predictability of the spoken syllable. Moreover, AV speech trials resulted in faster response times. Stekelenburg and Vroomen (2007) observed similar reductions in N1 amplitude and latency. Their study demonstrated that this effect seemed to be related to the visual cue temporally preceding the auditory cue, because they observed similar electrophysiological responses in trials involving non-speech stimuli, such as watching clapping hands.

The current study will address the question to what extent the behavioural as well as electrophysiological patterns during to AV speech perception in noisy environments differ between healthy younger and older adults. Given that the individual sensory modalities (i.e., vision and audition) in older age function less optimally than in younger adults, we predict, in line with the inverse effectiveness hypothesis, that older adults should benefit more from AV speech than younger adults. At the behavioural level, older adults should show faster and more accurate responses than younger adults during multisensory AV trials than during the unimodal conditions during which participants only listen to or watch someone talk. At the neural level we expect to see effects for early sensory components such as the auditory P1, N1, and P2. Similar to previous studies, we expect an amplitude reduction in these components during AV trials relative to the individual unisensory trials (A and V) and their summed response (A+V) as well as a latency shift of the N1 in that it will peak earlier during AV trials compared to unimodal trials. We hypothesize that this multisensory amplitude reduction and latency shift will be relatively larger in older adults than in younger adults.

The previous studies that concluded that the ability to integrate auditory and visual speech cues remains intact in older adults were based solely on behavioural findings. By measuring electrophysiological responses in addition to behavioural performance, the current study will be able to shed light on whether older adults recruit the same neural processes to integrate multisensory stimuli or whether they use different mechanisms than younger adults. Due to the high temporal resolution of ERP recordings, the current study will be able to pinpoint the time when differences in the processing stage occur in the range of ms.

In order to investigate AV speech integration in younger and older adults, ERPs were recorded while participants were asked to categorize spoken object names as natural or artificial. Participants were presented with stimuli under three conditions: auditory-only (A) trials during which they only heard the presenter speak, visual-only (V) trials during which they only watched the presenter speak (i.e., speechreading), and AV_{speech} trials during which they both heard and saw the speaker. We also included a fourth condition labelled AV_{photo} during which participants heard the speaker while looking at a photograph of her. In light of findings suggesting that the AV benefit might derive from visual speech cue preceding auditory speech information, this condition was included to determine whether the benefits associated with AV speech can be achieved just by a visual signal (i.e., a still face) preceding auditory speech information or whether it is necessary to have dynamic and congruent lip movements accompany the auditory speech signal in order to benefit from AV speech input.

3.3 Materials & Method

3.3.1 Participants

Twenty young and 19 older adults were tested; however, three younger and two older adults were excluded due to poor behavioural performance (reaction times and response accuracy differed from the group mean by more than two standard deviations) or due to electrophysiological recordings being too noisy. The final sample consisted of 34 individuals (N=17 in each age group) who were in reported good health. Participants were screened for intact sensory abilities. We assessed visual contrast sensitivity using the MARS Letter Contrast Sensitivity test (Haymes et al., 2006), auditory acuity by measuring pure tone averages (PTA; average hearing threshold for frequencies of 500, 1000 and 2000Hz), cognitive functioning using the Montreal Cognitive Assessment (MoCA; (Nasreddine et al., 2005)); these data plus important demographic information are summarized in Table 1. Although older adults had lower sensory functioning, both groups had age-appropriate and clinically normal contrast sensitivity scores (Haymes et al., 2006), and PTAs (ANSI, 1989). Only participants with a PTA below 20dB and PTA differences between the left and right ear of 10dB or less were included in the study. Participants gave their informed consent and the study was approved by the Concordia University research ethics board.

Table 1:

Demographics (mean (SD)) for Younger and Older Adults.

	younger adults	older adults	t-test
N	17 (12 female)	17 (12 female)	
Age	24.5 (3.43)	68.5 (5.03)	
Yrs. Of Education	17.0 (1.8)	15.1 (2.9)	*t*(32)= 2.4; *p*= .025
MoCA	28.4 (1.6)	27.3 (1.8)	*t*(32)= 1.8; *p*= .07
PTA	6.2 (4.3)	12.7 (4.4)	*t*(32)= 4.3; *p*< .001
MARS Contrast			
Sensitivity	1.7 (.04)	1.6 (0.1)	*t*(32)= 4.4; *p*< .001
S/N in dB	55/68	55/66	*t*(32)= 4.9; *p*< .001

MoCA= Montreal Cognitive Assessment; PTA= Pure Tone Average (Left & right ear); S/N= signal to noise ratio

3.3.2 Stimuli

The stimuli consisted of 80 spoken object names, 40 of which were natural objects (e.g., tree, pear, etc…) and 40 were artificial or man-made objects (e.g., bike, clock, etc…).

The items in the two categories did not differ on various psycholinguistic factors such as number of syllables (artificial: *mean*= 1.21 (*SD*= 0.41); natural: *mean*= 1.25 (*SD*= 0.44)), word frequency (artificial: *mean*= 645.1 (*SD*= 802.4); natural: *mean*= 454.0 (*SD*= 739.4)) and familiarity (artificial: *mean*= 558.3 (*SD*= 49.4); natural: *mean*= 536.8 (*SD*= 52.7)).

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

For all four presentation conditions (A, V, AV_{speech}, AV_{photo}), participants were exposed to background noise that was played at the same time the stimuli were presented. The background noise consisted of a multi-talker babble mask adapted from the Speech Perception in Noise test, Revised (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). We modified the original eight-speaker babble track by overlaying this track three times slightly shifted in time in order to create a background babble mask that was less variable in its intensity fluctuations. Importantly, the intensity of the background babble noise was individually adjusted relative to the word signals for each participant in order to assure an equivalent auditory perceptual load across the two age groups. To achieve the S/N adjustment, we played a list of object names that were not included in the experiment and asked participants to repeat the word that they have heard. We then adjusted the intensity of the babble noise until the participant identified about 55-60% of the words correctly. The S/N ratio was slightly more favourable for older adults (see Table 1) in order to achieve the same level of performance as the younger adults.

The experiment included four different conditions, namely auditory-only (A-only), visual-only (V-only), AV_{speech}, and AV_{photo}. In the AV_{speech} condition participants watched the video-clip of the woman speaking a stimuli word and heard the woman at the same time. Stimuli for the three other conditions were derived from these AV_{speech} stimuli. That is, the V-only condition consisted of the same stimuli as the AV_{speech} trials, but with the audio track removed. Likewise, the A-only trials were the same stimuli as AV_{speech} trials, but with the video removed. Similarly, the stimuli for the AV_{photo} condition were derived from the AV_{speech} stimuli; however, we replaced the dynamic video of the AV_{speech} trials with a series of still frames showing the image of the female speaker. That is, participants saw the face of the speaker but no lip-movements occurred.

In order to measure visual and auditory ERPs elicited by each stimulus, we inserted triggers at the onset of the lip movement and the onset of the sound, respectively, in all AV_{speech} stimuli. Since the AV_{speech} stimuli served as the basis for all other conditions, both trigger points were present in all four conditions. That is, the V-only condition included a trigger to mark the onset of the sound even though the sound was not audible to the participant. This was necessary in order assess multisensory interaction effects (see below).

3.3.3 Procedure

Participants were seated in a comfortable chair in a dimly lit room and informed consent was obtained before the testing session. Prior to the experimental task, we obtained sensory and cognitive performance scores and established the customized S/N. The experimental task consisted of a total of 640 trials with 160 trials in each of the four stimuli conditions. Each word was presented twice in each condition and the sequence of trial type was randomized. Stimulus presentation was controlled by software Inquisit 2.0 (2006) software. At the beginning of each trial a fixation dot was presented in the centre of the monitor for 200-300 ms (Figure 8). For A-only trials the dot was replaced by a blank screen and for trials involving visual information (i.e., V-only, AV_{speech}, and AV_{photo}) the fixation dot was replaced with a sequence of still frames of the speaker's face as lead-in (*mean*= 460 ms, *SD*= 55 ms), after which speaker's lips started to move in the V-only and AV_{speech} conditions. In the AV_{speech} condition, the lip movement preceded the first auditory speech cue on average by about 432 ms and varied from 36 to 600 ms (*SD*= 92 ms) depending on the word. In the V-only trials, no auditory speech was presented and in the AV_{photo} condition participants saw the same still frame for the entire duration of the trial. After the video had faded out, there was a 450ms inter-stimulus interval to give participants a sufficiently long response time window.

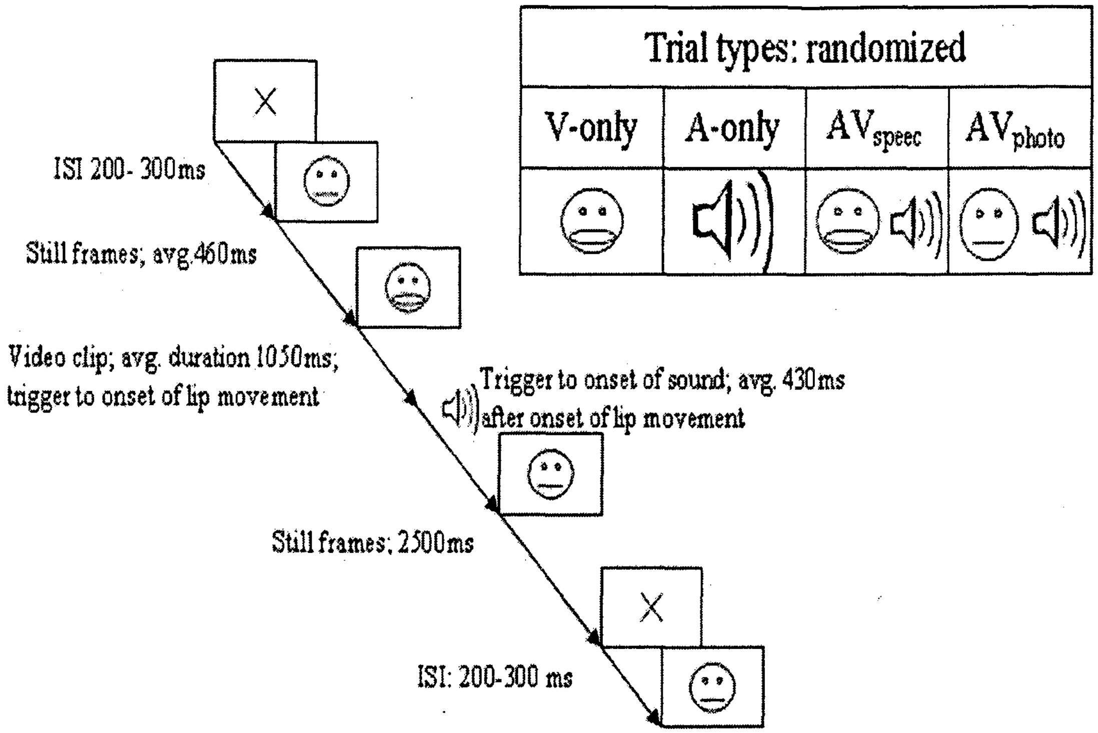
Trial types: randomized
V-only
A-only
AVspeec
AVphoto
ISI 200- 300ms
Still frames; avg.460ms
Video clip; avg. duration 1050ms;
trigger to onset of lip movement
Trigger to onset of sound; avg. 430ms
after onset of lip movement
Still frames; 2500ms
ISI: 200-300 ms

Figure 8: Schematic representation of a trial sequence. ISI= inter stimulus interval.

Participants were instructed to respond as to whether the stimulus word named a natural or man-made object by pressing one of two keys on a standard keyboard (i.e., *'S'* and *'L'* keys) with the side of response assignment counterbalanced across participants. Participants were instructed to respond as soon as they had identified the word. The stimulus onset asynchrony between the onset of the first video frame of consecutive stimuli was 4.5 seconds.

3.3.4 EEG Data Acquisition

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

All EEG data were re-referenced offline to linked ear lobes. The EEG signal was amplified using NeuroScan Synamps (Neuroscan, El Paso, TX, USA) and was recorded at a sampling rate of 500 Hz in a DC to 100 Hz bandwidth with electrical impedances kept below 5 kΩ. The continuous EEG was divided into 700 ms epochs defined by the onset of each stimulus trigger and included a 100 ms pre-stimulus baseline interval. EEG was filtered offline for frequencies between 1-30 Hz. Horizontal and vertical electrooculograms (HEOG and VEOG) were used to monitor eye movements and trials with HEOG activity exceeding +/- 50 μV were rejected. To assure a sufficient number of retained trials, excessive VEOG artefacts (i.e., eye blinks) were corrected using a spatial filter correction technique (Method 2, NeuroScan Edit 4.3 manual, 2003). Trials with EEG activity and other motion artefacts exceeding +/- 100μV were rejected. Furthermore, only trials with correct responses were included in our analyses. For a participant to be included in the analysis a minimum of 70 accepted trials per presentation condition had to be retained. As mentioned above, each stimulus contained two triggers, one to mark the onset of the lip movement and the other to mark the onset of the sound. This was the case even for A-only and AV_{photo} trials where no lip-movement was apparent and for V-only trials where no spoken word was audible.

This was important to assess multisensory interactions. To do so, we compared the ERPs to the AV_{speech} trials triggered by the onset of the sound (i.e., when signals from both modalities were available) to the sum of the ERPs to unisensory conditions (i.e., A+V). For this comparison to be valid, each of the triggers had to be aligned to the same point in time, namely the onset of the sound which was real in the case of AV_{speech} and A-only trials but virtual in the case of V-only. This careful alignment of time points allowed us to accurately assess any non-linear interaction effects present in the AV_{speech} trials (van Wassenhove et al., 2005). Having triggers placed at the onset of lip-movement and at the onset of the sound, we were able to measure visual and auditory evoked potentials, respectively. Onset of lip movement elicited clear visual evoked potentials in lateral occipito-temporal areas but because this study focused on auditory responses, visual evoked potentials are not discussed further.

As mentioned earlier, the electrophysiological response to an auditory stimulus typically consists of a series of early, sensory-driven and automatic ERPs referred to as P1-N1-P2 complex (Eggermont & Ponton, 2002; Näätänen & Picton, 1987). The amplitude of the auditory N1 was calculated by computing the absolute peak-to-peak microvolt difference between P1 and N1. The amplitude of the P2 was calculated by computing the absolute peak-to-peak microvolt difference between N1 and P2. Component latencies were recorded at the components' peaks relative to the 0 ms stimulus onset.

3.4 Results

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

3.4.1 Behavioural Results

Accuracy. Figure 9 presents the accuracy results for younger and older adults. In order to investigate an effect of age on accuracy, a 2 (Age Group; younger adults & older adults) x 4 (Condition; A-only, V-only, AV_{speech}, AV_{photo}) repeated measures ANOVA was conducted. The analysis revealed a main effect of Condition ($F(3,96)= 222.7$, $MSE= 59.5$, $\varepsilon= .49$, $p< .001$) in that responses to AV_{speech} trials were more accurate than responses in A-only and AV_{photo} trials which did not differ from each other. Responses to V-only were the least accurate of all the conditions.

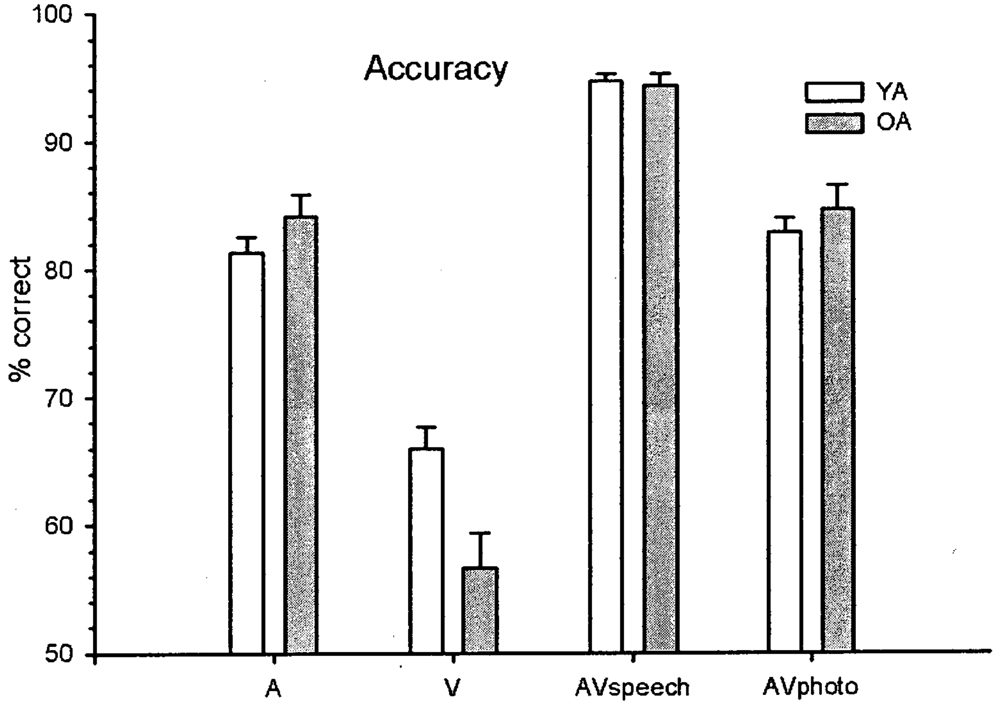

Figure 9: Mean accuracy data and standard error bars on the natural/man-made judgement task for younger adults (YA; white bars) and older adults (OA; grey bars) for the four presentation conditions: A= auditory only, V= visual only, AV_{speech}, AV_{photo}.

The analysis also revealed an Age Group by Condition interaction (*F*(3,96)= 8.8, *MSE*= 59.5, *p*= .002). Subsequent pairwise comparisons showed that the V-only condition was driving this interaction.

Subsequent pairwise comparisons showed that for the V-only condition, older adults performed less well than younger adults, indicating poorer lip-reading ability. No group differences were found for A-only due to the fact that we successfully equated the groups on listening performance. Interestingly, accuracy scores for AV_{speech} did not differ between groups, reflecting equivalent performance under multisensory conditions. No main effect of Age Group (*F*(1,32)= .61, *MSE*= 91.6, *p*= .44) was evident.

Response Time. Figure 10 presents the reaction time data for younger and older adults. To investigate an effect of Age on reaction time (RT), a 2 (Age Group) x 4 (Condition; A-only, V-only, AV_{speech}, AV_{photo}) repeated measures ANOVA was conducted, which revealed a main effect of Condition (*F*(3,96)= 824.9, *MSE*= 11444.5, ε= .46, *p*< .001). AV_{speech} trials resulted in the fastest responses relative to all other three conditions, whereas RTs in V-only trials were slower than RTs in the other three conditions (see Figure 10). RTs for A-only trials did not differ from AV_{photo} trials. A main effect of Age Group (*F*(1,32)= 5.6, *MSE*= 57266.01, *p*= .024) indicated that older adults responded more slowly than younger adults.

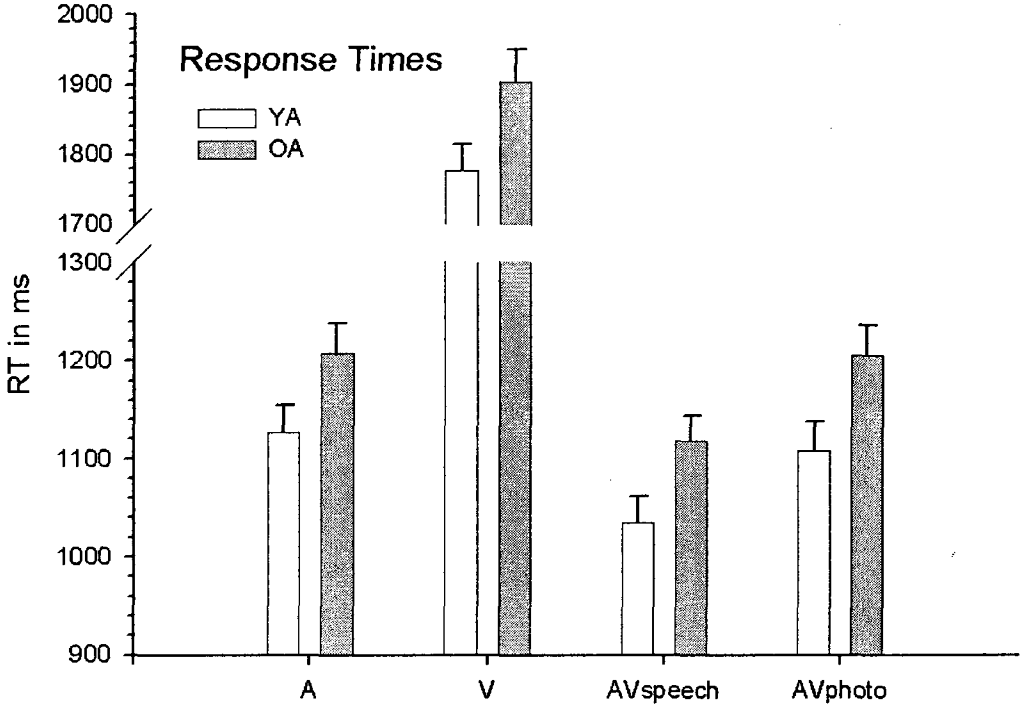

Figure 10: Mean response time data and standard error bars on the natural/man-made judgement task for younger adults (YA; white bars) and older adults (OA; grey bars) for the four presentation conditions: A= auditory only, V= visual only, AV_{speech}, AV_{photo}.

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

The race model is said to be violated when the probability of a particular response time is higher in the multisensory condition than the joint probability of the unisensory responses ((A+V)-(AxV)) for that given response time. A violation supports the co-activation model which states that RT facilitation is due to the interaction of the two sensory channels (Miller, 1982). To test for co-activation, the RT data are plotted as cumulative distribution functions (CDFs). We divided the RT interval from 0.4s to 2.5s into 10ms bins and calculated the likelihood that a response occurred at a given response time or faster. The CDFs of older adults and younger adults are plotted in Figure 11. These data were analysed by conducting paired t-tests at each time bin to determine if the observed AV_{speech} response time probabilities were higher than the joint probability of the unisensory responses ((A+V)-(AxV)) (i.e., test of race model violation).

In the younger adults, the CDF values for RTs to AV_{speech} trials were significantly larger (*p*< .05) than the CDF values of the joint probability of the unisensory responses for each time bin from 590ms to 1240ms. These data were remarkably similar to those of the older adults (*p*< .05; 600ms until 1260ms). Responses to AV_{speech} trials were faster than unisensory responses (A-only and V-only) and faster than the race model predictions which is shown in Figure 11 by the CDF of AV_{speech} response times shifted to the left relative to the other curves (Figure 11). To test whether multisensory integration occurred for AV_{photo} trials, we similarly compared CDFs of the AV_{photo} RTs to the CDF of the RTs from the unisensory conditions ((A+V)-(AxV)). Data of the younger adults revealed violations of the race model predictions during the AV_{photo} condition from 750 to 1100ms; however, for the older adults, no significant differences emerged (see Figure 11b & 11c).

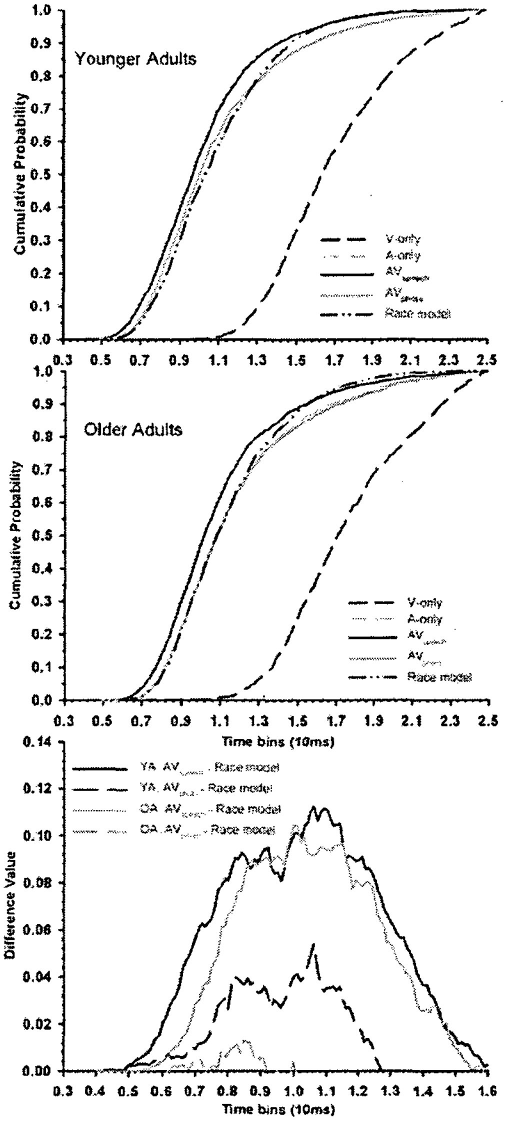

Figure 11: Cumulative Distribution Functions (CDFs) of reaction times obtained for younger adults (YA; top panel) and older adults (OA; bottom panel) in the four presentation conditions: A= auditory only (grey, dashed), V= visual only (black, dashed), AV_{speech} (black, solid), AV_{photo} (grey, solid). The predicted CDF from the Race model ((A+V)-(AxV)) is presented in the black, dashed-dotted line. The bottom panel presents the difference values between AV_{speech} & Race model predictions (solid) and AV_{photo} & Race model predictions (dashed) for younger (black) and older adults (grey).

Auditory and Visual Enhancement. To examine the benefit derived from combining information from two modalities, we calculated the visual enhancement (VE), which reflects the amount of benefit gained from the additional visual speech cues, separately for accuracy and RT values ((AV-A)/A). Additionally we analyzed auditory enhancement (AE; i.e., (AV-V)/V), which is the amount of benefit gained from the additional auditory information. Separate one-way ANOVAs were conducted to investigate age differences for AE and VE. The results indicated a significantly larger AE in the accuracy scores for older adults than younger adults ($F(1,32)= 6.4; p= .02$) (Figure 12). There was no reliable group difference for the VE ($F(1,32)= 3.3; p= .08$).

3.4.2 Electrophysiology of Auditory Evoked Potentials (AEP)

Multisensory interaction for neural responses can be assessed by comparing the multisensory response to the arithmetic sum of the individual unisensory responses (Calvert et al., 2001). Significant deviations from this sum (i.e., either response enhancement or reduction) signify non-linear interaction effects. In the current study, we compared the ERP responses to the multisensory AV_{speech} condition to the sum of the responses to the two unisensory conditions A-only and V-only (i.e., A+V). It is important to note that the waveform for the V-only condition was computed by averaging the EEG traces that were time-locked to the temporal point of the onset of the auditory signal (which was, of course, not audible to the participant in this condition). This allowed us to compare brain activity when information from both modalities was present (AV_{speech}) to the brain activity associated with the same point in time when information from only one modality was present (A and V). As expected, V-only trials did not elicit an AEP and are therefore not depicted in the figures.

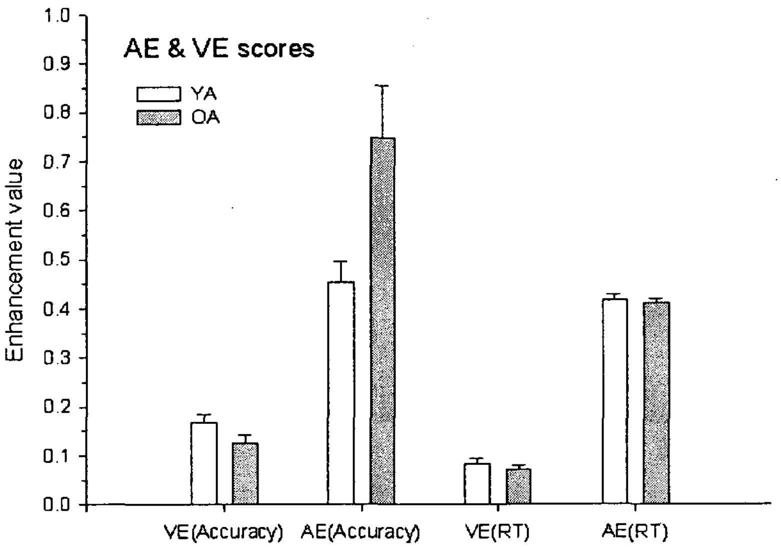

Figure 12: Visual (VE) and auditory enhancement (AE) values for accuracy and reaction time (RT) for younger adults (YA; white bars) and older adults (older adults; grey bars).

Furthermore, we limited our analyses to early sensory processes namely the AEPs P1, N1, and P2. As AEPs tend to be largest at the vertex, the figures in this section depict group average waveforms at site Cz only.

For each age group, a repeated measures ANOVA was conducted with factors Condition (A-only, A+V, AV_{speech}, AV_{photo}), Hemisphere (left & right), and Anteriority (6 sites from frontal to occipital lobes). Neither of the two groups showed a main effect of Hemisphere (younger adults: $F(1,16)= .78$, $MSE= 1.2$, $\varepsilon= 1.0$, $p= .39$; older adults: $F(1,16)= 2.2$, $MSE= 2.7$, $\varepsilon= 1.0$, $p= .16$) or interaction effects involving Hemisphere. Given that results from lateral sites did not yield additional information, subsequent ANOVAs included factors Condition and Site (6 midline sites: Fz, FCz, Cz, CPz, Pz, and Oz) and only results of AEPs from midline sites are reported here.

For each of the three auditory ERP deflections (i.e., P1, N1, P2) a separate ANOVA was conducted for peak latency, measured at the peak of the component of interest, with the factors Age Group, Condition, and Site. A similar ANOVA was conducted for the peak-to-peak amplitude differences between P1-N1 and N1-P2.

P1 latency. Analysis of P1 latency did not reveal a main effect of Condition ($F(3,96)= 2.9$, $MSE= 974.3$, $\varepsilon= .8$, $p= .06$) or Age Group ($F(1,32)= .12$, $MSE= 1811.8$, $p= .74$) nor an Age Group by Condition ($F(3,96)= 1.9$, $MSE= 974.3$, $p= .14$) interaction.

N1 latency. Analyses of N1 latency did not show a main effect of Age ($F(1,32)= .68$, $MSE= 2978.6$, $p< .42$). However, there was a main effect of Condition ($F(3,96)= 20.2$, $MSE= 1656.9$, $\varepsilon= .71$, $p< .001$) which showed that the N1 peaked significantly earlier during AV trials relative to the other three conditions and that the AV_{photo} condition did not differ reliably from A-only (see Figure 14 & 15). The N1 peaked slightly later at Oz relative to more anterior sites as seen in a main effect of Site (*F*(5,160)= 2.7, *MSE*= 249.7, ε= .64, *p*= .044). An Age x Condition interaction (*F*(3,96)= 3.9, *MSE*= 1656.9, *p*= .022) was due to a more pronounced N1 latency shift from A-only trials to AV_{speech} trials in older adults (see Figure 13). Subsequent planned comparisons showed that the N1 in response to A-only trials for older adults peaked significantly later than for younger adults at fronto-central sites (150ms vs. 135ms), but for AV_{speech} trials the latency of the auditory N1 did not differ between both age groups (120ms for both).

P2 latency. The P2 latency analysis revealed a main effect of Condition (*F*(3,96)= 4.2, *MSE*= 1634.6, ε= .925, *p*< .001) with the P2 peaking earlier during the AV_{speech} condition relative to the A-only condition. No main effect of Age for P2 latency was evident (*F*(1,32)= .32, *MSE*= 6499.98, *p*= .58). There was a Site x Age interaction (*F*(5,160)= 6.9, *MSE*= 942.1, *p*= .001) which showed that the P2 peaked later for older adults only at Oz.

N1 amplitude. The N1 amplitude, defined as the P1-N1 peak-to-peak amplitude difference, was subjected to the same ANOVA as used for previous analyses. The results revealed a main effect of Condition (*F*(3,96)= 37.6, *MSE*= 2.1, ε= .83, *p*< .001) which showed that the N1 amplitude in response to AV_{speech} trials was smaller than responses to A-only, AV_{photo} trials and to the sum of A+V, which was larger than the other three conditions (see Figures 14 & 15 for ERP responses from younger and older adults, respectively). The N1 amplitude in response to A-only trials did not differ from responses to AV_{photo} trials in either group. A main effect of Site (*F*(5,160)= 16.1, *MSE*= 1.1, ε= .56, *p*< .001) showed that amplitudes were largest at fronto-central and smallest at occipital sites. No main effect of Age or an interaction involving the factor Age was found.

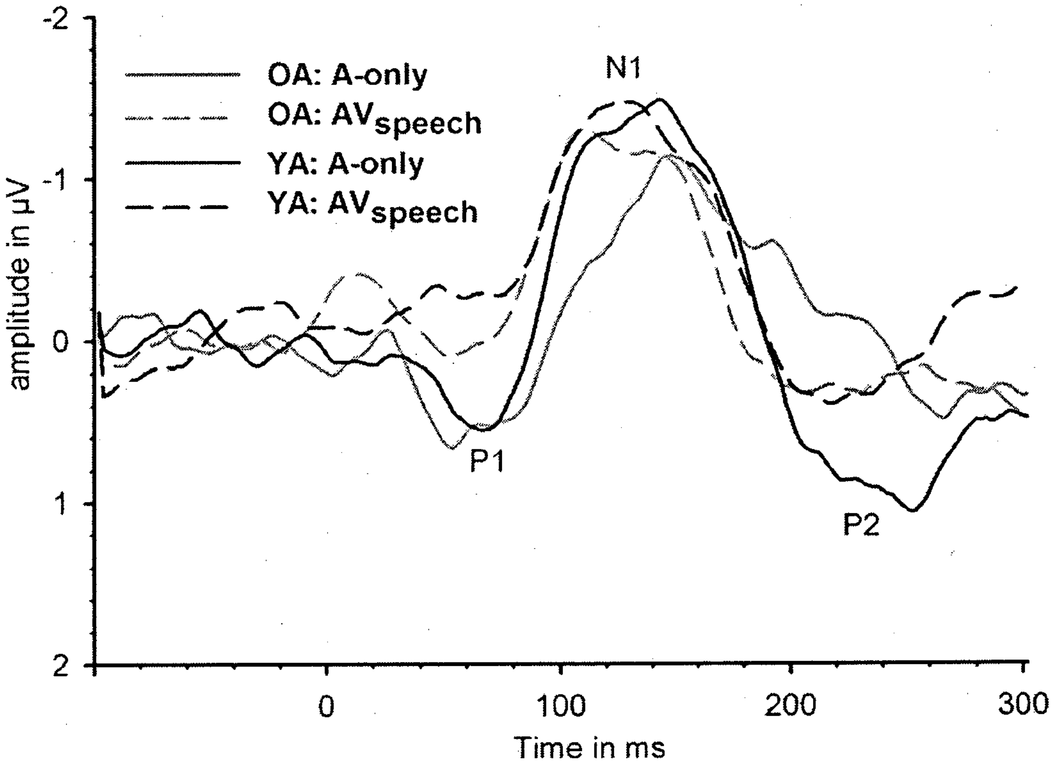

Figure 13: Group average waveforms of younger adults (YA; black) and older adults (older adults; grey) to auditory-only (A-only; solid lines) and AV$_{\text{speech}}$ trials (dashed lines) at site Cz.

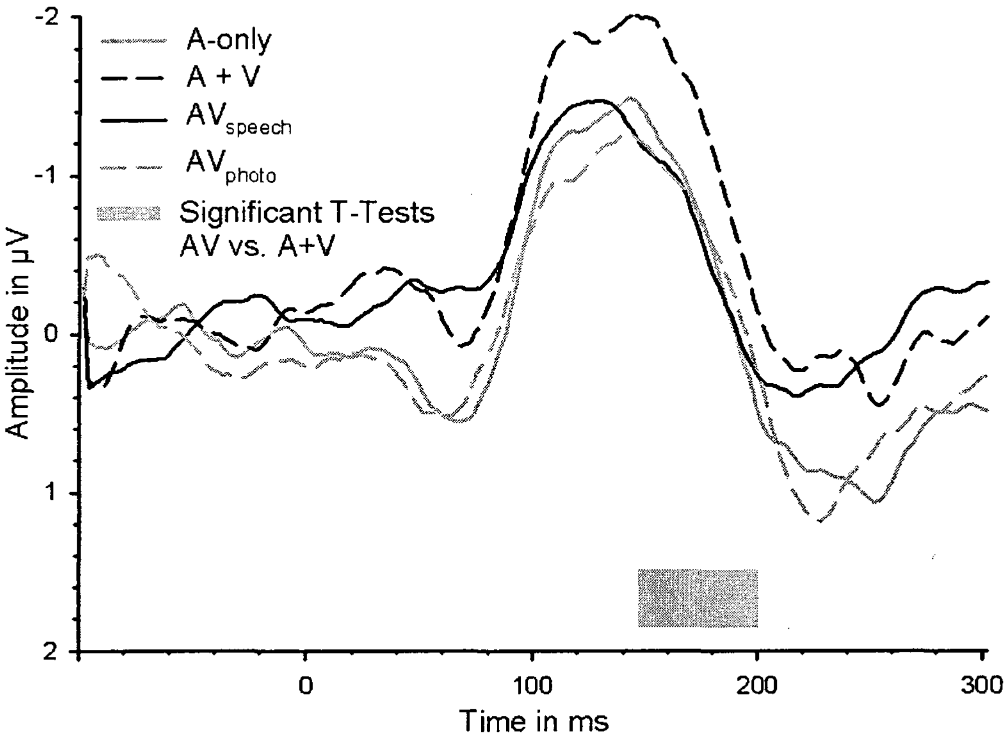

Figure 14: Group average waveforms of younger adults at Cz for conditions A-only (grey, solid), A+V (black, dashed), AV_{speech} (black, solid), and AV_{photo} (grey, dashed). Grey blocks indicate the time interval for which the AV_{speech} waveform differed significantly from the summed A+V waveform (*p*< .05).

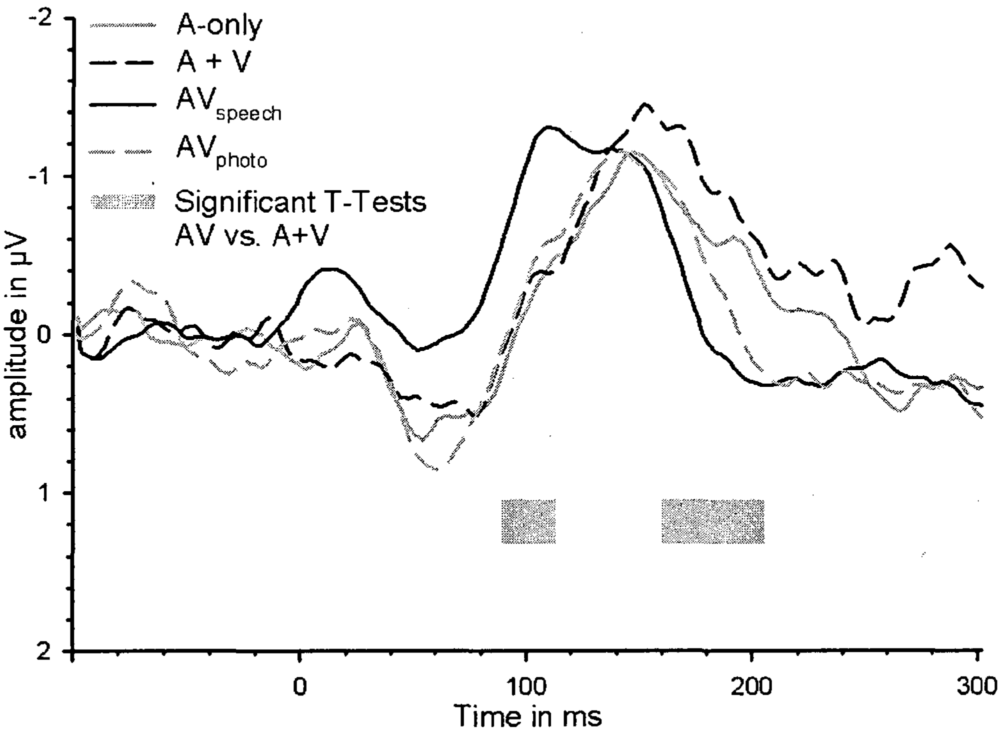

Figure 15: Group average waveforms of older adults at Cz for conditions A-only (grey, solid), A+V (black, dashed), AV_{speech} (black, solid), and AV_{photo} (grey, dashed). Grey blocks indicate the time interval for which the AV_{speech} waveform differed significantly from the summed A+V waveform ($p< .05$).

P2 amplitude. Analysis of the P2 amplitude, defined as the N1-P2 peak-to-peak amplitude difference, revealed a main effect of Condition (*F*(3,96)= 19.8, *MSE*= 2.4, ε= .78, *p*< .001) which showed that the summed response of A+V yielded the largest P2 amplitude relative to the other conditions which did not differ from each other. A main effect of Site (*F*(5,160)= 28.9, *MSE*= 3.1, ε= .53, *p*< .001) showed that amplitudes were largest at fronto-central and smallest at occipital sites. A main effect of Age (*F*(1,32)= 5.3, *MSE*= 21.8, *p*= .028) showed that P2 amplitudes were smaller for older adults than for younger adults.

Time point of multisensory interaction. Our analyses indicated that AV speech led to multisensory interaction at the level of early sensory processes such as the auditory P1-N1-P2 complex. To assess the time point of multisensory interaction more closely, we computed ERP difference waveforms by subtracting the responses to AV speech trials from the summed response of A+V trials. At each of the six midline electrodes we then conducted a t-test at each time point from 0-300ms after stimulus onset (i.e., 150 time points) and applied the most conservative criterion for significance proposed by Guthrie and Buchwald (1993), namely a minimum of 12 consecutive t-values larger than the critical value of 2.14. Cz, which is where AEPs were most prominent, older adults revealed significant differences from 88 to 114 ms after stimulus onset which is around the time period of the P1 and from 160 to 208 ms corresponding to the N1-P2 ERP complex (Figure 15). For the group of younger adults significant differences between AV_{speech} and A+V at Cz emerged only for the later time window, namely at 142-198ms after stimulus onset (see Figure 14).

3.4.3 The Role of Sensory Functioning

Predicated on the inverse effectiveness idea and its predictions related to sensory effectiveness, we examined the relations between basic sensory functioning (i.e., visual contrast sensitivity and auditory PTA thresholds) and our dependent variables. Initial calculations of correlations between contrast sensitivity and various dependent outcome measures revealed that contrast sensitivity correlated only with accuracy performance on AV_{speech} trials ($r(32)= .36, p= .037$). The relation suggested that higher contrast sensitivity led to better AV_{speech} perception but interestingly not to better lipreading (V-only) per se. However, a standard multiple regression with AV_{speech} accuracy as the dependent variable and age, cognitive functioning, hearing level, and contrast sensitivity as independent variables did not reach significance. A standard multiple regression analysis was conducted between the N1 latency shift from A-only to AV trials as dependent variable and age, contrast sensitivity, cognitive functioning and PTAs as independent variables. Table 2 shows the results of the analysis, including the bivariate correlations between the independent variables and the dependent variable, the unstandardized regression coefficients (B), the standardized regression coefficients (β), the squared semipartial correlations (sr^2), the intercept, R and R^2 (Tabachnik & Fidell, 2001). This regression revealed that hearing level was the only significant predictor of the size of the auditory N1 latency reduction, predicting almost 20% of the variance in N1 latency shift (Table 2). Figure 16 shows that higher hearing thresholds (i.e., poorer auditory functioning) led to a greater reduction in N1 latency on AV trials compared to A-only trials.

Table 2:

Regression on N1 Latency Shift from A-only to AV_{speech} Trials

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

* $p < .01$

CS= Contrast sensitivity, MoCA= Montreal Cognitive Assessment, PTA= Pure Tone average.

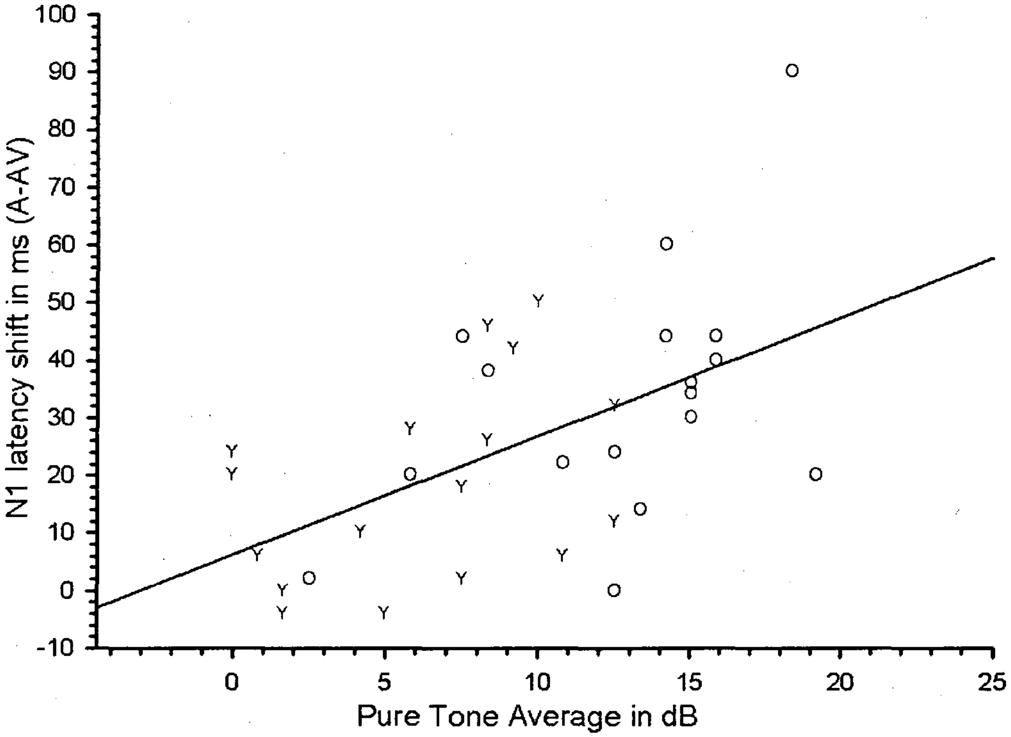

Figure 16: Regression of auditory functioning as measured by listening thresholds (PTA= Pure Tone Average) on the shift in the auditory N1 latency from A-only to AV_{speech} trials (A-AV). O= older adults; Y= younger adults. Regression equation: N1 latency shift = .54*PTA+57.4 ms.

3.5 Discussion

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

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

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

According to the inverse effectiveness hypothesis (Stein & Meredith, 1993), namely that the gain derived from a multisensory stimulus should be larger the less effective the unisensory stimuli are on their own, we hypothesized a relatively larger multisensory benefit in older adults than in younger adults. Our older adults exhibited poorer visual and auditory sensory functioning than the younger adults and could be considered to be in a 'permanently' suboptimal environment. Thus, they should benefit relatively more from AV speech (Hugenschmidt et al., 2009; Laurienti et al., 2006). The RT data did not support the inverse effectiveness hypothesis, because the amount of improvement from A-only to AV speech trials did not differ for younger and older adults (93 ms and 89 ms, respectively), nor did the visual enhancement and auditory enhancement effects. Interestingly, the RTs for older adults during AV trials were as fast as the RTs for younger adults during A-only trials. In other words, the addition of visual speech cues brought older adults to the hearing performance (A-only) of younger adults, a finding that has also been shown by Laurienti and colleagues (2006). Moreover, the cumulative distribution functions of the RTs of older adults during the AV_{speech} condition overlapped with those of the A-only condition for younger adults.

However, the accuracy data partially supported the inverse effectiveness hypothesis. The improvement from A-only to AV_{speech} trials and the magnitude of the visual enhancement effect was the same for younger adults and older adults. It is possible that the older adults did not reveal a larger visual enhancement because we titrated the auditory S/N so that both age groups were matched on auditory perceptual load and accuracy. This means that when the listening condition was manipulated to produce an equivalent auditory perceptual load, older adults were as efficient as younger adults in integrating visual speech cues to enhance speech perception (Sommers et al., 2005). This is interesting given that the other index of multisensory benefit, the auditory enhancement effect, was significantly larger for older adults. That is, even though older adults performed significantly worse than younger adults on the lipreading task, they were as efficient in integrating the auditory and visual speech cues. Interpreted in the context of the inverse effectiveness hypothesis, the older adults showed a larger multisensory gain relative to younger adults even though their baseline visual information processing was less effective. Consequently when both information channels were combined performance of both groups was identical. Similarly, a target detection study that simulated myopia in young participants showed a multisensory benefit over unisensory (i.e., auditory and visual alone) performance (Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003), in line with inverse effectiveness. As was the case for young participants with simulated myopia, the older adults of the current study showed marked improvement in performance under V-only to multisensory AV_{speech} trials indicating that visual deficits could be offset by additional, congruent auditory information.

Turning to the ERP data, we focused our analyses on early sensory ERP responses of the auditory system namely the P1, N1 and P2. In both younger and older adults, we demonstrated an amplitude reduction of the auditory N1 in response to AV_{speech} trials relative to the unisensory A-only condition and the summed response of the two unisensory conditions, A+V. This finding corresponds to previous studies on AV speech processing in younger adults (Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005) and, importantly, extends it to older adults. Näätänen and Picton (1987) have shown that N1 amplitude becomes smaller if the auditory stimulus is predictable. In the context of AV speech, van Wassenhove and colleagues (2005) explain the phenomenon of an N1 amplitude reduction with the increased predictability of the auditory speech sound due to the visual speech cue which precedes the auditory signal.

The N1 amplitude reduction in the present study reflects multisensory interaction in form of a response reduction in the AV condition compared to the sum of the unisensory responses (i.e., AV < A+V) and, based on previous research, suggests that visual information interacted with auditory cues at the level of the auditory cortex (Besle et al., 2004; Campbell, 2008; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). Interestingly, the size of the amplitude reduction from A-only and A+V to AV trials was the same for younger and older adults (Figure 17a), suggesting that the neural processes underlying AV speech processing were intact in older adults. This finding is in line with our behavioural data.

In addition to the amplitude reduction, both groups exhibited a significant latency shift, with the multisensory N1 response peaking earlier than that of the unisensory A-only and the summed A+V response. This is in line with previous findings (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). Interestingly, this facilitation of auditory processing speed was larger in older than in younger adults (Figure 17b). According to Van Wassenhove and colleagues (2005), N1 latency shifts in response to AV stimuli depend on the degree of predictability of the visual speech cue. With this in mind, our findings indicate that older adults were more apt than younger adults in extracting useful information from visual speech cues to predict or supplement the upcoming spoken utterance. Also, whereas younger adults showed multisensory interaction effects from 142-198 ms after stimulus onset, older adults showed multisensory interaction effects even earlier, namely between 88 and 114 ms, corresponding to the latency window of the P1. Again, this suggests that the neural processes underlying AV speech perception of older adults use the visual speech cues more effectively.

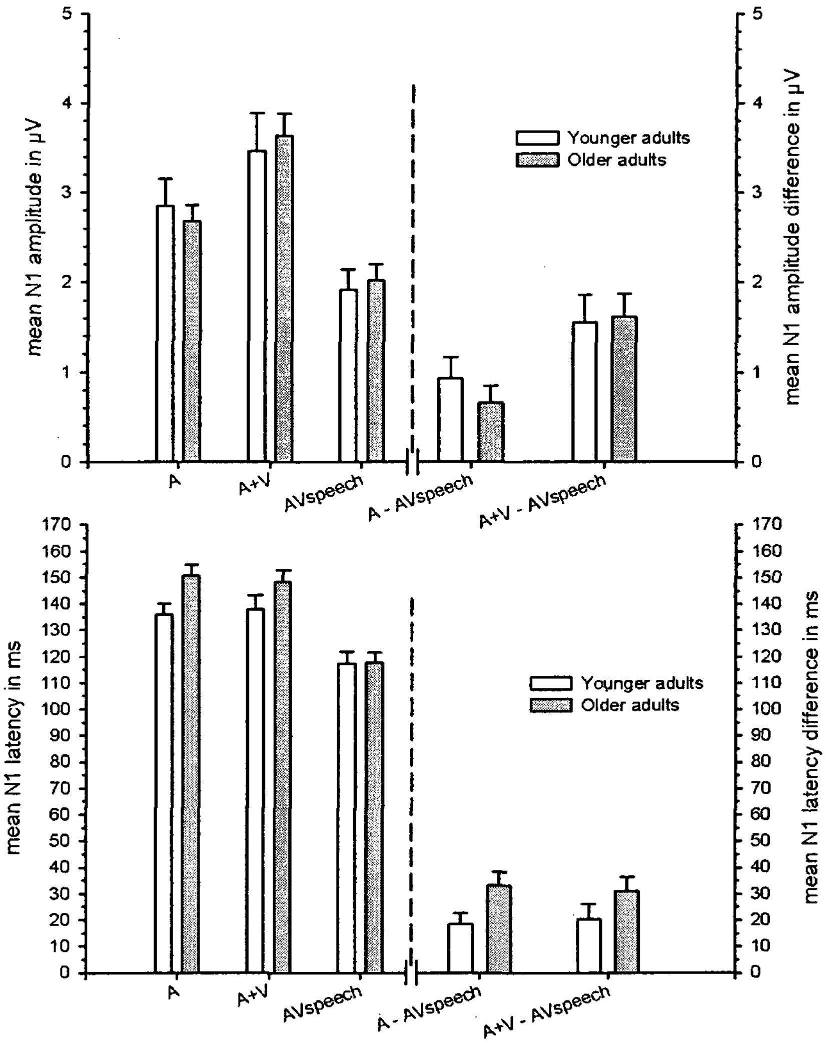

Figure 17: a) left half: Mean auditory N1 amplitude values (i.e., P1-N1 amplitude difference) with standard error bars at electrode Cz of older (grey) and younger adults (white) for conditions A, A+V, and AV_{speech}. Right half: Mean auditory N1 amplitude difference plus standard error bars for A-AV_{speech} and (A+V)-AV_{speech}. b) left half: Mean auditory N1 latency values with standard error bars at electrode Cz of older (grey) and younger adults (white) for conditions A, A+V, and AV_{speech}. Right half: Mean auditory N1 latency difference plus standard error bars for A-AV_{speech} and (A+V)-AV_{speech}.

Our results suggest that older adults, compared to younger adults, are not better lipreaders per se but rather better "lip/speech integrators". One explanation for this could be impoverished auditory functioning. The hearing thresholds, although clinically normal, were higher in the older adults than in the younger adults. Interestingly, hearing level predicted the size of N1 latency shifts from A-only to AV trials in all participants, regardless of age. In other words, participants with poorer auditory functioning exhibited a more pronounced speeding of auditory processing at the neural level when visual speech cues were made available. Our interpretation is that individuals with less optimal hearing compensate for diminished auditory function by making more efficient use of visual speech cues. The idea that older adults rely to a larger extent on additional visual speech cues is supported by other studies on AV speech perception in older adults (Cienkowski & Carney, 2002; Thompson & Malloy, 2004).

In the current study, both RT and accuracy findings revealed AV speech benefits in conjunction with the electrophysiological results which showed an amplitude reduction of the auditory N1 in response to AV speech trials. Interestingly, this indicates that fewer neural resources were expended to achieve better performance, suggesting that AV_{speech} was processed more efficiently than auditory or visual speech alone in both younger and older adults. The idea of efficiency is very intriguing as it leads to some interesting implications. Assuming that the brain has only a finite amount of neural resources available to perform both sensory as well as cognitive processes (Just & Carpenter, 1992; Rabbitt, 1968), efficiency in processing is crucial. Speech perception in noisy environments is more effortful for older adults (CHABA - Committee on Hearing and Bioacoustics, 1988; Pichora-Fuller et al., 1995; Schneider & Pichora-Fuller, 2000). If signal processing is effortful, more processing resources have to be devoted to sensory encoding. This, in turn, leads to fewer resources available for higher level processing such as working memory (WM). Research has shown that WM performance declines with age in general (Park et al., 2002; Wingfield & Tun, 2001) and especially for auditory stimuli presented in background noise (Pichora-Fuller et al., 1995; Schneider & Pichora-Fuller, 2000). If AV_{speech} signals make speech processing more efficient at the sensory level, which was demonstrated in the current study, resources that are not used could be recruited to improve higher level processes such as WM (Just & Carpenter, 1992; Schneider & Pichora-Fuller, 2000). Whether this hypothesis holds true is still an open question but preliminary findings suggest that this seems to be the case (Baranyaoiva, Winneke, & Phillips, submitted).

In addition to age differences in AV speech processing, the current study also investigated a basic property of the mechanisms of AV speech. To address the question of whether dynamic visual speech cues are necessary to achieve an AV benefit, we included an AV_{photo} condition. In that condition, the auditory speech signal was presented alongside a static photograph of the speaker. That is, no visual speech signals were available to cue the onset of auditory speech information. Neither the accuracy nor the RT data of older adults revealed an AV_{photo} benefit. However, younger adults showed a modest improvement in RT from having a photo available as their data showed violations of predictions made by the race-model suggesting that younger adults integrated the auditory speech stimulus and the photograph of the speaker. It should be noted that this effect was not as large as the benefit they demonstrated for the AV_{speech} condition. What might account for this finding? Recall that all stimuli were presented in the stream of ongoing phonological masking, making it difficult to know when an auditory stimulus might occur. Possibly, the appearance of the photograph made younger adults more attentive to a potentially upcoming auditory stimulus. Older adults, on the other hand, may have taken a more conservative approach in their resource allocation and only relied on cues that were strongly informative of the onset and nature of an upcoming speech signal, namely the actual onset of lip movement. For both age groups, the ERPs in response to the AV_{photo} trials did not differ significantly from those of A-only trials in their peak amplitude and latency. Thus, the behavioural and electrophysiological data for older adults clearly showed that just looking at a still image of a speaker was not effective enough to elicit a perceptual benefit over only listening to the speaker. For younger adults, the presence of a photograph might have been sufficient to raise the global level of attention which in turn led to a small performance benefit. Since the AV_{photo} condition led to RT benefits in younger adults but not to electrophysiological interaction effects it could be argued that this condition primed the behavioural response system but did not lead to genuine multisensory interactions at the sensory-perceptual level as was the case for the AV_{speech} condition.

3.6 Conclusion

This study demonstrated that AV speech perception remained intact in older age and facilitated speech perception in a noisy environment. Despite the fact that older adults were less skilled in reading lips, they performed as well as the younger adults during AV speech trials. Interestingly, despite a similar pattern in behavioural measures, the electrical brain responses indicated that AV speech resulted in earlier multisensory interaction effects and relatively larger N1 latency shifts in older adults. This suggests that in the brains of older adults visual speech cues were used more effectively to improve auditory speech processing in the presence of background noise. One explanation for this age-related benefit is that the availability of visual speech cues compensated for less-than-optimal auditory processing. That is, the additional visual speech cues made older adults' ears hear "younger". Overall, younger and older adults manifested reduced neural activity and better behavioural performance during AV speech trials compared to unisensory trials. The possibility that increased efficiency under multi-sensory conditions could have important implications for resource allocation and higher-level cognitive performance is currently a focus of our research attention.

Chapter 4: General Discussion

Through a series of experiments this dissertation addressed a fundamental question as well as a second, more applied issue, both of which are relevant to further our understanding of multisensory perception in humans. The first study aimed to establish whether neural processes underlying audiovisual (AV) speech are fundamentally different than those involved in AV non-speech perception or in other words, whether AV speech holds a special place relative to other multisensory processes. The second study approached the issue of AV speech from a more applied direction. Given the knowledge of age-related declines in speech perception, I investigated the extent to which older adults benefit from AV speech and how this behavioural benefit would be reflected in the brain. By comparing older and younger adults I was able to address age differences underlying AV speech processing.

4.1 AV Speech vs. AV Non-Speech

4.1.1 Same principle, different processes

Using an object identification task the results from the first study revealed behavioural benefits associated with AV stimuli over unisensory stimuli such as improvements in response accuracy as seen for the first experiment (AV non-speech). Electrophysiologically those behavioural benefits were accompanied by modulations of the visual N1 at occipital electrode sites. More specifically, the AV interaction effect was evident as an amplitude reduction of the visual N1. No multisensory interaction was apparent for auditory evoked potentials such as the auditory N1.

The second experiment of the first study (AV speech stimuli) revealed faster response times to AV speech stimuli relative to only seeing (V-only) or only hearing someone speak (A-only). Analyses of the response time data testing the race model (Miller, 1982) indicated violations of the prediction of sensory signals being processed independently. Rather, visual and auditory speech cues interacted to promote better performance than would be predicted by the race model, suggesting that neural integration took place (Bucur, Allen et al., 2005; Laurienti et al., 2006; Miller, 1982). In addition to behavioural improvements, Experiment 2 showed AV-induced amplitude reductions of the auditory N1. Data from the first study suggest that AV speech modulated auditory processes whereas identification of AV non-speech objects influenced visual processes.

Why might this difference occur?

The reason for this differentiation could have to do with sensory dominance. Speech perception is an inherently auditory task making audition the more dominant modality during AV speech processing rather than vision (Easton & Basala, 1982). For stimuli outside the domain of speech, vision seems to be the more dominant sense (Posner et al., 1976). For example, the Colavita effect suggests that the visual signal is more potent than auditory cues in terms of accessing the response system (Colavita, 1974; Koppen & Spence, 2007a, 2007b; Sinnett, Spence, & Soto-Faraco, 2007). In a series of experiments Colavita (1974) demonstrated that participants responded more often to a visual flash than an auditory tone presented simultaneously. Interestingly, in about 18% of those trials participants reported to be unaware that an auditory stimulus was presented. The Colavita effect has also been shown in a target detection task using more complex stimuli like line drawings of objects and naturalistic sounds (Sinnett et al., 2007). Functionally the Colavita effect translates to humans trusting their eyes more than their ears – think about it the next time you try to pick out a ripe watermelon!

The question that arises is why during multisensory perception the dominant modality is affected or modulated by the less dominant. One explanation is that the less dominant modality carries information that is to some degree redundant but it also contains complementary information. This is especially the case when the signal in the dominant modality is ambiguous. If the pictures in Experiment 1 had not been blurred, the auditory signal would not have supplied any or very little additional information. In terms of AV speech it can be argued that visual speech cues are not entirely redundant, but actually provide complementary information such as cues about place of articulation (Campbell, 2006, 2008; Grant & Seitz, 2000b; Munhall & Vatikiotis-Bateson, 1998; Summerfield, 1979, 1983). These cues augment the auditory signal in ideal listening environments but should do so even more when listening to speech takes place in a noisy environment or when hearing is impaired.

Even though the results of the first study indicate processing differences for AV speech and AV non-speech object recognition there are some key aspects common to both AV conditions. First, both experiments revealed superior behaviour under AV conditions. Second, both showed modulations of sensory specific processes in form of amplitude reductions. This leads to the third common aspect which is of a more theoretical nature. The combined finding of superior behavioural performance (i.e., faster reaction times and/or higher accuracy) of AV trials over unisensory trials along with reduced ERP amplitudes indicates an increase in processing efficiency. In other words, fewer neural resources were recruited, yet performance was the same or even better, making multisensory processing more efficient than unisensory processing. Combing results from electrophysiological recordings from single cells in auditory cortical areas of animals and human ERP responses allows making inferences regarding the neural basis of sensory evoked potentials. For example, it has been shown that spike firing rate in the cat primary auditory cortex increases with increasing stimulus intensities (Schreiner, 1998) and similarly, the auditory N1 amplitude has been shown to increase with stimulus intensity (Antinoro, Skinner, & Jones, 1969; Beagley & Knight, 1967; Näätänen & Picton, 1987; Picton, Woods, Baribeau-Braun, & Healey, 1976). These parallels between human and animal auditory response functions suggest that changes in ERP amplitudes could be due to changes in the level of activity of neurons in the auditory cortex. If applied to the findings presented here, the auditory N1 reduction in response to AV trials was potentially due to reduced firing rates (i.e., fewer neural resources). Combined with better behavioural performance this reduction could reflect multisensory efficiency. Multisensory efficiency and its potential implications are discussed in more detail further below.

The finding of reduced ERP amplitudes in response to AV stimuli has been reported by numerous EEG studies in the multisensory literature (e.g.: Besle et al., 2009; Besle et al., 2008; Besle et al., 2004; Giard & Peronnet, 1999; Pilling, 2009; Reale et al., 2007; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). I should be noted though, that animal studies recording from single cells in the superior colliculus as well as cortex commonly reported enhanced responses to multisensory stimuli together with enhanced behavioural performances (e.g.: Meredith & Stein, 1986; Stanford et al., 2005;

Stein et al., 1988; Stein & Meredith, 1993; Wallace et al., 1996). One potential explanation for the discrepancy of multisensory effects might be due to the different levels of analyses; that is, single cell recordings on the one hand and aggregate activity from large populations of cells on the other. Whether this is the driving factor for this difference remains to be determined.

Findings of the first study taken together suggest that the general principle of multisensory efficiency was common to AV speech and non-speech perception, but which specific processes implement this principle was dependent on the dominant modality. Which modality is more dominant depends on the signals and the task. That is, during object identification tasks AV speech and AV non-speech stimuli were both processed more efficiently, but the former caused modulations of auditory processes whereas the latter modulated visual processes.

4.1.2 Multiple stages of multisensory interaction

Given the high temporal resolution of ERPs, the results from the first study shed light on the processing stages at which multisensory interactions occurred. The data showed modulations at sensory specific stages suggesting that information from one modality influenced signal processing in the other. This happened fairly early and given what is known about the ERP components that were modulated, these interactions could represent tuning of feature analyses. In AV speech for example, the visual speech signal might serve as a frequency filter for the auditory modality. This possibility will be discussed further below. The AV non-speech study (Experiment 1) revealed that congruency did not affect those early feature analysis stages, which suggests that object identification was not completed that early. The absence of early congruency effects have been demonstrated by other AV studies (Lebib et al., 2004; Yin, Qiu, Zhang, & Wen, 2008; Yuval-Greenberg & Deouell, 2007). However, in Experiment 1 differences between AV_{match} and $AV_{mismatch}$ trials emerged starting at around 350ms after stimulus onset and were largest at the vertex. The increased negativity for mismatching A and V stimulus pairs relative to matching ones resembled that of an ERP component called N400. Other studies using pictorial stimuli have reported N400-like responses to incongruent stimulus pairs with a more frontal distribution (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999; West & Holcomb, 2002). A similar frontal negativity for $AV_{mismatch}$ trials was observed in Experiment 1 as well (see Figure 2).

The N400 is said to reflect assessment of the semantics or meaning of a word (or object) and evaluates how well it fits within a given context (Connolly & Phillips, 1994; Holcomb & McPherson, 1994; Kutas & Hillyard, 1980; Kutas et al., 2006; McPherson & Holcomb, 1999; Sitnikova et al., 2008; Sitnikova et al., 2003; West & Holcomb, 2002). The less congruent an object (or word) is with its context the larger the amplitude of the N400. The 'N400'-like effect observed in Experiment 1 could therefore indicate a second AV interaction stage. This particular stage might represent the timing when the concepts conveyed by the auditory and visual signals were integrated. If this integration or combination poses difficulties, as is the case when seeing a cat but hearing a 'moo', an N400-like is elicited. Similar N400 effects have been found by other studies using mismatching AV non-speech stimuli (Lebib et al., 2004; Molholm et al., 2004; Yin et al., 2008; Yuval-Greenberg & Deouell, 2007).

These sequential interaction effects can be aligned with fMRI data revealing a complex network of regions that were differentially activated by AV stimuli relative to unisensory stimuli (Amedi et al., 2005; Bushara et al., 2003; Callan et al., 2003; Calvert, 2001; Calvert, Brammer, & Iversen, 1998; Calvert et al., 2001; Calvert & Thesen, 2004; Driver & Spence, 2000; Macaluso & Driver, 2005; Macaluso et al., 2004; Saito et al., 2005). Neural connectivity studies with primates provide further support to the notion that modalities are interconnected and that senses interact in several cortical regions (Cappe & Barone, 2005; Ghazanfar & Schroeder, 2006). In humans as well as animals, areas identified as multisensory interaction sites were sensory-specific as well as hierarchically higher up in the processing sequence. This is congruent with the results of Experiment 1. The results showed early, sensory modulations of the visual N1 involved in feature analysis and feature discrimination at sensory-specific cortices followed by later effects at the level of conceptual or semantic processing, namely the N400. Results regarding the exact neural source of the N400 vary but it likely lies within the left temporal lobe and possibly in the superior temporal sulcus (Van Petten & Luka, 2006). These sequential multisensory effects have also been documented in a MEG study on AV speech perception by Möttönen and colleagues (2004). The first effect was a reduction of the magnetic counterpart to the auditory N1 (i.e., M1 or M100) which was localized in the primary auditory cortex subsequent to which further AV interaction effects were found in the posterior part of the superior temporal sulcus between 250 – 600m after stimulus onset. A study measuring EEG coherence found early and late multisensory modulations (Yuval-Greenberg & Deouell, 2007) which provides further support for sequential AV interaction effects observed in Experiment 1. According to the authors the earlier effect was likely related to low-level feature processing whereas the later modulation at around 300ms might have indicated higher level feature binding and multisensory object formation (Yuval-Greenberg & Deouell, 2007). Sequential interaction effects were not apparent for Experiment 2 (AV speech) of the first study because it was not designed to do so. Nevertheless, early interaction effects at sensory specific stages were evident for Experiment 2 as well.

To my knowledge this was the first ERP study on AV speech perception that used complete words as stimuli, which is ecologically more valid than individual syllables. Nevertheless the finding of early amplitude reductions of the auditory N1 in response to AV speech tokens is consistent with other studies using ERPs to assess AV speech perception (Besle et al., 2009; Besle et al., 2004; Pilling, 2009; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). This is important as it suggests that early multisensory modulations underlying AV syllable perception are similar to the effects observed when processing entire words and possibly sentences as well.

To briefly sum up the main findings of the first study; it was shown that the principle underlying AV speech and AV non-speech object recognition seemed to be the same, namely that of multisensory efficiency. However, the principle was implemented differently depending on which modality was more dominant for a given class of objects. For spoken objects as during AV speech perception, audition was more dominant and consequently early auditory processes, likely in the auditory cortex, were modulated during AV speech trials. Object recognition using AV non-speech stimuli was dominated by vision and therefore modulations of early visual processes, likely in extrastriatal areas, were observed. This suggests that AV speech was processed differently than non-speech stimuli which stands in contrast to observations by Stekelenburg and Vroomen (2007). The authors of that study did not find differences between AV speech perception and audiovisually presented actions like hand-clapping in terms of the underlying electrophysiological responses. This led the authors to conclude that AV speech is not special. According to them, the auditory N1 amplitude reduction was observed due to a visual cue preceding, and hence predicting, the onset of the auditory stimulus.

Re-analysis of the data of Experiment 2, however, showed that the presence of a preceding visual speech cue cannot be the whole story. Analyses of a subset of the AV speech stimuli for which the onset of the first lip movement coincided with the onset of the sound induced the same auditory N1 amplitude reduction as did stimuli for which the auditory onset lagged the visual cues. One alternative explanation for the reduction of the auditory N1 during AV speech is the nature of the visual speech cues. As already mentioned, visual speech information during AV speech perception is not completely redundant. Possibly it was the complementary information derived from the concurrent visual speech cues that allowed the auditory system to process the acoustic speech signals more efficiently. The role of visual cues in augmenting auditory processes are described further below.

4.2 AV Speech in Younger and Older Adults

4.2.1 Implications for multisensory efficiency

A similar auditory N1 amplitude reduction for AV_{speech} trials relative to unisensory trials was also evident in the ERP results of the second study. Additionally, a speeding of the auditory N1 during AV_{speech} trials was observed which is consistent with previous studies (Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005). These electrophysiological multisensory interaction effects were seen in younger and older adults. AV interaction effects were also reflected in behavioural variables as both accuracy and reaction time data showed significant improvements during AV_{speech} trials relative to unisensory responses. To my knowledge this is the first study to look into ERPs during AV speech processing in noise, and it is also the first study to investigate ERPs during AV speech perception in older adults.

The concept or principle of multisensory efficiency recurred in the second study. Multisensory efficiency resembles the principle of neural efficiency (for review see: Neubauer & Fink, 2009). Neural efficiency essentially means that fewer neural resources, for example in form of smaller brain areas and/or reduced activity, are recruited in order to handle specific task demands. For example, intelligence was found to correlate negatively with brain activity (Haier, Siegel, Nuechterlein, & Hazlett, 1988) Since then various neuroimaging studies have been able to extend the concept of neural efficiency to other cognitive tasks such as working memory and executive functioning (Neubauer & Fink, 2009).

Related to the principle of neural efficiency the results of the experiments conducted for this dissertation indicated multisensory efficiency. For instance, in the second study it was shown that behaviour improved significantly (i.e., higher accuracy and faster response times) even though the auditory system recruited fewer neural resources (i.e., smaller N1 amplitude) and actually processed the auditory information faster which was seen in the N1 peak latency shift for AV_{speech} trials. This multisensory efficiency could have important implications for speech perception in younger and older adults but particularly for the latter group.

Working memory (WM) capacity in OA declines with increasing age (Park et al., 2002; Waters & Caplan, 2005) which in turn can negatively influence auditory speech perception and comprehension. WM is important for speech comprehension (Caplan & Waters, 1999; Just & Carpenter, 1992), because in order to understand a sentence at the end, it is necessary to keep track of the sequence of spoken words, to store what was being said in the beginning and to integrate the information.

One explanation for this WM deficit in OA is based on the limited resource hypothesis (Just & Carpenter, 1992; Kahneman, 1973; Rabbitt, 1968). According to this hypothesis all cognitive and sensory processes access a common pool of 'mental resources'. Age related sensory difficulties are compensated for by allocating more resources to these processes which in turn take away valuable resources that are required for successful higher-order functions such as WM (Hasher & Zacks, 1979; Just & Carpenter, 1992; Schneider & Pichora-Fuller, 2000).

With respect to the results of increased multisensory efficiency, I hypothesize that the reduced sensory demand during AV speech requires fewer resources than during unimodal speech perception, which means that more resources could be assigned to higher level cognition. Speech perception under adverse conditions (i.e., hearing deficit or background noise) taxes the auditory system which means that more resources have to be devoted to basic signal encoding. It has been shown that for example memory is worse under noisy conditions (Cervera et al., 2009; Pichora-Fuller et al., 1995; Rabbitt, 1968; Tun et al., 2009; Yampolsky et al., 2002). However, there is evidence that if sensory encoding can be improved, for example by making efficient use of the context of a discourse or a sentence, then not only perception improves but also working memory (Pichora-Fuller et al., 1995). This shows an intrinsic link between sensory functioning and higher level cognition (Schneider & Pichora-Fuller, 2000; Tun et al., 2009; Wingfield et al., 2005).

If this idea is applied to this dissertation and in particular to the second study, the visual speech cues during AV speech perception could alleviate sensory processing load and boost cognitive functions. Support for this idea comes from an AV study involving young participants whose task was to memorize final words of sentences (Pichora-Fuller, 1996). Sentences were presented in a multi-talker babble background and memory performance improved significantly relative to an auditory speech condition, when visible speech cues were presented as well. Possibly, the additional visual speech made auditory processing more efficient and took away perceptual stress from the auditory modality. In turn, this led to a surplus of neural resources used to memorize the words. This explanation is similar to a study that showed that the use of semantic context improved auditory speech comprehension and memory (Pichora-Fuller et al., 1995). Due to sensory aging (e.g., presbycusis) older adults experience permanently noisy sensory channels (i.e., sensory-perceptual stress) (Erber, 2002; Schneider & Pichora-Fuller, 2000) and the addition of visual speech cues could enhance speech perception similar to what was seen in the younger adults in the study by Pichora-Fuller and colleagues (1996).

Research that has looked at the effect of multisensory stimuli on WM has shown that WM improves under bimodal stimulation (Foos & Goolkasian, 2005; Goolkasian & Foos, 2002, 2005; Mastroberardino et al., 2008). It is not certain though how this benefit is brought about. The findings from this dissertation provide the basis for an explanation in terms of efficient use of neural resources. Preliminary results of ongoing research investigating the relation between WM and AV speech revealed behavioural benefits of

AV over unisensory trials. Analysis of whether this benefit is correlated with electrophysiological modulations is underway (Baranyaoiva et al., submitted).

The results presented in this dissertation consistently showed ERP amplitude reductions in response to AV stimuli. But how is this accomplished? What are the neural mechanisms which allow for and implement multisensory efficiency?

4.2.2 Neural basis of multisensory efficiency

A potential mechanism that achieves the amplitude reduction of the N1 could be like a frequency filter (van Wassenhove et al., 2005). There is no 1:1 match between a viseme and corresponding phoneme but rather one viseme corresponds to a class of phonemes (e.g., /p/, /b/, and /m/), which is the reason why speechreading is so challenging for hearing individuals (Campbell, 2008; Erber, 1974; Summerfield, 1983). In numbers, an estimated 60% of speech sounds are not available via visual speech cues (Easton & Basala, 1982). Nevertheless, a viseme provides some constraint regarding the number of possible phonemes that might enter the auditory system. Grant and Seitz (2000b) analyzed visual speech cues with the corresponding auditory speech signal and demonstrated correlations between lip movements and second and third formant frequencies. A vowel can be characterized by a number of formants which are frequency bands with the most energy. Also, head movement accompanying natural speech has been shown to correlate with the fundamental or first formant frequency of the speech signal (Munhall et al., 2004). It could be speculated, that visual speech cues provide information about upcoming auditory frequency. The auditory system could make use of this information by lowering the activity level of those cells sensitive to the frequency bands predicted by the viseme in order to avoid redundant information processing.

Essentially this translates to 'listening with your eyes'. As a result, fewer neurons would respond to the speech sound than if no constraint were placed on the auditory system as when only listening to someone speaking. If fewer neurons were active this should reduce the overall level of neural activity and, by extension, the amplitude of ERP components like the N1. Intracranial recordings in epilepsy patients provided direct evidence that the activity level of neurons in the auditory cortex was reduced in response to AV speech stimuli (i.e., syllables) relative to auditory-only speech trials (Besle et al., 2008; Reale et al., 2007).

Research on neuronal energy consumption in rats has shown that action potentials account for 50% of the total energy consumption (Niven & Laughlin, 2008). This high energy cost is mainly due to activity of the sodium-potassium pump responsible for maintaining and re-establishing the resting potential of the neuron. Therefore, decreasing the amount of neuronal activity leads to less energy expended. If this reduction goes together with equivalent or even superior behaviour such a process can be considered to operate efficiently. Energy efficiency of the sensory systems is determined by the ratio of information encoded to energy expended (Niven & Laughlin, 2008). Sparse coding is one way to make signal transfer more efficient. Sparse coding refers to a principle in which only a small group of neurons represent information rather than a large neuronal population (Laughlin, 2001) and the existence of sparse coding has been shown to be present in the auditory cortex of rats (Hromádka, DeWeese, & Zador, 2008). According to Niven and Laughlin (2008) a sensory signal processor can be more efficient if it can reduce redundant signal processing. As mentioned above, visual speech and auditory speech signals carry partially redundant information. The ability to filter out these redundant signals might be a strategy used by the auditory system to operate more efficiently during AV speech perception.

4.2.3 Older adults use lip cues more effectively

Results of the second study revealed, in addition to an auditory N1 amplitude reduction, an N1 peak latency shift during AV trials relative to auditory-only trials. These electrophysiological AV modulations are similar to those reported by van Wassenhove and colleagues (2005). Importantly, the experiment in the second study used a total of 80 different words whereas previous ERP studies on AV speech used just a few syllables. Therefore, results of this experiment can be seen as an extension to previous findings and as an indication of their generalization, but they also provided information regarding age-related differences in electrophysiological processing of AV speech perception.

The findings of the second study line up nicely with an analysis-by-synthesis model (van Wassenhove et al., 2005). The visual speech cues that preceded the auditory speech signal aided auditory processing. To some degree the benefit may be derived from attentional cueing but visual speech provides complementary information that is not readily available to the auditory modality, such as place of articulation (Campbell, 2008; Summerfield, 1983). Wassenhove and colleagues (2005) demonstrated that this speeding of the ERP component could be related to predictability. That is, the more predictable the visual speech cue is of the upcoming auditory speech stimulus the larger the N1 latency shift in AV trials. The fact that older adults showed more pronounced latency shifts during AV_{speech} trials as compared to younger adults suggests that older adults made better use of the predictive value of the visemes. The poor lip reading performance during the V-only condition illustrates that older adults were not better speechreaders but given the larger N1 latency shift, older adults seemed to be better lip-speech integrators. Such a latency shift means that auditory processing is faster under AV conditions for older adults.

Another difference between younger and older adults emerged with respect to the use of visual speech cues. The inclusion of the AV_{photo} condition enabled the assessment of whether seeing a static face alongside auditory speech is sufficient to achieve AV speech benefits. For older adults, no AV_{photo} benefit emerged. However, younger adults showed AV benefits in terms of faster reaction times. This dissociation suggests that older adults required dynamic visual speech cues to boost performance more so than younger adults. One could speculate that it had to do with cognitive resource conservation. Younger adults might have used the still face as a cue to raise their level of attention globally. Older adults on the other hand might have been more conservative and only increased their attention when a highly predictive cue was available as during AV_{speech} trials. This more cautious use of precious resources can be linked back to the limited resource hypothesis mentioned earlier (Just & Carpenter, 1992). One possible reason for why older adults are faced with reduced resources for cognitive processes is an enhanced demand of resources for sensory signal processing. Age-related changes to the sensory systems reduce the effectiveness of sensory signal processing (Bergman & Rosenhall, 2001; Divenyi et al., 2005; Erber, 2002; Schieber, 2006; Schneider & Pichora-Fuller, 2000; Wingfield et al., 2005). However, reduced sensory effectiveness could open the door for the principle of inverse effectiveness during multisensory perception.

4.2.4 Sensory functioning and inverse effectiveness

The idea of inverse effectiveness is prominent in the literature on multisensory interaction. It states that the multisensory response is largest, the less effective the unisensory stimuli are by themselves (Stein & Meredith, 1993). If this principle is translated to older adults it would be predicted that older adults should benefit more from AV stimuli than younger adults due to sensory aging. Previous studies have provided support for this claim (Hugenschmidt et al., 2009; Laurienti et al., 2006). Although, recent data from an ongoing project in our laboratory on AV speech comprehension in noise showed larger multisensory benefits in younger than in older adults, but nevertheless, older adults showed significant AV enhancements as well (N. A. Phillips et al., 2009). Findings of the second study can accommodate the inverse effectiveness hypothesis with respect to older adults. Compared to younger adults, older adults were less skilled in the V-only (i.e., speechreading) condition, but groups were indistinguishable from each other in terms of accurate responses to AV_{speech} trials. This finding was reflected in the significantly higher auditory enhancement scores for older adults. Auditory enhancement values reflect the amount of benefit that is gained from adding auditory signals to the V-only baseline performance (Sommers et al., 2005). Support for the inverse effectiveness hypothesis was also provided by the auditory N1 latency shifts which were larger in older than in younger adults. Conducting a multiple regression analysis revealed that hearing level thresholds, a measure of auditory functioning, predicted the size of the N1 latency shift. Auditory functioning was worse in older adults (i.e., elevated hearing thresholds), yet the benefit in terms of speeded auditory processing was larger for older adults when hearing and seeing a person speak relative to just hearing someone speak. This can be interpreted in a theoretical framework as inverse effectiveness or, in practical terms, as compensation for sensory deficits. Possibly older adults made more efficient use of available speech cues to compensate for decreased auditory functions, which would in turn decrease perceptual processing load.

It should be noted though, that the relation between auditory functioning and increase in auditory processing speed under AV_{speech} conditions remained after controlling for the independent variable of age. In other words, this relation did not only apply to older adults but also to younger adults with elevated hearing thresholds (Figure 16). However, given that hearing thresholds were on average significantly higher in older adults, the N1 latency shift was more pronounced for the group of older adults.

The fact that the relation between sensory functioning and neural responses remained after controlling for age highlights the importance of taking sensory functioning into account when designing a study on multisensory perception. Studies on AV processing involving older adults usually assess and control for sensory intactness, but, the current data clearly show that sensory functioning is not just an issue for older adults. Therefore, sensory testing should be common practice in multisensory interaction studies even if the cohort consists of only younger adults. In addition to this methodological implication, results of this dissertation are of relevance to some theoretical and practical issues.

4.3 Theoretical and Practical Implications

Results from all three experiments provided additional support to the possibility that one sensory modality can influence processes in another, and that these interactions are likely to take place in areas that have traditionally been considered as unisensory or sensory-specific. With respect to Fodor's *"Modularity of the Mind"* (1983) these findings indicate that our modalities, or at least vision and audition, might not be independent modules that are domain specific and, depending on the definition, are not or are only partially informationally encapsulated. Given the multisensory nature of the primate cortex Ghazanfar and Schroeder (2006, p. 278) even proposed "...to abandon the notion that the senses ever operate independently during real-world cognition".

This dissertation adds to the growing body of evidence that multisensory perception is associated with superior performance over unisensory perception. If the speculation is true that multisensory efficiency leads to more resources available for higher level cognition, it could have important implications for learning and teaching. Studies on perceptual learning have shown that participants learned faster during multimodal than unisensory condition (Seitz, Kim, & Shams, 2006; Shams & Seitz, 2008). Learning and recognizing someone's voice has also been shown to benefit when during the learning phase the voice is paired with the corresponding face (von Kriegstein & Giraud, 2006). Functional imaging data revealed that the face-selective area of the fusiform gyrus was activated upon hearing a familiar voice, indicating a functional coupling between voice and face areas (von Kriegstein & Giraud, 2006; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). Despite the emerging evidence in support of better learning through multisensory stimulation and association, more research is needed to empirically solidify the effects of multisensory learning and teaching (Shams & Seitz, 2008). Particularly the question whether multisensory learning is beneficial for more complex processes (e.g., second language acquisition) needs to be carefully explored. There is some evidence that adding visual speech to a speech sound in a second language improves phonemic identification (Navarra & Soto-Faraco, 2007), yet whether this applies to sentences or entire conversations needs to be explored. Assuming that second language comprehension is effortful, the principle of multisensory efficiency might also apply to perception in the non-native language. That is, complementary visual speech cues would reduce signal processing demands so that more resources can be devoted to higher level cognition. Other areas where audiovisual signal processing might be applicable are public announcements via video screens in noisy environments like subways or airports.

4.4 Future Directions

Based on other neuroimaging studies, multisensory interaction is not restricted to sensory specific areas but likely requires a complex network of brain areas. These networks are likely to consist of dynamic feedback and feedforward connections. Also, intersensory connections seem to be bidirectional given that findings from the AV non-speech study indicated modulations of vision through audition and the opposite seemed to be the case for AV speech perception. Further studies are needed to establish which areas are involved, which functions they have, in which sequence they are activated and how they are connected.

The findings of the second study are very promising as they indicate that efficient use of visual speech cues might help to reduce hearing deficits experienced by older adults. Future studies should address whether training programs on speechreading lead to improved speech comprehension during face-to-face conversations. The ability to communicate effectively and effortlessly is closely linked to one's perceived quality of life (Erber, 2002). For example, caregivers of patients with Alzheimer's disease report difficulties in effectively communicating with the patients which puts more strain on the relationship (Orange, 2001; Orange & Colton-Hudson, 1998; Orange, Lubinski, & Higginbotham, 1996; Richter, Roberto, & Bottenberg, 1995). It has been suggested that one effective strategy to improve communication between caregiver and patient is to maintain eye-contact (Small, Gutman, Makela, & Hillhouse, 2003). A recent empirical investigation of AV speech and dementia indicated that speech perception in patients with Alzheimer's disease improved significantly when auditory speech was complemented by visual speech cues (Phillips, Baum, & Taler, 2009). This strengthens the notion that facing the conversation partner improves communication with patients in particular, but also with healthy individuals in general.

Chapter 5: Conclusion

The first question this dissertation addressed was whether processes for AV speech and AV non-speech perception are the same or different. According to the results from the first two studies, multisensory interaction modulated different processes during recognition of AV speech stimuli than AV non-speech stimuli. More precisely, multisensory stimuli seemed to affect the processes of the dominant modality for a given task; vision for non-speech and audition for speech perception. Even though different processes were modulated, multisensory interaction manifested itself in form of multisensory efficiency for both classes of stimuli. That is, AV conditions (speech and non-speech) led to behavioural improvements even though fewer neural resources were recruited. One potential explanation for this effect is that complementary information provided by signals in the non-dominant modality (i.e., audition for non-speech and vision for speech perception) constrained signal processing in the dominant modality for a given task. Given that AV trials modulated early, sensory specific ERP components, the results add further support, albeit indirectly, to the notion that multisensory interactions take place in sensory-specific cortices. In addition to early interactions, signals from different modalities are likely to interact at later stages in the information processing stream as well.

Early sensory-specific AV modulations as well as the principle of multisensory efficiency were also evident in younger and older adults in the second study, which investigated the question whether AV speech is processed differently in younger and older adults. The answer is that there are strong indications that older adults made better or more effective use of visual speech cues than younger adults. This age-related benefit is interpreted in terms of compensation for sensory aging. The finding that older adults made effective use of visual speech cues could provide a relatively easy (and cost-effective) way to cope with age-related hearing deficits. Improvement in hearing would lead to better and less effortful communication which in turn would lead to an improvement in the experienced quality of life.

Even though our sensory modalities have highly specialized receptors sensitive to a particular type of signal, the results of all three experiments presented here, add to the increasing amount of data suggesting that our senses do not operate independently. Not only are there signs of multisensory interaction, but the findings shown here, in addition to findings reported in the literature, demonstrate that one modality seems to be able to influence early, sensory-specific processes of another. More research is needed to solidify this notion but if true, it would have important implications for our understanding and conceptualization of sensory processing and perceptual mechanisms in the brain.

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*(3), 131-138.

Alho, K., Woods, D. L., & Algazi, A. (1994). Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiology, 31*(5), 469-479.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research, 166*(3-4), 559-571.

Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: An event-related brain potential study. *Psychophysiology, 32*, 177-190.

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

Antinoro, F., Skinner, P. H., & Jones, J. J. (1969). Relation between sound intensity and amplitude of the AER at different stimulus frequencies. *J Acoust Soc Am, 46*(6), 1433-1436.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28*(2), 105-167.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339-355.

Baranyaoiva, J., Winneke, A. H., & Phillips, N. A. (submitted). *The effect of auditory-visual stimuli on working memory: An ERP study.* Paper presented at the Cognitive Aging, Atlanta, GA.

Barth, D. S., Goldberg, N., Brett, B., & Di, S. (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Research, 678*(1-2), 177-190.

Beagley, H. A., & Knight, J. J. (1967). Changes in auditory evoked response with intensity. *J Laryngol Otol, 81*(8), 861-873.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41*(5), 809-823.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Bergeson, 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.

Bergman, B., & Rosenhall, U. (2001). Vision and hearing in old age. *Scandinavian Audiology, 30*(4), 255-263.

Besle, J., Bertrand, O., & Giard, M. H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hearing Research*, in 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*(52), 14301-14310.

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

Besle, J., Fort, A., & Giard, M. H. (2005). Is the auditory sensory memory sensitive to visual information? *Experimental Brain Research, 166*(3-4), 337-344.

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

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebral Cortex, 17*(9), 2172-2189.

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

Boothroyd, A. (1988). Linguistic factors in speechreading. *Volta Review, 90*, 77-87.

Bucur, B., Allen, P. A., Sanders, R. E., Ruthruff, E., & Murphy, M. D. (2005). Redundancy gain and coactivation in bimodal detection: Evidence for the preservation of coactive processing in older adults. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 60*(5), 279-282.

Bucur, B., Madden, D. J., & Allen, P. A. (2005). Age-related differences in the processing of redundant visual dimensions. *Psychology and Aging, 20*(3), 435-446.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience, 143*(4), 1065-1083.

Budinger, E., Laszcz, A., Lison, H., Scheich, H., & Ohl, F. W. (2008). Non-sensory cortical and subcortical connections of the primary auditory cortex in Mongolian gerbils: bottom-up and top-down processing of neuronal information via field AI. *Brain Research, 1220*, 2-32.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience, 6*(2), 190-195.

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M.-J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience, 2*(8), 759-766.

Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport, 14*(17), 2213-2218.

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience, 16*(5), 805-816.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110-1123.

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., Brammer, M. J., & Iversen, S. D. (1998). Crossmodal Identification. *Trends in Cognitive Sciences, 2*(7), 247-253.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*(5312), 593-596.

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *Journal of Cognitive Neuroscience, 15*(1), 57-70.

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.

Calvert, G. A., & Lewis, J. W. (2004). Hemodynamic studies of audiovisual interactions. In G. A. Calvert, C. Spence & B. E. Stein (Eds.), *The Handbook of Multisensory Processes* (pp. 483-502). Cambridge, Ma: The MIT Press.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal de Physiologie-Paris, 98*(1-3), 191-205.

Campbell, R. (1992). The neuropsychology of lipreading. *Philosophical Transactions of the Royal Society of Londond Series B-Biological Sciences, 335*(1273), 39-44; discussion 44-35.

Campbell, R. (2006). Audiovisual speech processing. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistics* (2 ed., pp. 562-569). London, UK.: Elsevier.

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

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading. A neurological dissociation. *Brain, 109 (Pt 3)*, 509-521.

Campbell, R., Zihl, J., Massaro, D., Munhall, K., & Cohen, M. M. (1997). Speechreading in the akinetopsic patient, L.M. *Brain: A Journal of Neurology, 120*(10), 1793-1803.

Campi, K. L., Bales, K. L., Grunewald, R., & Krubitzer, L. (2009). Connections of auditory and visual cortex in the Prairie Vole (Microtus ochrogaster): Evidence for multisensory processing in primary sensory areas. *Cerebral Cortex*, in press.

Capek, C. M., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P., & Neville, H. J. (2004). The cortical organization of audio-visual sentence comprehension: an fMRI study at 4 Tesla. *Cognitive Brain Research, 20*(2), 111-119.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22*(1), 77-94; discussion 95-126.

Cappe, C. l., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience, 22*(11), 2886-2902.

Cervera, T. C., Soler, M. J., Dasi, C., & Ruiz, J. C. (2009). Speech recognition and working memory capacity in young-elderly listeners: effects of hearing sensitivity. *Canadian Journal of Experimental Psychology, 63*(3), 216-226.

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

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

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping, 2*(3), 170-187.

Clarke, S., Bellmann, A., de Ribaupierre, F., & Assal, G. (1996). Non-verbal auditory recognition in normal subjects and brain-damaged patients: Evidence for parallel processing. *Neuropsychologia, 34*(6), 587-603.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*(2), 409-412.

Coles, M. G. H., & Rugg, M. D. (1995). ERPs: An introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition* (pp. 1-26). Oxford, U.K.: Oxford University Press.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology, 113*(4), 495-506.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*(3), 256-266.

Craik, F. I. M., & Salthouse, T. A. (2000). *Handbook of Aging and Cognition* (2 ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Davis, C., & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition, 100*(3), B21-BB31.

Davis, H., Osterhammel, P. A., Wier, C. C., & Gjerdingen, D. B. (1972). Slow vertex potentials: interactions among auditory, tactile, electric and visual stimuli. *Electroencephalography and Clinical Neurophysiology, 33*(6), 537-545.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*(10), 460-467.

Di Russo, F., Martinez, A., & Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex, 13*(5), 486-499.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2001). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*(2), 95-111.

Divenyi, P. L., Stark, P. B., & Haupt, K. M. (2005). Decline of speech understanding and auditory thresholds in the elderly. *Journal of the Acoustical Society of America, 118*(2), 1089-1100.

Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology, 10*, R731-R735.

Easton, R. D., & Basala, M. (1982). Perceptual dominance during lipreading. *Perception & Psychophysics, 32*(6), 562-570.

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

Eimer, M. (2000). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Cognitive Brain Research, 10*(1-2), 145-158.

Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: Evidence from event-related brain potential studies. *Neuroscience & Biobehavioral Reviews, 25*(6), 497-511.

Eimer, M., & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology, 35*(3), 313-327.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research, 12*(2), 423-425.

Erber, N. P. (1974). Visual perception of speech by deaf children: Recent developments and continuing needs. *Journal of Speech & Hearing Disorders, 39*(2), 178-185.

Erber, N. P. (2002). *Hearing, Vision, Communication and Older People*. Clifton Hill, Vic., Australia: Clavis Publishing.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*(13), 5749-5759.

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Foos, P. W., & Goolkasian, P. (2005). Presentation format effects in working memory: the role of attention. *Memory and Cognition, 33*(3), 499-513.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research, 14*(1), 20-30.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). *Cognitive Neuroscience: The Biology of Mind* (2 ed.). New York, NY: W W Norton & Company.

Ghazanfar, A. A., & Logothetis, N. K. (2003). Neuroperception: facial expressions linked to monkey calls. *Nature, 423*(6943), 937-938.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience, 25*(20), 5004-5012.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences, 10*(6), 278-285.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490.

Giard, M. H., & Perronet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience, 11*(5), 473-490.

Goldstein, E. (2007). *Sensation & Perception* (7 ed.). Belmont, CA: Thomson Wadsworth.

Goolkasian, P., & Foos, P. W. (2002). Presentation format and its effect on working memory. *Memory and Cognition, 30*(7), 1096-1105.

Goolkasian, P., & Foos, P. W. (2005). Bimodal format effects in working memory. *American Journal of Psychology, 118*(1), 61-77.

Government of Canada. (2002). *Canada's Aging Population.* Retrieved. from http://www.hc-sc.gc.ca/seniors-aines.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America, 104*(4), 2438-2450.

Grant, K. W., & Seitz, P. F. (2000a). The recognition of isolated words and words in sentences: individual variability in the use of sentence context. *Journal of the Acoustical Society of America, 107*(2), 1000-1011.

Grant, K. W., & Seitz, P. F. (2000b). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108*(3 Pt 1), 1197-1208.

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 the Acoustical Society of America, 103*(5), 2677-2690.

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(24), 95-112.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*(2), 240-244.

Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech, 26*(2), 117-129.

Haier, R. J., Siegel, B. V., Nuechterlein, K. H., & Hazlett, E. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence, 12*(2), 199-217.

Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., & Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Experimental Brain Research, 152*(3), 404-408.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*(3), 356-388.

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

Hay, I. S., & Davis, H. (1971). Slow cortical evoked potentials: interactions of auditory, vibro-tactile and shock stimuli. *Audiology, 10*(1), 9-17.

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*(6), 2739-2745.

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech, Language and Hearing Research, 40*(2), 432-443.

Helfer, K. S. (1998). Auditory and auditory-visual recognition of clear and conversational speech by older adults. *Journal of the American Academy of Audiology, 9*(3), 234-242.

Hertrich, I., Mathiak, K., Lutzenberger, W., & Ackermann, H. (2009). Time course of early audiovisual interactions during speech and nonspeech central auditory processing: a magnetoencephalography study. *Journal of Cognitive Neuroscience, 21*(2), 259-274.

Hertrich, I., Mathiak, K., Lutzenberger, W., Menning, H., & Ackermann, H. (2007). Sequential audiovisual interactions during speech perception: a whole-head MEG study. *Neuropsychologia, 45*(6), 1342-1354.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 781-787.

Hillyard, S. A., Mangun, G. R., Luck, S. J., & Heinze, H. J. (1990). Electrophysiology of visual attention. In E. R. John, T. Harmony, L. S. Prichep, M. Valdes-Sosa & P. Valdes-Sosa (Eds.), *Machinery of the mind* (pp. 186-205). Boston: Birkhauser.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24*(2), 259-276.

Hopf, J. M., Vogel, E., Woodman, G., Heinze, H. J., & Luck, S. J. (2002). Localizing visual discrimination processes in time and space. *Journal of Neurophysiology, 88*(4), 2088-2095.

Hromádka, T., DeWeese, M. R., & Zador, A. M. (2008). Sparse Representation of Sounds in the Unanesthetized Auditory Cortex. *PLoS Biology, 6*(1), e16.

Hugenschmidt, C. E., Mozolic, J. L., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport, 20*(4), 349-353.

Hummert, M. L., & Nussbaum, J. F. (2001). *Aging, Communication, and Health: Linking Research and Practice for Successful Aging*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hyde, M. (1997). The N1 response and its applications. *Audiology and Neuro-otology, 2*(5), 281-307.

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

Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience, 12 Suppl 2*, 35-51.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America, 96*(16), 9379-9384.

Jagger, C., Spiers, N., & Arthur, A. (2005). The role of sensory and cognitive function in the onset of activity restriction in older people. *Disability and Rehabilitation: An International Multidisciplinary Journal, 27*(5), 277-283.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology, 85*(2), 506-522.

Jousmaki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Current Biology, 8*(6), R190.

Jung, R., & Berger, W. (1979). [Fiftieth anniversary of Hans Berger's publication of the electroencephalogram. His first records in 1924--1931 (author's transl)]. *Archiv für Psychiatrische Nervenkrankheiten, 227*(4), 279-300.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kahneman, D. (1973). *Attention and effort*. New York: Prentice-Hall.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: gamma-band activity during audiovisual speech perception. *Cerebral Cortex, 15*(5), 646-653.

Kang, E., Lee, D. S., Kang, H., Hwang, C. H., Oh, S. H., Kim, C. S., et al. (2006). The neural correlates of cross-modal interaction in speech perception during a semantic decision task on sentences: a PET study. *Neuroimage, 32*(1), 423-431.

Kawashima, R., Imaizumi, S., Mori, K., Okada, K., Goto, R., Kiritani, S., et al. (1999). Selective visual and auditory attention toward utterances-a PET study. *Neuroimage, 10*(2), 209-215.

Kim, S. H., Frisina, R. D., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2006). Effect of age on binaural speech intelligibility in normal hearing adults. *Speech Communication, 48*(6), 591-597.

Kislyuk, D. S., Mottonen, R., & Sams, M. (2008). Visual processing affects the neural basis of auditory discrimination. *Journal of Cognitive Neuroscience, 20*(12), 2175-2184.

Koppen, C., Alsius, A., & Spence, C. (2008). Semantic congruency and the Colavita visual dominance effect. *Experimental Brain Research, 184*(4), 533-546.

Koppen, C., & Spence, C. (2007a). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research, 1186*, 224-232.

Koppen, C., & Spence, C. (2007b). Seeing the light: exploring the Colavita visual dominance effect. *Experimental Brain Research, 180*(4), 737-754.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 11*, 203-205.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83-143). San Diego: Academic Press. Inc.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics Electrified II. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of Psychlinguistics* (2 ed.). New York: Elsevier Press.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology, 11*(4), 475-480.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27*(8), 1155-1163.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience, 14*(3), 420-429.

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research, 166*(3-4), 289-297.

Lebib, R., Papo, D., Douiri, A., de Bode, S., Gillon Dowens, M., & Baudonniere, P. M. (2004). Modulations of 'late' event-related brain potentials in humans by dynamic audiovisual speech stimuli. *Neuroscience Letters, 372*(1-2), 74-79.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human Brain Regions Involved in Recognizing Environmental Sounds. *Cerebral Cortex, 14*(9), 1008-1021.

Li, K. Z. H., & Lindenberger, U. (2002). Connections among sensory, sensorimotor, and cognitive aging: Review of data and theories. *Neuroscience and Biobehavioral Reviews, 26*, 777-783.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Lobaugh, N. J., Chevalier, H., Batty, M., & Taylor, M. J. (2005). Accelerated and amplified neural responses in visual discrimination: Two features are processed faster than one. *Neuroimage, 26*(4), 986-995.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

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

Luck, S. J., Heinze, H. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and Clinical Neurophysiology, 75*(6), 528-542.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*, 291-308.

Luck, S. J., & Hillyard, S. A. (1995). The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. *International Journal of Neuroscience, 80*(1-4), 281-297.

Macaluso, E. (2006). Multisensory processing in sensory-specific cortical areas. *Neuroscientist, 14*(4), 327-338.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in Neurosciences, 28*(5), 264-271.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage, 21*(2), 725-732.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*(3), 253-257.

MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., et al. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage, 22*(4), 1605-1618.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology, 32*(1), 4-18.

Mangun, G. R., Hillyard, S. A., & Luck, S. J. (1993). Electrocortical substrates of visual selective attention. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 219-243). Cambridge, Ma.: MIT Press.

Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., & Belardinelli, M. O. (2008). How the bimodal format of presentation affects working memory: An overview. *Cognitive Processing, 9*(1), 69-76.

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology, 58*(1), 22-33.

McDowd, J. M., & Shaw, R. J. (2000). Attention and aging: a functional perspective. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of Aging and Cognition* (2 ed., pp. 221-292). Mahwah, NJ: Lawrence Erlbaum Associates.

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

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*(1), 53-65.

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research, 14*, 31-40.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389-391.

Meredith, M. A., & Stein, B. E. (1985). Descending efferents from the superior colliculus relay integrated multisensory information. *Science, 227*(4687), 657-659.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*(3), 640-662.

Mierzejewski, K. (2009). How to pick a ripe watermelon Retrieved Oct. 15th, 2009, from http://www.gardeningknowhow.com/fruit-gardening/pick-a-watermelon.htm

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247-279.

Möbes, J., Lambrecht, J., Nager, W., Büchner, A., Lesinski-Schiedat, A., Lenarz, T., et al. (2006). Die audiovisuelle Sprachverarbeitung bei Patienten mit Cochlea-Implantat. *Zeitschrift für Neuropsychologie, 17*(1), 25-34.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. *Cerebral Cortex, 14*(4), 452-465.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research, 14*, 115-128.

Morrot, G., Brochet, F., & Dubourdieu, D. (2001). The color of odors. *Brain and Language, 79*(2), 309-320.

Möttönen, R., Schürmann, M., & Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. *Neuroscience Letters, 363*(2), 112-115.

Munhall, K., & 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.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Research Report Visual Prosody and Speech Intelligibility Head Movement Improves Auditory Speech Perception. *Psychological Science, 15*(2), 133-137.

Murray, M. M., Wylie, G. R., Higgins, B. A., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). The spatiotemporal dynamics of illusory contour processing: combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *Journal of Neuroscience, 22*(12), 5055-5073.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of component structure. *Psychophysiology, 24*, 375-425.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences Review, 24*(5), 283-288.

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*(4), 695-699.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research/Psychologische Forschung, 71*(1), 4-12.

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience and Biobehavioral Reviews, 33*(7), 1004-1023.

Niven, J. E., & Laughlin, S. B. (2008). Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology, 211*(Pt 11), 1792-1804.

Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage, 25*(2), 333-338.

Orange, J. B. (2001). Family caregivers, communcation, and Alzheimer's disease. In M. L. Hummert & J. F. Nussbaum (Eds.), *Aging, Communication, and Health: Linking Research and Practice for Successful Aging* (pp. 225-248). Mahwah, NJ: Lawrence Erlbaum Associates.

Orange, J. B., & Colton-Hudson, A. (1998). Enhancing communication in dementia of the alzheimer's type. *Topics in Geriatric Rehabilitation, 14*(2), 56-75.

Orange, J. B., Lubinski, R. B., & Higginbotham, D. J. (1996). Conversational repair by individuals with dementia of the Alzheimer's type. *Journal of Speech & Hearing Research, 39*(4), 881-895.

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.

Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., et al. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport, 16*(2), 125-128.

Pérez, E., Meyer, G., & Harrison, N. (2008). Neural correlates of attending speech and non-speech: ERPs associated with duplex perception. *Journal of Neurolinguistics, 21*(5), 452-471.

Phillips, N. A., Baum, S., & Taler, V. (2009). *Audio-visual speech perception and grammatical prosody perception in healthy older adults, mild cognitive impairment (MCI) and Alzheimer disease (AD).* Paper presented at the International Conference on Auditory-Visual Speech Processing (AVSP), Norwich, U.K.

Phillips, N. A., Gagné, J.-P., Basu, M., Copeland, L., Gosselin, P., Saint-Pierre, A., et al. (2009). *Audio-visual speech perception in younger and older adults: Effects on word identification and memory.* Paper presented at the 10th Annual meeting of the International Multisensory Research Forum (IMRF), New York City.

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., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*(1), 593-608.

Pichora-Fuller, M. K., Schneider, B. A., Macdonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: a simulation of auditory aging. *Hearing Research, 223*(1-2), 114-121.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neuro-otology, 4*(2), 64-79.

Picton, T. W., Woods, D. L., Baribeau-Braun, J., & Healey, T. M. (1976). Evoked potential audiometry. *J Otolaryngol, 6*(2), 90-119.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language and Hearing Research, 52*(4), 1073-1081.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83*(2), 157-171.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20*(3), 241-248.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A, 97*(22), 11800-11806.

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*(1), 162-184.

Reisberg, D., McLean, J., Goldfield, A., Dodd, B., & Campbell, R. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In *Hearing by eye: The psychology of lip-reading.* (pp. 97-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Richter, J. M., Roberto, K. A., & Bottenberg, D. J. (1995). Communicating with persons with Alzheimer's disease: Experiences of family and formal caregivers. *Archives of Psychiatric Nursing, 9*(5), 279-285.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188-194.

Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience, 27*, 169-192.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience, 2*(12), 1131-1136.

Rönnberg, J., & Lyxell, B. (1998). Conceptual constraints in sentence-based lipreading in the hearing-impaired. In R. Campbell, B. Dodd & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech.* (pp. 143-153). Hove England: Psychology Press/Erlbaum (UK) Taylor & Francis.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*(5), 1147-1153.

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage, 20*(3), 1609-1624.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America, 104*(17), 7295-7300.

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia, 45*(3), 587-597.

Saito, D. N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. (2005). Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. *Cerebral Cortex, 15*(11), 1750-1760.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., et al. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*(1), 141-145.

Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T., & Munhall, K. (2003). Perceiving Biological Motion: Dissociating Visible Speech from Walking. *Journal of Cognitive Neuroscience, 15*(6), 800-809.

Schieber, F. (2006). Vision and aging. In J. E. Birren & K. Warner Schaie (Eds.), *Handbook of the Psychology of Aging* (6 ed., pp. 129-154). London, UK: Academic Press.

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

Schreiner, C. E. (1998). Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. *Audiol Neurootol, 3*(2-3), 104-122.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology, 50*(1-2), 5-17.

Schwartz, J.-L., Berthommier, F. d. r., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition, 93*(2), B69-BB78.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences, 26*(2), 100-107.

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology, 16*(14), 1422-1427.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport, 16*(16), 1781-1785.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411-417.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics, 69*(5), 673-686.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience, 20*(11), 2037-2057.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage, 25*(1), 76-89.

Small, J. A., Gutman, G., Makela, S., & Hillhouse, B. (2003). Effectiveness of communication strategies used by caregivers of persons with Alzheimer's disease during activities of daily living. *Journal of Speech, Language, and Hearing Research, 46*(2), 353-367.

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

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience, 25*(28), 6499-6508.

Statistics Canada. (2005). Population projections for Canada. Retrieved October 19th, 2009, from http://www.statcan.gc.ca/daily-quotidien/051215/dq051215b-eng.htm

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: The same rules of multisensory integration apply. *Brain Research, 448*(2), 355-358.

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

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*(12), 1-10.

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

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36*(4-5), 314-331.

Summerfield, Q. (1983). Audio-visual speech perception. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing Science and Hearing Disorders* (pp. 131-182). New York: Academic Press.

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

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

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging, 10*(2), 215-228.

Thompson, L. A., & Malloy, D. (2004). Attention resources and visible speech encoding in older and younger adults. *Experimental Aging Research, 30*(3), 241-252.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522.

Tillberg, I., Ronnberg, J., Svard, I., & Ahlner, B. (1996). Audio-visual speechreading in a group of hearing aid users. The effects of onset age, handicap age, and degree of hearing loss. *Scandinavian Audiology, 25*(4), 267-272.

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging, 24*(3), 761-766.

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

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

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*(2), 157-165.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279-293.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Science, 102* (4), 1181-1186.

Vaughan, H. G., Jr., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology, 28*(4), 360-367.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*(2), 190-203.

von Kriegstein, K., & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*(10), e326.

von Kriegstein, K., Kleinschmidt, A., Sterzer, P., & Giraud, A. L. (2005). Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience, 17*(3), 367-376.

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1993). Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *Journal of Neurophysiology, 69*(6), 1797-1809.

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*(2), 1246-1266.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neuroscience, 9*, 79.

Watanabe, J., & Iwai, E. (1991). Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Research Bulletin, 26*(4), 583-592.

Waters, G., & Caplan, D. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory, 13*(3-4), 403-413.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*(3), 363-375.

Wingfield, A., & Tun, P. A. (2001). Spoken language comprehension in older adults: Interactions between sensory and cognitive change in normal aging. *Seminars in Hearing, 22*(3), 287-301.

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*(3), 144-148.

Yampolsky, S., Waters, G., Caplan, D., Matthies, M., & Chiu, P. (2002). Effects of acoustic degradation on syntactic processing: implications for the nature of the resource system used in language processing. *Brain and Cognition, 48*(2-3), 617-625.

Yin, Q., Qiu, J., Zhang, Q., & Wen, X. (2008). Cognitive conflict in audiovisual integration: an event-related potential study. *Neuroreport, 19*(5), 575-578.

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience, 27*(5), 1090-1096.

Yvert, B., Fischer, C., Bertrand, O., & Pernier, J. (2005). Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. *Neuroimage, 28*(1), 140-153.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *Hanbook of Aging and Cognition* (2 ed., pp. 293-357). Mahwah, NJ: Lawrence Erlbaum Associates.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature, 274*(5670), 423-428.

Zifkin, B. G., & Avanzini, G. (2009). Clinical neurophysiology with special reference to the electroencephalogram. *Epilepsia, 50 Suppl 3*, 30-38.