Contributions of visual speech, visual distractors, and cognition to speech perception in noise for younger and older adults

Julianne M. Beadle
BA (Honours) Psychology

A Thesis Submitted for the Degree of
Doctor of Philosophy

The MARCS Institute for Brain, Behaviour, and Development
Western Sydney University
December 2019

Abstract

Older adults report that understanding speech in noisy situations (e.g., a restaurant) is difficult. Repeated experiences of frustration in noisy situations may cause older adults to withdraw socially, increasing their susceptibility to mental and physical illness. Understanding the factors that contribute to older adults' difficulty in noise, and in turn, what might be able to alleviate this difficulty, is therefore an important area of research. The experiments in this thesis investigated how sensory and cognitive factors, in particular attention, affect older and younger adults' ability to understand speech in noise. First, the performance of older as well as younger adults on a standardised speech perception in noise task and on a series of cognitive and hearing tasks was assessed. A correlational analysis indicated that there was no reliable association between pure-tone audiometry and speech perception in noise performance but that there was some evidence of an association between auditory attention and speech perception in noise performance for older adults.

Next, a series of experiments were conducted that aimed to investigate the role of attention in gaining a visual speech benefit in noise. These auditory-visual experiments were largely motivated by the idea that as the visual speech benefit is the largest benefit available to listeners in noisy situations, any reduction in this benefit, particularly for older adults, could exacerbate difficulties understanding speech in noise. For the first auditory-visual experiments, whether increasing the number of visual distractors displayed affected the visual speech benefit in noise for younger and older adults when the SNR was -6dB (Experiment 1) and when the SNR was -1dB (Experiment 2) was tested. For both SNRs, the magnitude of older adults' visual speech benefit reduced by approximately 50% each time an additional visual distractor was presented. Younger adults showed the same pattern when the SNR was -6dB, but unlike older adults, were able to get a full visual speech benefit when one distractor

was presented and the SNR was -1dB. As discussed in Chapter 3, a possible interpretation of these results is that combining auditory and visual speech requires attentional resources.

To follow up the finding that visual distractors had a detrimental impact on the visual speech benefit, particularly for older adults, the experiment in Chapter 4 tested whether presenting a salient visual cue that indicated the location of the target talker would help older adults get a visual speech benefit. The results showed that older adults did not benefit from the cue, whereas younger adults did. As older adults should have had sufficient time to switch their gaze and/or attention to the location of the target talker, the failure to find a cueing effect suggests that age related declines in inhibition likely affected older adults' ability to ignore the visual distractor.

The final experiment tested whether the visual speech benefit and the visual distraction effect found for older adults in Chapter 4 transferred to a conversation-comprehension style task (i.e., The Question-and-Answer Task). The results showed that younger and older adults' performance improved on an auditory-visual condition in comparison to an auditory-only condition and that this benefit did not reduce when a visual distractor was presented. To explain the absence of a distraction effect, several properties of the visual distractor presented were discussed. Together, the experiments in this thesis suggest that the roles of attention and visual distraction should be considered when trying to understand the communication difficulties that older adults experience in noisy situations.

Acknowledgements

First, I would like to extend my sincerest thanks to all of the older (but young at heart) adults who participated in this research. Your commitment to the project was inspiring and motivating.

To my parents, thank you for always answering the phone and listening, even if the time change meant it was really late (or early) for you. Thank you for being my editors on call, for sending care packages, for coming to visit, for helping me come home to visit and for always supporting my education.

To my supervisors Prof. Chris Davis and Prof. Jeesun Kim, thank you for your continuous support and guidance throughout my candidature. I am grateful that you liked my Skype interview and that we were able to work together on this project.

To Dr. Erin Austen, thank you for sending me the advertisement for this PhD position. Working in the Perception Lab and taking your classes gave me a solid foundation in research that supported me through every stage of my candidature.

To the Multisensory Communication Team: April Ching, Leo Chong, Daniel Hochstrasser, Sonya Karisma, Saya Kawase, Yatin Mahajan, Simone Simonetti, and Mandy Visser. I have learned so much from each of you and I am very grateful for your support.

Thank you to the MARCS technical team members; particularly Johnson Chen, and Ben Binyamin for their assistance throughout my candidature. Last but not least, a special thank you to the HEARing CRC and The MARCS Institute for the financial support that I received during my candidature.

Statement of Authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

[redacted]..........

(Julianne M. Beadle)

Dedication

This thesis is dedicated to Jean Marion Beadle

January 26, 1924 – December 8, 2015

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Motivation

Over the next 30 years, the proportion of the world's population over 60 is expected to nearly double, from 12% to 22% (i.e., approx. 900 million to 2 billion; WHO, 2015). In order to maximize seniors' quality of life, and minimize the financial and often emotional burden of illness on seniors themselves, caregivers, and publicly funded social services such as hospitals, societies should consider prioritizing the health and wellbeing of older adults. According to the World Health Organisation, one functional ability that fosters wellbeing in old age is building and maintaining relationships (WHO, 2015). This recommendation is consistent with research that has identified an association between social engagement and positive health outcomes later in life (Cherry et al., 2011; Gilmore, 2012). For example, in old age, individuals with higher levels of social engagement tend to have better memory (semantic, episodic, and working; James, Wilson, Barnes, & Bennet, 2001; Krueger et al., 2009), mobility (Mendes de Leon, Glass, & Berkman, 2003), and a reduced risk of mental illness such as depression (Hajek et al., 2017). In summary, communicating and connecting with others seems to be an important aspect of healthy ageing.

Speech perception (i.e., hearing, recognising, and understanding what someone is saying) is a critical component of successfully communicating with others. However, speech perception can become less efficient later in life, especially in situations with background noise (Pichora-Fuller, Alain, & Schneider, 2017). Studying the processes involved in speech perception in noise, and how ageing impairs these processes, is an important step in developing strategies to help individuals build and maintain relationships across the lifespan.

The experimental research of this thesis focused on the contributions of visual speech, visual distractors, and cognition to speech perception in noise for younger and older adults. The next section of this introduction summarises the process of speech perception in quiet and in noise. This is followed by a brief history of the field of cognitive hearing science and a discussion of the behavioural methods that have been used to index speech perception within this field. Lastly, the introduction will provide an outline of the experiments included in this thesis.

1.2 Speech Perception in Quiet and in Noise

1.2.1 Auditory Speech Perception

In ideal conditions (i.e., when a listener has no auditory pathology and is in a quiet environment) speech perception occurs relatively effortlessly and automatically. In general, speech perception is the result of bottom up processing (based on the acoustic signal) and top down processing (based on knowledge and meaning of language) working together (Goldstein, 2016). From an auditory perspective, the processing of speech can be understood in terms of the transmission of the acoustic signal from the environment to the brain. First, continuous and spectrally diverse acoustic speech signals travel through the outer, middle and inner ear. Within the inner ear, outer and inner hair cells in the cochlea convert sound vibrations into electrical signals. These signals are transmitted to the auditory nerve, the brain stem, and then to the auditory cortex for further processing. Next, listeners must map acoustic speech signals onto discrete meanings. This involves holding the speech signals in short term memory, recognizing specific phonetic categories (e.g., vowels and consonants), and matching these phonetic categories to higher level language representation (i.e., words and sentences) in sematic long-term memory. Prior knowledge and experience (i.e., top down processing) facilitates the mapping process, allowing listeners to better understand language spoken in their environment (Denes & Pinson, 1993).

However, speech is rarely produced in isolation; rather, day-to-day life requires speech to be understood in the presence of background noise. That is, other talkers and/or sounds from the environment often mask acoustic speech signals. In noise, listeners are charged with selectively attending to a talker of interest and segregating this speech signal from irrelevant sounds in the environment. Top down processing becomes more important for processing speech in noise, as context, knowledge, and experience help listeners to gain the gist of a message when the speech signal is degraded.

While younger adults are able to perceive speech fairly well in noisy situations, many older adults experience difficulties understanding speech in noise, especially when the noise consists of competing meaningful speech (i.e., informational masking; Helfer & Freyman, 2008). As older adults' often experience difficulties perceiving speech in noise to a greater extent than what might be expected based on their audiometric assessment, the underlying etiology of a speech in noise impairment likely goes beyond audibility (Spankovich, Gonzalez, Su, & Bishop, 2018; Tremblay, Pinto, Fischer, Klein, Levy, Tweed & Cruickshanks, 2015). Thus, the field of cognitive hearing science, which is discussed in more detail in section 1.3 of this introduction, aims to understand how both auditory and cognitive systems contribute to successful speech perception in noise, and how age-related changes to either of these systems may contribute to the communication challenges experienced by older adults.

1.2.2 Multimodal Speech Perception

When visual information from a talker is available, speech perception is multisensory. This is demonstrated by results showing that the visual input in audio-visual speech facilitates speech understanding in quiet (Davis & Kim, 2004) and in noise (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954) for listeners with and without hearing loss (Grant, Walden, & Seitz, 1998). Not only does the presentation of visual speech permit the listener to overcome ambiguities in the auditory signal and detect speech information unavailable to the auditory system (Jongman,Wang & Kim, 2003), but it appears to help with parsing speech input (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Visual speech also has a positive effect on the cognitive processes that are involved in speech processing and understanding, such as working memory (Frtusova & Phillips, 2016).

Given the benefits for speech processing, cognitive hearing researchers have incorporated multimodal speech input into their models of language understanding (e.g., The Ease of Language Understanding (ELU) Model – Rönnberg et al, 2008; 2013); however, these models do not provide thorough explanations of the role of cognition in multimodal speech perception or any potential effects of cognitive ageing on multimodal speech perception. This thesis research is grounded in the key concepts of cognitive hearing science that are presented in models like the ELU, but this research also aims to extend the field of cognitive hearing science by taking a multimodal (i.e., auditory-visual) approach to exploring the cognitive factors involved in speech perception.

1.3 Cognitive Hearing Science

1.3.1 History

As described by Arlinger, Lunner, Lyxell, and Pichora-Fuller (2009), cognition and hearing were mostly studied separately for the last quarter of the 20th century. Cognitive psychologists studied language processing largely by investigating reading, as visual stimulus presentation allowed for a high degree of control. Hearing researchers focused on the cochlea and peripheral, bottom-up processing rather than the cortex and top-down processing. Furthermore, audiologists focused on indexing hearing ability (or disability) by measuring the perception of simple tones and words (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009).

Within the last 20 years, the field of cognitive hearing science has emerged due to a need to understand how individuals cope in more ecologically valid conditions (e.g., in noise). Importantly, audiologists and hearing researchers have identified a difference between hearing and listening, whereby listening involves both auditory and cognitive factors (Kiessling et al., 2003). The study of ageing has been a source of unity between the once disparate hearing and cognitive science communities. Indeed, an increasingly senior population has been a catalyst for the development of a cognitive hearing science, as the need to understand how speech perception changes across the lifespan, and to design better technologies to assist older adults, becomes a more pressing issue.

A coming together of the hearing and cognitive sciences is evident in the early 21st century (Schneider and Pichora-Fuller, 2000; Wingfield, Tun, & McCoy, 2005). Schneider and Pichora-Fuller (2000), for example, argued that in order to understand how perception and cognition are affected by age, they must be considered as an integrated system. Wingfield, Tun, and McCoy (2005) also summarised the auditory and cognitive factors involved in language comprehension and identified how aging could affect different components of this auditory-cognitive system. Wingfield, Tun, and McCoy (2005) highlighted the role of working memory and age-limited processing resources (i.e., attentional resources), two cognitive constructs that continue to be key concepts within the field of cognitive hearing science.

1.3.2 Key Concepts

1.3.2.1 Working Memory

The most prominent cognitive factor studied within cognitive hearing science is working memory. For this thesis, working memory is defined as a limited capacity system that is responsible for processing and temporary storage of information for complex cognitive tasks (Baddeley, 2012). The role of working memory in understanding speech in adverse listening conditions is proposed in the ELU Model (Rönnberg et al., 2013; Rönnberg, Holmer, & Rudner, 2019), which is described in more detail in Chapter 2 of this thesis. Notably, the ELU model does not propose a distinct model of working memory in itself, rather, core components of the ELU model (i.e., the multimodal episodic buffer and the domain free general capacity system) are consistent with core components of Baddeley's (2012) model (i.e., the episodic store and the central executive; Baddeley, 2012; Rönnberg et al., 2013; Wingfield, 2016).

Although several measures of working memory capacity exist (Conway, Kane, Bunting et al. 2005), the tasks most commonly used within cognitive hearing science are auditory-verbal, complex-span tasks (i.e., The Reading Span and the Listening Span). Performance on these tasks has been associated with performance on speech recognition in noise tasks for older adults with and without hearing loss (Akeroyd, 2008; Gordon-Salant & Cole, 2016). These studies suggest that, in addition to (or instead of) hearing sensitivity, age-related declines in working memory capacity could contribute to older adults' difficulty understanding speech in noise.

1.3.2.2 Attention

Attention is a complex domain of cognition that is conceptualized and studied differently across different research disciplines (Neumann, 1996). Within the field of cognitive hearing science, attention has been conceptualized in three ways: as a limited capacity of resources, as multidimensional, and as an element of the working memory system. Although described individually below, these conceptualizations are not necessarily mutually exclusive. For example, a capacity of attentional resources could control how attention is divided between two tasks, and in turn how efficiently and accurately information from each task is encoded in working memory. Further, all three conceptualizations imply that attention facilitates performance on a task (e.g., speech understanding) by allowing a perceiver to "withdraw from some things [e.g., noise] in order to deal effectively with others [e.g., speech perception]" (James, 1890).

1.3.2.2.1 Attention as a Limited Capacity of Resources

Kahneman's (1973) Capacity Model of Attention suggests that each person has a limited capacity of attentional resources, that the maximum capacity varies between individuals, and that the amount of resources supplied to a task increases as the task becomes more difficult or "effortful". This understanding of attention has influenced the development of the Framework for Understanding Effortful Listening (FUEL), which defines listening effort as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task" (Pichora-Fuller et al., 2016, p. 11). This definition highlights a critical relationship between the allocation of *mental resources* (which are interchangeably referred to as attentional resources, processing resources, cognitive resources and resources within the literature) and cognitive effort. That is, measuring the extent to which attentional resources have been allocated should index the degree of effort required to complete a [listening] task (Pichora-Fuller et al., 2016).

The dual-task paradigm (Koch, Poljac, Muller, et al., 2018), has been widely adopted as a method to estimate the amount of mental resources allocated to listening (i.e., the degree of listening effort; Gagné, Besser, & Lemke, 2017; McGarrigle, Munro, Dawes, et al., 2014; Strand, Brown, Merchant, et al., 2018). For the dual-task paradigm, participants' performance on a primary task (e.g., speech recognition in noise) and a secondary task (e.g., counting how many times a particular shape appears on the screen) are measured when each task is presented alone (i.e., single-task performance) and when the tasks have to be completed simultaneously (i.e., dual-task performance). The difference between single task performance and dual-task performance for the secondary task is referred to as the dual-task cost, and is interpreted as a measure of the amount of cognitive resources spent on the primary task (Koch, Poljac, Muller, et al., 2018). Older adults with and without hearing loss tend to have greater dual-task costs than younger adults with normal hearing (Degeest, Keppler, & Corthals, 2015; Tun, McCoy, & Wingfield, 2009; Ward, Shen, Souza, Grieco-Calub, 2017, Xia, Nooraei, Kalluri, & Edwards, 2015).

Although dual-task methodology is grounded in the Kahneman's (1973) Capacity Model of Attention, the most suitable dual-task paradigm to measure listening effort has not been determined. As such, there is considerable variability in how dual-task procedures are designed, particularly in the selection of secondary tasks (Gagné, Besser, & Lemke, 2017). Rather than have participants perform two tasks concurrently (which inherently different than performing one task) the experiments presented in chapters three, four, and five of this thesis aimed to manipulate the attentional demands of the listening task itself, and compare younger and older adults performance across conditions that differ in their demand on attentional resources.

1.3.2.2.2 Attention as a Multidimensional Construct

Within the FUEL, attention is defined as “a multidimensional construct that includes orienting, selecting, and/or focusing on environmental stimuli (e.g., speech) or internal representations (e.g., thoughts) for varying periods of time”. Although this is a very broad definition of attention, it effectively captures that attention is often understood as multiple different processes. Examples of these processes include selective attention (i.e., focusing on one aspect of a stimulus input while ignoring or filtering out another; Phillips, 2016), divided attention (i.e., processing two or more tasks or sources of information at the same time; Phillips, 2016), sustained attention (i.e., maintaining focus over time), and attention switching (i.e., switching the focus of attention from one object or location to another). These dimensions are often further divided according to the modality of stimuli being studied (e.g., auditory selective attention) as an effect of attention in one modality does not necessarily generalise to a different modality (Guerreiro, Murphy, & Van Gerven, 2013; Shinn-Cunningham, & Best, 2008). The Test of Everyday Attention (TEA) has been used within cognitive hearing science to index these sub-processes of attention separately, in both auditory and visual modalities (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994). The TEA is used to measure sub-dimensions of attention in Chapter 2 of this thesis.

The concept of orienting is also included in the FUEL’s definition of attention. Orienting is one of three attention networks proposed by Posner and Peterson (1990;

2012) and represents the selection of a spatial location (or modality) to be the focus of one's perception, which in turn enhances the processing of stimuli at that location (or within that modality). Depending on the stimuli and task, orienting can be automatic or voluntary, and, in a visual context, overt (i.e., performed with head and/or eye movements) or covert (performed without a change in head or eye movement; Erel & Levy, 2016; Posner & Petersen, 1990). The attention network theory proposes three subprocesses of orienting: disengaging from the current focus, shifting to the new location or modality, and engaging attention at the new location or modality (Erel & Levy, 2016; Posner & Petersen, 1990). Although they are not specifically referenced within the FUEL's definition of attention, the remaining attention networks included in Posner and Petersen (1990) are alerting (i.e., achieving and maintaining a state of optimal vigilance for detecting and processing stimuli), and executive control (i.e., co-ordinating goal directed behaviour by managing information selection during complex tasks, resolution of conflict between competing cognitive processes, and co-ordinating process switching; Fan, Gu, Guise et al., 2009).

Support for Posner's theory of attention networks comes from the well-established attentional cueing paradigm (Posner, Snyder, & Davidson, 1980). As such, one strategy used by cognitive hearing scientists to measure the role of attention is speech processing has been to adapt the original visual cueing paradigm to an auditory cueing paradigm that includes a speech perception component (Kidd, Arbogast, Mason, & Gallun, 2005; Singh, Pichora-Fuller, & Schneider, 2008; Singh, Pichora-Fuller, & Schneider, 2013). Studies that have taken this approach suggest that knowing where to listen helps older adults to understand speech in noisy situations, but that having to switch attention (i.e., disengage and reengage) to different locations in an auditory scene may be particularly challenging for older adults (Singh, Pichora-

Fuller, & Schneider, 2013). The experiment presented in Chapter 4 of this thesis uses an adapted visual cueing paradigm to investigate the role of attention in auditory-visual speech perception in noise.

1.3.2.2.3 Attention as an Element of Working Memory

Attention has also been conceptualised as an element of the working memory system (Wingfield, 2016). That is, some cognitive scientists have suggested that the successful storage and manipulation of information within the working memory system is moderated by an individual's ability to control and sustain attention (Cowan, 2005; Engle, 2002; Wingfield, 2016). Indeed, Engle (2010) argues that individual differences in performance on complex span tasks reflect the ability of the central executive component of the working memory system to focus attention to stimuli and/or representations critical to a task, keep that information available in active memory or easily and quickly retrievable from inactive memory, and inhibit any stimuli or representations that would interfere with this process.

Engle's (2010) proposition is supported by studies that have found a strong association between performance on complex span tasks and measures of attentional control. That is, individuals with high working memory capacities (as measured by complex span tasks) perform significantly better on dichotic listening tasks and the executive control portion of the Attention Network Task, in comparison to individuals with low working memory capacities (Conway, Cowan & Bunting 2001; Redick & Engle, 2006). McCabe, Roediger, McDaniel, Balota and Hambrick (2010) also found that performance on complex span tasks was highly correlated to several measures of executive functioning (i.e., attentional control as conceptualised from a neuropsychologist perspective). McCabe et al. (2010) concluded that a common attention-related component, which they labelled "executive attention", likely underlies performance on both complex-span and executive functioning tasks. Most experiments in this thesis included an auditory-verbal complex span task (i.e., the listening span task) primarily as a measure of working memory capacity, however, it is recognised that performance on this task may be influenced by the ability to control attention.

1.3.3 Ecological Validity

Both this thesis research and the field of cognitive hearing science have a deeper motivation – that of trying to ensure that their endeavours relate to real-life communication concerns. In this regard, it is important to briefly review the concept of ecological validity within the constraints of experimental research paradigms.

Ecological validity is an abstract concept that consists of various dimensions (e.g., the nature of Research Setting or Context; the nature of the Stimuli and the nature of the Task, Behaviour, or Response). Thus, it is difficult to arrive at general criteria for determining whether any given experiment is ecologically valid or not. What is important is whether the context, stimuli, or responses have captured the critical aspects of the phenomena in question (Schmuckler, 2001).

In an effort to understand the role of cognition in communication challenges that listeners face during day-to-day interactions, researchers have taken standard speech recognition tasks and adapted certain components of these tasks (i.e., the stimuli presented or the response format) to incorporate features of real-life listening. In using this controlled approach, the goal is not necessarily to create a speech perception test that is identical to real-life, but to systematically test how critical components of real-life listening, in their most basic form, affect the skill under examination (i.e., speech perception). For example, to examine difficulties in speech perception in noise, an experimenter might use a noise stimulus (e.g., speech shaped noise) and signal-to-noise ratios (SNRs) that do not often occur in daily-life but that nevertheless capture important aspects of the phenomenon – i.e., a challenging situation in which the speech signal is degraded.

However, there are critical aspects of the phenomenon in question (i.e., communication in challenging conditions) that are not typically incorporated into research designs. For example, considering the nature of the stimuli presented, visual distraction is not usually included in experiments testing auditory-visual speech perception in noise (i.e., the visual speech benefit). Although these studies extend the ecological validity of auditory-only speech perception tasks, they could be overestimating the benefit that visual speech provides in a more complex visual scene. Furthermore, when considering response measures, almost all speech perception in noise research has used word recall as a response measure. This effectively measures word recognition, but minimizes the importance of having to actually understand what was said.

The following section summarises the two key components of behavioural speech perception in noise tests (i.e., the stimuli presented and the response format), and how each component has been manipulated in order to incorporate components of real-life listening into experimental paradigms. Strengths and limitations of the approaches that have been used are also discussed.

1.4. Behavioural Measures of Speech Perception in Noise

1.4.1 Auditory Stimuli

Two essential components of any speech perception in noise task are the auditory speech signal and the auditory masker (i.e., noise). The linguistic complexity of the input signal used for speech perception in noise tasks varies across studies and has been shown to influence the degree to which cognitive processes are engaged.

That is, more linguistically complex signals (e.g., sentences) tend to be more strongly associated with cognitive processes than less complex speech signals (e.g., phonemes, syllables, words; Heinrich, Henshaw, & Ferguson, 2015). Studies that aim to assess the role of processes that may be engaged in real-life listening (especially the role of cognition) have typically used more complex linguistic inputs, such as IEEE sentences (Dryden, Allen, Henshaw & Heinrich, 2017; Rothauser et al., 1969; Schoof & Rosen 2014). These standardised sentences are presumed to measure the sensory and informational processing skills necessary for real-life listening while controlling for potential confounding variables such as context and speech clarity. That is, standardised sentences typically have low predictability, consist of read speech that is clearly articulated, are produced in quiet with noise added at a later time, and do not have conversational context. The experiments presented in the first three experimental chapters used standardised sentences (and thus the controls listed above) as the speech signal so that the results from these experiments could be compared to the majority of previous studies on ageing and speech perception in noise.

The type of noise used to mask speech signals also varies across studies but has been broadly classified into two categories: energetic and informational. Energetic masking refers to when the temporal and spectral properties of the speech signal and noise overlap, limiting the audibility of the signal (Cooke, Garcia-Lecumberri & Barker, 2008; Kidd, Mason, Richards, Gallun, & Durlach, 2008; Pollack,1975). The effect of energetic masking on speech perception is therefore highly dependent on the degree of acoustic overlap between the speech signal and the noise masker. That is, the greater the acoustic overlap, the greater the effect that energetic masking will have on speech perception.

Informational masking is a term used to describe all remaining sources of interference with a speech signal when the effects of energetic masking have been accounted for (Cooke, Garcia-Lecumberri & Barker, 2008). In general, it is proposed to occur in conditions where speech is masked by another source of speech (e.g., a single talker, multi-talker babble; Kidd, Mason, Richards, Gallun, & Durlach, 2008; Pollack, 1975). Informational masking is arguably more cognitively demanding than energetic masking, as listeners need to differentiate between what the talker of interest is saying versus another irrelevant talker. However, depending on the spectral characteristics of the informational masker, it is possible to have both energetic and informational masking at the same time. Although differentiating between the effects of informational and energetic masking is a debated topic (e.g., Durlach, 2006), the proposed effects of each masker type are summarized in Figure 1.1.

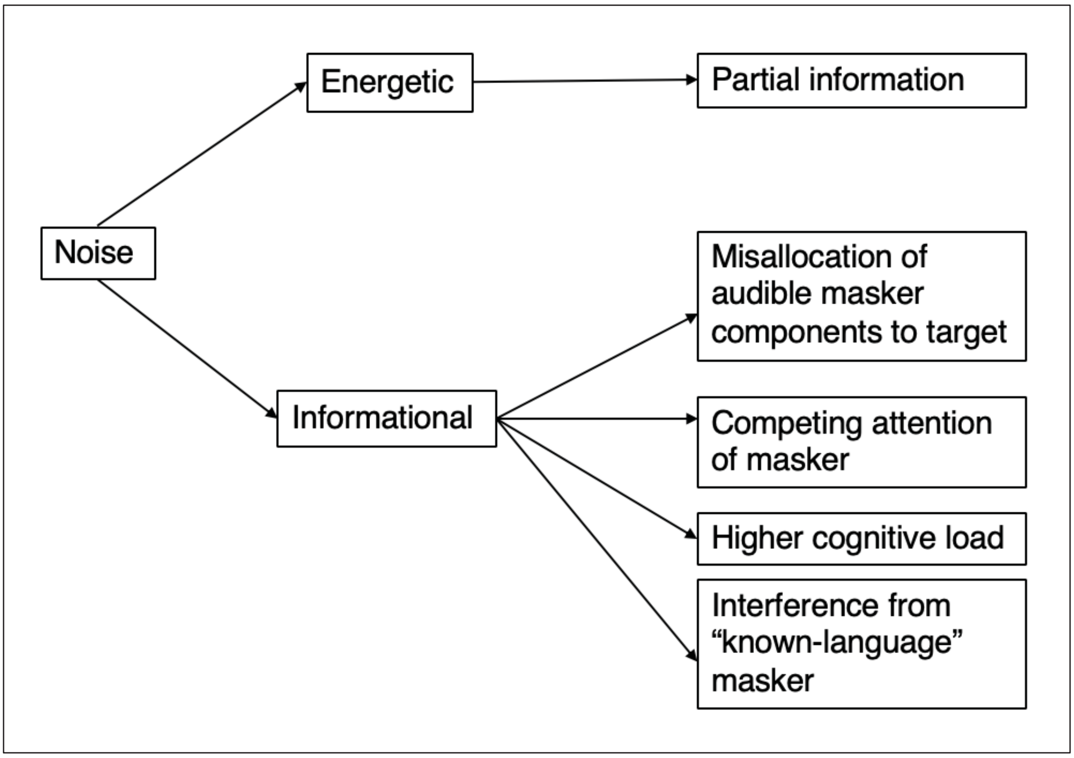

Figure 1.1. Energetic and Informational Masking. Figure adapted from Cooke, Garcia-Lecumberri & Barker, 2008.

Informational masking (i.e., eight talker babble) was used for the auditory-only speech perception in noise task presented in the first experimental chapter (i.e., Chapter 2). However, given that the focus of the thesis was on manipulating visual information, I controlled the characteristics of the auditory masker as much as possible by using energetic masking. That is, for the experiments presented in Chapters 3,4, and 5, speech-shaped noise was derived from the long-term average spectrum of the speech signals recorded and then mixed (with the same speech signals) at specific SNRs.

In order to avoid ceiling effects, the SNRs employed for the experiments included in this thesis, and for most research on ageing and speech perception in noise, are generally more adverse than SNRs that have been estimated for real-life communication situations (To & Chung, 2014; Weisser & Buchholz, 2019). Although the use of particularly adverse SNRs likely represents relatively rare conditions in daily life, it is possible that performance at these adverse SNRs draws on similar processes to those used during more favourable conditions (Weisser & Buchholz, 2019). However, it is also possible that the involvement of cognitive abilities during speech perception in noise is modulated by specific characteristics of a listening situation, such as the SNR (Helfer & Freyman, 2014; Henrich & Knight, 2016; Naylor, 2016). Additional research is needed to understand the precise situations in which cognitive abilities come into play in listening.

1.4.2 Visual Stimuli

Some speech perception in noise tasks do incorporate visual stimuli in that a video of a face uttering the auditory signal is presented (in addition to the auditory signal and noise) for some trials. The results of these studies indicate that seeing a talker's face is one of the largest intelligibility benefits available to a perceiver. That is, when a talker's face can be seen, speech recognition in noise improves by approximately 10-15dB in comparison to an auditory-only baseline (i.e., the visual speech benefit; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954). As even the most advanced hearing aid technology is only able to provide approximately a 3-5 dB benefit in noise (e.g., Wu et al., 2019), visual speech is clearly an important source of speech-related information.

1.4.2.1 Do older and younger adults get a visual speech benefit in noise when visual distractors, in addition to visual speech that matches the auditory signal, are within the visual field?

Existing research suggests that older and younger adults have equal access to the visual speech benefit in noise (Cienkowski & Carney, 2002; Jesse & Janse, 2012; Middelweerd & Plomp, 1987, Sommers, Tye-Murray, Spehar, 2005; Tye-Murray, Spehar, Myerson, Hale & Sommers, 2016; Winneke & Phillips, 2011); however, as the visual stimuli presented for these studies was not demanding on visual-spatial attention (i.e., there was only one face to look at/attend to), any consequences of cognitive ageing on the ability to gain a visual speech benefit would not have been observed. As research suggests that auditory-visual processing requires visual-spatial selective attention (Andersen et al., 2009; Tiippana, Anderson, & Sams, 2004; Alsius & Soto-Faraco, 2011), and that older adults have limited attentional resources and control (Craik & Byrd, 1982; Craik & Salthouse, 2011; Greenwood & Parasurman, 2004; Lustig, Hasher, & Zacks, 2007; Madden, Connelly & Pierce, 1994), it is possible that the presence of visual distractors within the visual field may interfere with older adults' ability to get a visual speech benefit.

This proposal was explored in the current thesis by comparing younger and older adults' speech recognition in noise performance on standard auditory-only and auditory-visual conditions, to performance on auditory-visual conditions with additional visual information (i.e., visual distractors). That is, by incorporating a component of real-life listening that is not usually included in speech perception in noise tasks (i.e., visual distractors), the experiments presented in Chapters 3, 4, and 5 aimed to test how age-related changes in attentional resource capacity and visual-spatial attention affect older adults' ability to gain a visual speech benefit in noise.

1.4.3 Response Format

Although the majority of speech perception in noise tasks use verbatim recall as a response format, researchers have started to test alternative response formats that may be more representative of speech perception in real-life. That is, during day-to-day interactions, listeners are not typically asked to report back exactly what they have heard. Rather, listeners need to be able to understand the gist of what was said and make an appropriate response within a socially acceptable amount of time (Best, Streeter, Roverud, Mason, & Kidd Jr., 2016).

One method researchers have used to measure speech comprehension rather than speech recognition is to present participants with multiple-choice and/or short answer style questions during or after the presentation of a speech stimulus (i.e., a passage or discourse; Best, Keidser, Buchholz, Freeston, 2016; Best, Keidser, Freeston, & Buchholz, 2016; Gordon, Daneman, & Schneider, 2009; Schneider, Daneman, Murphy, & See, 2000; Sommers, Hale, Myerson et al., 2011; Tye-Murray, Sommers, Spehar et al., 2008). Although this method captures the speech comprehension component of listening, it also places demands on skills that may not be particularly relevant to speech perception, such as reading ability.

The Question-and-Answer Task (which is discussed in more detail in Chapter 5), on the other hand, does not require participants to read text in order to select a response as it uses a true/false response format (Best, Streeter, Roverud, Mason, & Kidd Jr., 2016). One practical benefit of a true/false response format is that it facilitates the presentation of visual speech stimuli, as participants should be able to keep their eyes focused on a computer monitor rather than a keyboard (Best, Streeter, Roverud, Mason, & Kidd Jr., 2016). A true/false response format also allows researchers to obtain both accuracy and appropriate response time data. Response time, although not typically measured for speech perception tasks due to response format limitations, could be useful for measuring the cognitive demands of listening (i.e., listening effort; van den Tillaart-Haverkate, de Ronde-Brons, Dreschler, & Houben, 2017).

1.5 Overview of Thesis

At a broad level, the research in this thesis aimed to expand the agenda of cognitive hearing science to encompass multimodal speech perception. More specifically, the experiments focused on the contributions of visual speech, visual distractors, and cognition to speech perception in noise for younger and older adults. A focus on these factors is rare, but in my view, concern with the influence of such factors fits squarely with the underlying goal of cognitive hearing science (i.e., a greater understanding real-life listening).

First, I evaluate the performance of 30 younger and 30 older adults on a standard, auditory-only, speech perception in noise task and a series of tasks measuring auditory, cognitive, and lifestyle factors that contribute to speech perception in noise ability (Chapter 2). The results from this study characterize the participant groups that were generally tested for the subsequent auditory-visual experiments included in this thesis. In Chapter 3, I test how seeing multiple talkers affects the visual speech benefit in noise for younger and older adults. For this experiment, I adapted the standard speech in noise paradigm, where participants are required to identify speech in noise for auditory-only and auditory-visual conditions, by manipulating the number of talking faces (1, 2, 4, or 6) presented for the auditory-visual conditions. In Chapter 4, I test whether younger and older adults can gain a standard visual speech benefit when visual speech that matches the auditory signal and visual speech that does not match the auditory signal (i.e., a visual speech distractor) are presented, and the location of the matching talker's face is visually cued. Lastly, in Chapter 5, I present an auditory-visual version of a conversation-comprehension style task (i.e., the Question-and-Answer Task) and test whether visual distraction affects younger and older adult's performance (response time and accuracy) on this task. Together, the aim of these chapters is to increase our understanding of the role that cognition plays in auditory-visual speech perception in noise, and how cognitive ageing affects speech perception in noise for older adults. It should be noted that as the experiments have been written as a series of papers, there is some necessary overlap between the introductions of each chapter. The candidate is preparing manuscripts for submission to academic journals.

Chapter 2

Exploring the effect of age on auditory, cognitive, and life-experience factors that contribute to speech perception in noise ability

2.1 Introduction

A common complaint among older adults is that understanding speech in noisy environments is challenging (CHABA,1988; Pichora-Fuller, Alain & Schneider, 2017). As communication often takes place in locations with some degree of noise (e.g., a café), and as social interaction is related to positive mental and physical health outcomes (Cherry et al., 2013; Gilmore, 2012), strategies to help older listeners overcome this challenge would be beneficial for an ageing population (Heinrich et al., 2016). A critical first step in developing these strategies, however, is understanding the factors that contribute to the communication difficulties that many older adults experience.

It has been proposed that understanding the basis of older adults' speech perception difficulties requires a consideration of peripheral, central, and cognitive factors (CHABA, 1988). Based on this division, researchers have tested the peripheral, central, and the cognitive hypotheses separately, with the majority of work focusing on the peripheral hypothesis (Humes et al., 2013; Van Rooij et al, 1989). Within the last decade, however, there has been a greater focus on the idea that an individual's difficulty in noise could be due to a combination of contributions from peripheral, central, and/or cognitive factors, and that the distinction between these three factors may become less clear as a listening task becomes more complex (Akeroyd, 2008; Heinrich, Henshaw, & Ferguson, 2015; Pichora-Fuller, Alain & Schneider, 2017).

The main aim of this chapter is to characterise the auditory-peripheral and cognitive functioning of the participant groups (i.e., 30 younger and 30 older adults) that completed the speech perception in noise tasks included in this thesis. Although the individual participants were not exactly the same for each experiment, participant groups in chapters 2, 3, and 4 are likely similar to the samples tested for the present study, as identical recruitment methods were used for all experiments (i.e., younger adults were students from Western Sydney University and older adults were recruited from community groups such as computer clubs). The remainder of this introduction will review existing research on ageing and speech perception in noise, starting with a brief discussion of the role of the peripheral and temporal processing components of the auditory system and then focusing on the role of three cognitive factors: processing speed, working memory, and attention. The potential modulating effects of life experiences (e.g., exercise) on cognition and audition are also discussed.

2.1.1 Auditory Ageing

2.1.1.1 Hearing Sensitivity

Although age-related hearing loss (i.e., presbycusis) typically presents as increased hearing thresholds (particularly at high and low frequencies) and poor frequency resolution, different cochlear pathologies can underlie these symptoms (Gates & Mills, 2005; Yamasoba et al., 2013). That is, sensory, neuronal, and/or metabolic damage to the cochlea can all result in hearing loss and consequently, difficulties understanding speech. The cause and severity of an individual's hearing loss can vary in relation to particular environmental, life-experience, and genetic risk factors. For example, cardiovascular health has been related to metabolic damage (i.e., the functioning of the stria vascularis), whereas noise exposure has been related to sensory hair cell loss (Gates & Mills, 2005; Yamasoba et al., 2013).

The most widely used measure of hearing loss is the pure-tone audiogram. For this measure, listeners are presented with pure-tones at varying frequencies, and the volume at which a person is able to detect the sound at each frequency is determined. However, research investigating how age-related changes in hearing sensitivity (measured by the pure-tone audiogram) affect speech perception in noise has produced mixed results. Some studies indicate that older adults' average hearing loss at moderately-high frequencies (i.e., 1000Hz, 2000Hz, 4000Hz) can account for a significant portion of variance in performance on speech perception in noise tasks (i.e., 50-75%; Amos & Humes 2007; Humes 1994, Humes, 2013; Schoof & Rosen, 2014; Van Rooij et al, 1989), whereas other studies have only found a significant correlational relationship between speech perception in noise and hearing sensitivity at particularly high frequencies (i.e., 6000 Hz, 8000 Hz, 10,000 Hz; Besser, Festen, Goverts, Kramer, Pichora-Fuller, 2015). Due to this discrepancy, the current study will examine moderately-high and high hearing thresholds separately for both younger and older adults. However, as there are studies that have not found any relationship between speech perception in noise and moderately-high or high pure-tone thresholds (e.g., Dubno 1984; Duquesnoy, 1983; Jerger, 1992; Vermiglio, Soli, Freed, & Fisher, 2012), and as younger and older adults with clinically normal pure-tone audiometric thresholds still display variance in performance on speech perception in noise tasks (e.g., Füllgrabe, Moore & Stone, 2015; Schoof & Rosen, 2014), the audibility of a speech signal cannot be the only requirement for accurate speech perception in noise.

2.1.1.2 Temporal Processing

Independent of hearing sensitivity, an age-related decline in auditory temporal processing could contribute to older adults' difficulties understanding speech in noise (Pichora-Fuller & MacDonald 2007; Gordon-Salant, Fitzgibbons, Yeni-Komshian, 2011). Moore (2008) describes auditory temporal processing as the decomposition of sound in the time domain onto a slowly varying envelope (ENV) superimposed on a more rapidly varying temporal fine structure (TFS). Both ENV and TFS processing are important for perceiving the phonetic, phonemic, and prosodic components of speech, particularly in noise (Pichora-Fuller & MacDonald 2007, Schnieder & Pichora-Fuller, 2001).

Support for age-related declines in ENV processing comes from studies measuring gap detection thresholds. For these studies, listeners are asked to detect a brief silent interval inserted in an otherwise continuous tone or noise burst. The shortest silent interval that a listener can detect is the gap detection threshold (GDT, Gordon-Salant, Fitzgibbons, Yeni-Komshian, 2011). Gap detection thresholds tend to increase with age, however more robust age differences seem to occur when the gap is located near the stimulus onset or offset (He *et al*., 1998) rather than centrally (He *et al*., 1998; Schneider *et al*., 1994; Snell, 1997), and when the duration of markers is short (i.e., < 500ms; Schnieder & Hamstra, 1999). Older adults' GDTs have been associated with consonant identification and speech perception in noise performance (Pichora-Fuller & MacDonald 2007; Snell, Mapes, Hickman & Frisina, 2002; Tyler, Summerfield, & Wood, 1982; Vermeire, 2016).

Several studies also suggest that TFS processing declines with increasing age (Ross et al., 2007; Grose & Mamo, 2010; Hopkins and Moore, 2011; Moore et al., 2012a, 2012b; Füllgrabe, 2013; King et al., 2014; Füllgrabe et al., 2015; He, Mills, Dubno, 2007). For example, studies measuring frequency modulation (FM) detection (He et al., 2007), pitch discrimination (Füllgrabe, 2013), and inter-aural phase or time difference detection (Grose & Mamo, 2010), all support an age-related deficit in TFS processing. Furthermore, TFS processing has been related to performance on concurrent vowel identification and speech perception in noise tasks (Sheft, 2012; Pichora-Fuller & MacDonald 2007, Snyder & Alain, 2005).

2.1.2 Cognitive Ageing

To test the cognitive hypothesis (i.e., that age-related declines in cognitive functioning contribute to older adults' difficulties understanding speech in noise; CHABA, 1988) researchers have focused on measuring three cognitive factors: Processing Speed, Working Memory, and Attention. This section will define each of these cognitive factors, describe how they are typically measured in behavioural research, and provide an overview of the results from research that has investigated the role of each factor in speech perception in noise ability.

2.1.2.1 Processing Speed

Processing Speed Theory was developed by Salthouse (1996) and refers to the proposal that a major factor underlying age-related differences in cognitive functioning between older and younger adults is the reduction in the speed with which cognitive operations can be executed that occurs in old age. This theory would therefore predict that older adults' speech in noise difficulties are in part due to this general decline in the speed at which cognitive operations can be successfully completed. With regard to speech processing, this proposal has been supported by studies showing that older adults understand less speech than younger adults when the rate at which speech is presented is increased, particularly when the target speech is presented with noise (Tun, 1998). Further support for processing speed theory comes from studies that have found measures of processing speed (e.g., the Digit Symbol Substitution Test; Wechsler, 1981) to be reliably associated with older adults' performance on speech perception in noise tasks (Helfer, 2014; Tun, 1999; van Rooij, Plomp, & Orlebeke, 1989; Zekveld, Kramer, & Festen, 2011). It is worth noting, however, that other studies have reported that measures of processing speed are not significantly predictive of (or correlated with) speech perception in noise performance (Gordon Salant & Cole, 2016; Schoof and Rosen, 2014), and suggest that other auditory (e.g., temporal processing) and cognitive (e.g., working memory) skills may be more critical for explaining variance in speech perception in noise abilities.

2.1.2.2 Working Memory

Working memory is the most prominent cognitive construct within the field of cognitive hearing science. Working memory is defined as a limited capacity system that is responsible for processing and temporary storage of information during complex cognitive tasks, such as language comprehension. According to Baddeley's multi-component model of working memory, the system is comprised of a control process (i.e., the central executive) and multiple storage processes (i.e., the visual spatial sketchpad and the phonological loop) that connect to long-term memory (Baddeley, 2012). Since working memory is understood as a limited capacity system, there is always a trade-off between storage and processing; if a task requires more processing, then less information can be stored, and vice versa. Furthermore, when an individual's working memory capacity is reached, then both storage and processing are impaired.

An individual's working memory capacity is typically measured by performance on a complex span task (i.e., The Reading Span or The Listening Span), which index the ability to simultaneously store and process information. In the most widely used version of the reading span test, a participant reads a set of sentences and determines if each sentence makes sense or not (e.g., "the girl sang a song" vs "the train sang a song"). After a set of sentences have been presented, the participant is asked to recall the final word from each sentence in a set. The set size is gradually increased and an individual's working memory capacity is determined by the largest set size that a participant can correctly recall a specified proportion of the final words (Daneman & Carpenter, 1980; Rönnberg et al., 2013). Older adults tend to perform worse on the reading span test than younger adults, although there is considerable within-sample variability for both age groups (Souza, Arehart, & Neher, 2015).

2.1.2.2.1 How Working Memory Contributes to Speech Perception in Noise

One prominent proposal for why and how working memory is involved in speech understanding (particularly in adverse listening conditions) is the Ease of Language Understanding (ELU) model (Figure 2.1; Rönnberg et al. 2008, 2013, 2019). The key proposal of this model is that when speech is degraded in some way, working memory resources are recruited to aid understanding. If an individual has a reduced working memory capacity (e.g., as in an older adult), then her/his ability to recruit top-down working memory resources will be less efficient, resulting in reduced speech comprehension in difficult listening situations.

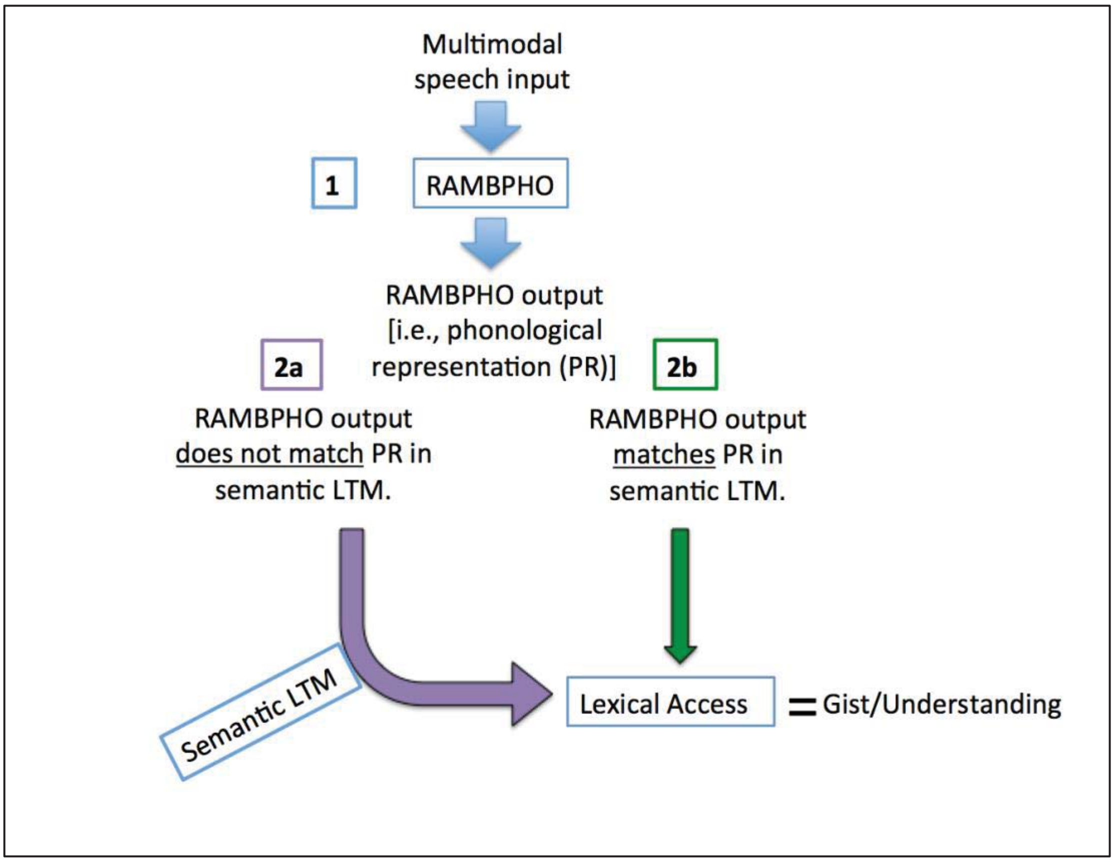

Figure 2.1. Summary of the Ease of Language Understanding (ELU) Model (Ronnberg, 2008 and 2013). [1] Multimodal speech input is **R**apidly, **A**utomatically, and **M**ultimodally **B**ound into **Pho**nological representations by an episodic buffer termed **RAMBPHO**. [2b] When an incoming signal matches a stored representation in the long-term memory, language understanding is automatic or implicit. [2a] If an incoming signal does not match a stored representation in Long Term Memory (LTM), due to hearing loss or a degraded signal, a mismatch occurs. In this case, working memory resources are engaged in order to support understanding. The slower, explicit processing loop recruits information from the semantic LTM in an attempt to fill in missing information.

2.1.2.2.2 Testing the ELU Model

2.1.2.2.2.1 Complex Span Tasks

The ELU model suggests that working memory capacity is important for speech understanding in noise for all listeners. However, research suggests that age and hearing status may combine to moderate the effect of working memory on speech perception in noise (Füllgrabe & Rosen, 2016). For example, verbal measures of working memory capacity (e.g. The Reading Span; Daneman & Carpenter, 1980) have been consistently related to performance on speech perception in noise tasks for older adults with hearing loss. Studies testing younger and older adults with normal hearing, however, have produced mixed results (Füllgrabe & Rosen, 2016; Gordon-Salant & Cole, 2016; Schoof and Rosen, 2014). Similarly, Schoof and Rosen (2014) did not find a significant relationship between reading span scores and speech perception in noise performance for younger and older adults with normal hearing. Indeed, a meta-analysis conducted by Füllgrabe and Rosen (2016) indicated that individual variations in working memory capacity account for less than 2% of variance in speech perception in noise identification scores for younger adults with normal hearing. However, there are studies that suggest that working memory is important in speech understanding even for those with normal hearing if a sensitive measure is used. For instance, Gordon-Salant and Cole (2016) divided normal hearing participants into groups based on age and performance on an auditory-verbal working memory capacity measure (i.e., The Listening Span), and found that younger and older adults with high working capacities had lower SRTs (i.e., performed better) than younger and older adults with low working memory capacities. From this, it would seem that the listening span measure may be more sensitive to age-related changes in working memory than other working memory tasks.

2.1.2.2.2.2 Simple Span Tasks

Although the reading span and listening span are the measures most often used to operationalize working memory within the field of cognitive hearing science, several studies have also used simple span tasks such as the digit span. Using the digit span in addition to complex span tasks is advantageous as this allows for a more direct investigation of the subcomponents of working memory (Millman & Mattys, 2017). That is, forward digit span is thought to measure the phonological storage component of working memory; whereas backwards digit span is thought to measure both phonological storage and executive control (i.e., storage and processing; Millman & Mattys, 2017). A meta-analysis conducted by Bopp and Verhaeghen (2005) suggests that older adults tend to perform worse than younger adults on both the forward digit span and the backwards digit span; however, there is typically a greater age difference for backwards digit span (i.e., when both storage and processing are required).

Studies that have investigated the relationships between ageing, performance on digit span tasks, and speech recognition in noise have produced variable results. In contrast to Bopp and Verhaeghen (2005), Füllgrabe et al. (2015) did not find a significant difference between younger and older adults' performance for forward or backward digit span. However, for the older adults in Füllgrabe et al.'s (2015) study (who had normal hearing) both forward digit span and backward digit span were significantly related to performance on a sentence recognition in noise task (DSF: $r = 0.76$, DSB: $r = 0.59$; $p < 0.05$). These correlations remained significant when the younger and older adults were grouped together (with age partialled out; DSF: $r = 0.65$ DSB: $r = 0.47$, $p < 0.05$). Heinrich et al., (2015) also administered both the backwards and forwards digit span measures to older adults (i.e., 44 subjects aged 50-74, $M = 65.3$, $SD = 5.7$), however, they only found a significant relationship between backwards digit span and the speech recognition in noise task ($r = -0.32$, $p < 0.05$).

That is, Heinrich et al's. (2015) results suggest that older adults who had higher scores on backwards digit span performed better on a speech recognition in noise task (i.e., lower speech reception thresholds on the Adaptive Sentence List). In contrast, Millman and Mattys (2017) tested 30 adults, 31-67 (M = 53.5, SD = 9.4) years of age and found that age was predictive of performance on forward digit span (R_2 = 0.15, p = 0.03); however, age was not predictive of performance on backwards digit span (R_2 = 0.05, p = 0.25). Further, neither of the digit span tests improved the fit of the regression models for predicting speech recognition in noise for any of the SNR's tested.

2.1.2.3 Attention

The role of attention in understanding speech in difficult listening situations has been proposed in the Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller et al., 2016). The FUEL (See Figure 2.2) is based on Kahneman's (1973) Capacity Model of Attention which proposed that individuals have a limited capacity of attentional resources and that the decision to allocate attentional resources to a specific task (e.g., listening) depends on both the demands of the task and the motivation of the individual. In the context of listening, the FUEL suggests that as the quality of a speech signal becomes poorer (due to hearing loss, noise, or accented speech) the demand for attentional resources increases, which subsequently increases the degree of listening effort (i.e., "the allocation of mental resources in goal pursuit when carrying out a listening task" Pichora-Fuller et al., 2016, pp.13S). Thus, according to FUEL, age-related declines in the overall capacity and/or ability to control attentional resources could contribute to older adults' difficulties understanding speech in noise (Phillips, 2016). Reduced motivation and general fatigue could also affect how older adults allocate attentional resources to listening tasks (Pichora-Fuller et al., 2016).

Since the FUEL was published, research into the validity and reliability of various methods to index listening effort has increased (e.g., pupillometry, EEG, questionnaires;

Hughes, Hutchings, Rapport, McMahon, Boisvert, 2018; Miles et al., 2017). However, less work has focused specifically on the relationship between attention and speech perception in noise. Indeed, Dryden, Allen, Henshaw, and Heinrich (2017) concluded that attention was not featured in a sufficient number of studies to be included in a meta-analysis reviewing the association between cognition and speech perception in noise. As cognitive ageing research suggests that various attentional processes decline with age (Craik & Byrd, 1982; Zanto & Gazzaley, 2014), further research is needed to understand how age-related declines in attention could be contributing to the difficulties that older adults experience in noisy situations.

As attention is such a multifaceted construct, it is possible that uncertainty regarding the best way to behaviourally measure attention may have contributed to the lack of research focused on ageing, attention, and speech perception in noise. One approach has been to use the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994). The TEA consists of a series of auditory and visual tasks that claim to measure selective attention, sustained attention, attention switching, and working memory, respectively. Füllgrabe, Moore, and Stone (2015), for example, administered a comprehensive cognitive test battery that included the TEA to younger and older adults with normal hearing. They found that younger adults performed significantly better than older adults on the selective attention and attention switching measures. However, they found that for an older adult participant group, none of the TEA measures were significantly related to consonant or sentence recognition in noise after applying a Holm Bonferroni Correction (See Füllgrabe et al., 2015 for a summary of *r* values).

Alternatively, Anderson, White-Schwoch, and Parbery-Clark (2013) extracted the auditory attention quotient (AAQ) from the Integrated Visual and Auditory Continuous Performance Test of Attention (IVA+) as an index of sustained attention. For this study, the

IVA+, as well as a variety of other hearing, cognitive, and speech perception tasks, were administered to 120 older adults with normal hearing. The AAQ was positively correlated with the sentence recognition in noise task (i.e., the HINT; $r = 0.265$, $p < .01$). However, the AAQ was not a significant predictor of HINT performance in a hierarchical regression model ($B = 0.159$, $SE\ B = 0.106$, $\beta = 0.152$, $p = 0.138$).

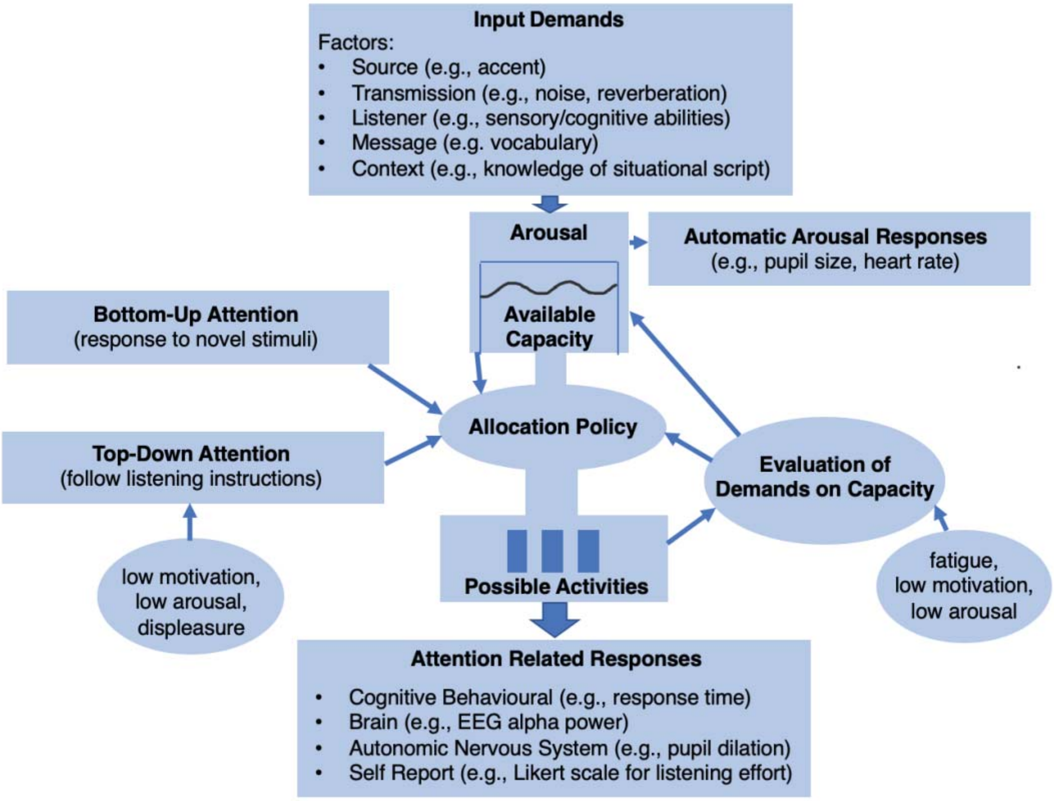

Figure 2.2. An adaptation of Pichora-Fuller et al.,'s (2016) interpretation of Kahneman's (1973) Capacity Model of Attention in relation to listening effort. Available cognitive capacity and arousal both increase or decrease depending on the current input related demands (e.g., speech in noise). The allocation policy and the evaluation of demands work together to govern which activities need capacity and how much they will receive. According to Kahneman (1973), the allocation policy is influenced by 1) Enduring dispositions (i.e., bottom- up or "automatic" attention), 2) Momentary intentions (i.e., top-down or "intentional" attention), 3) The evaluation of demands, and 4) Effects of arousal. Fatigue, low motivation, low arousal, or displeasure can influence a) the evaluation of demands on cognitive capacity and b) intentional (i.e., top-down) attention. Attention related responses can be used to index listening effort.

2.1.2.4 Life-Experience Factors

Anderson, White-Schwoch, and Parbery-Clark (2013) outline four life-experience factors (i.e., Social Economic Status, Intellectual Engagement, Musical Training, and Physical Activity) that may moderate older adults' speech perception in noise abilities. In summary, each of these factors has been associated with aspects of cognition such as working memory and attention (Luo, 2005; Netz et al., 2011; Parbery-Clark, Anderson, Hittner, Kraus, 2012; Parbery Clark, Strait, Anderson, Kraus, 2011; Salthouse, 2006). As these life-experience factors may play a role in offsetting cognitive decline and in turn supporting speech perception in noise later in life, they are worthwhile including in any general characterization of an older adult cohort.

2.1.2.5 The Present Study

The present study aimed to evaluate 30 younger and 30 older adults' performance on a standard speech in noise task and to evaluate the same participants' performance on a series of tasks measuring auditory and cognitive factors that have been previously shown to affect speech perception in noise ability (as outlined above). The research inventory included measures of hearing sensitivity, auditory temporal processing, processing speed, (short-term and working) memory, and (selective and sustained) attention. Questionnaires were also administered to measure life-experience factors (e.g., physical activity) that could moderate the relationships between age, cognition, and speech perception in noise. The principle motivation in conducting the current study was to provide a detailed characterisation of an older adult cohort typical of that tested in subsequent experiments included in this thesis.

2.2 Method

2.2.1 Participants

Thirty older adults (17 Females, $M_{Age} = 70$) were recruited from the Western-Sydney region by advertisements in public places (e.g., libraries) and appeals to community clubs (e.g., The Rotary Club). Older adults received $50 for their participation. Younger adults (n = 30; 20 Females, $M_{Age} = 21$) were students at Western Sydney University and participated for course credit. All participants passed a screening test for cognitive impairment (i.e., Addenbrooke's Cognitive Examination; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006) and did not wear hearing aids at the time of testing.

2.2.2 Equipment

A laptop PC (Windows 7) was used in all computer-based tasks unless otherwise stated. Sound was delivered through Sennheiser HD280pro headphones. Depending on the task, response collection was made via either a mouse click, a button press, or orally. For cognitive tests (paper-based) participants and the experimenter sat across from each other at a table in a sound attenuated booth and completed the tasks. Depending on the task, participants gave their response orally or provided a written response.

2.2.3 Speech Perception in Noise

Speech Reception Thresholds in Noise (SRTn) were assessed using National Acoustic Laboratories' Beautifully Adaptive Speech Test (BASTE; Keidser, Dillon, Meja, Nguyen, 2013). The program presents Bamford-Kowal-Bench (BKB) sentences in 8 talker babble noise. After each sentence is presented, participants orally repeat as many words as possible and the researcher selects all of the morphemes correctly reported via a mouse click (morphemes for each sentence are visible to the researcher on the monitor). The algorithm uses this morphemic score to adapt the noise level for the next item(s). The program was set to determine the SRTn at 50% correct for each participant. A lower SRTn indicates better speech perception in noise abilities.

2.2.4 Hearing

2.2.4.1 Pure-tone Audiometry (PTA)

Pure-tone thresholds for both ears (Diagnostic Audiometer, AD229e) were measured at 7 different frequencies (0.25,0.5, 1, 2, 4, 6, and 8 kHz). Thresholds from moderately-high frequencies (1, 2, and 4 kHz) and high frequencies (6 & 8 kHz) were averaged to produce a mean moderately-high threshold and a mean high threshold respectively.

2.2.4.2 Temporal Processing

2.2.4.2.1 Gap Detection threshold (GDT)

The stimuli for estimating Gap Detection thresholds (GDTs) were created using a custom written MATLAB script. The GDTs were estimated for the broadband noise probe which were 750 ms long and were generated at a sampling frequency of 44100Hz. The broad band noise was further band pass filtered between 100-7999Hz with a 10ms raised cosine ramp applied at the onset and offset of the noise. The gap was introduced at the temporal centre of the noise. The onset and offset of the gap had a 0.5 ms raised cosine ramp. The overall duration of noise with and without the gap was same.

To estimate the gap detection thresholds, a transformed 2-down-1-up adaptive psychophysical procedure was used (Levitt, 1971). The responses from the participants were acquired using a 2-interval, 2 alternate forced choice (2I2AFC) method that estimated the 70.7% point on a psychometric function. Among the two intervals, one contained a continuous 750 ms broadband noise burst with no gap and another interval contained a gap. The gap in the noise was randomly assigned to one of the intervals. Both the intervals were presented in succession with an inter-stimulus interval of 250ms. A participant's task was to attend to both the intervals and to identify the interval with the gap by clicking on the options

1 or 2 on the computer screen. Response accuracy feedback was provided after each trial. The starting duration of the gap was 10 ms for both age groups. Subsequently, following two correct responses, the duration of the gap was reduced by 1.25 ms and duration was increased by 1.25 ms after an incorrect response. The task was stopped after eight reversals. A geometric mean of the mid-points of the last six reversals was taken as the gap detection threshold. The measurement was repeated twice for all the participants and the average of the two runs was considered as the final gap detection threshold.

2.2.4.2.2 Frequency Modulation Difference Limen (FM Detection Task)

The sinusoidal FM signals were created using a custom written MATLAB script. The audio signals were sampled at 44100 Hz. The carrier frequency of the signal was 500 Hz. The modulation frequency was always kept at 5 Hz (modulation cycle period of 200 ms) with an initial modulation index of 10 Hz resulting in frequency variation of 10 Hz around the center frequency. The modulation index was subsequently varied based on the participant's responses. All the FM sounds were 500 ms in duration with a cosine-square rise/fall time of 10 ms and were presented at a constant level of 75 dB SPL. The time domain waveform of frequency modulated signal is represented as:

$$x(t) = \sin\left(2\pi f_c t + \frac{\Delta f}{f_m}\sin(2\pi f_m t)\right)$$

where 'fc' is carrier frequency, 'fm' is modulation frequency and 'Δf' is peak frequency deviation.

To estimate the frequency modulation differential limens (FMDLs) or FM detection thresholds, a 2-down-1-up adaptive psychophysical procedure and a two alternate forced-choice (2AFC) task were used to estimate the 70.7% point on a psychometric function. In this procedure, each trial consisted of two 500 ms sounds of which one was frequency modulated and the other was a 500 Hz pure-tone separated by an inter-stimulus interval of 1s. A participant's task was to attend to both of the sounds which were presented sequentially (1 and 2) and to click on the options 1 or 2 on a computer screen to identify which one they thought contained the FM signal. After each response, accuracy feedback was provided. The starting value for the Δf was 25 Hz for both younger and older adults. Subsequently, the Δf was reduced by a factor of 1.25 Hz following two consecutive correct responses and Δf was increased by the factor of 1.25 Hz after an incorrect response. The experiment stopped after eight reversals. A geometric mean of the mid-points of the last six reversals was taken as the FM detection threshold. The measurement was repeated twice for each participant and the average of the two runs was considered as the final FM detection threshold.

2.2.5 Cognition

2.2.5.1 Working Memory

2.2.5.1.1 The Digit Span: Forwards and Backwards

Participants completed two simple span tasks from the Wechsler Adult Intelligence Scale –Fourth Edition (IV): Digits Forwards and Digits Backwards (Wechsler, 2008). These tasks measure short-term memory (i.e., temporary storage of information) and working memory (i.e., storage and processing of information). For the Digits Forwards task, the experimenter verbally presented digit sequences for the participant to immediately recall. After one practice trial, two trials were presented with a sequence length of two. If both or one of the trials were recalled correctly, two trials from the next sequence length (3) would be administered. This procedure continued until two incorrect answers were given for a particular sequence length, or the participant correctly recalled both trials with the maximum sequence length of 9. The score corresponded to the sum of correctly recalled trials. The maximum score was 16. The Digit Backwards Task was administered in the same manner except participants were instructed to recall the digits in the opposite order from what had been spoken. The maximum score was also 16.

2.2.5.1.2 The Listening Span (LSPAN)

The Listening Span was used to measure working memory capacity (Conway et al., 2005). For this task, participants listened to letter sequences ranging from 3-7 letters. Each letter in a sequence was preceded by an auditory semantic categorization task in which a sentence was presented (e.g. the train sang a song) and the participant judged whether the sentence made sense or not. At the end of each sequence, participants were instructed to recall each letter from that sequence using a letter matrix. The researcher performed all of the mouse clicking during the task while the participant provided oral responses (i.e., true, false, and letter sequences). Participants were instructed to adjust the volume to a comfortable level during the practice session. The LSPAN was calculated as the sum of all perfectly recalled sequences (i.e., the absolute scoring method). For example, if an individual recalled 2 letters in a set of 2, 3 letters in a set of 3, and 4 in a set of 5, their absolute score would be 5 (i.e., 2+3+0). The maximum score was 75.

2.2.5.2 Attention

2.2.5.2.1 The Test of Everyday Attention (TEA)

Two subsets of the Test of Everyday Attention (TEA) were administered (Robertson, Ward, and Ridgeway, 1994). Version one was always used for each subset. Practice sessions for each subset were completed according to the TEA instruction manual.

2.2.5.2.1.1 Subset Six Telephone Search (Selective Attention)

Participants visually searched as quickly as possible for specific symbols (i.e., pairs of circles, squares, and stars) on a telephone directory. The number of correctly identified symbols was recorded and the time taken/ symbol was calculated.

2.2.5.2.1.2 Subset Seven Telephone Search While Counting (Sustained Attention)

Participants performed the telephone search task (using a different version of the telephone directory). At the same time, they performed the elevator counting task (with different sequences of tones). The number of correctly identified symbols was recorded and the time taken/symbol was calculated. It was also possible to compare scores from the original telephone search task and calculate a dual task decrement as outlined in the TEA manual.

2.2.5.2.2 The Integrated Visual and Auditory Continuous Performance Test (IVA+)

A version of the IVA+ Continuous Performance Task available in the Millisecond Software library was used to measure auditory and visual sustained attention (Inquisit 5, 2016; Borchert, 2018). The IVA+ is a computer based "go - no go" task. Following a practice run, participants complete 500 trials that consist of either the number one (go cue) or two (no go cue) presented in a pseudorandom order in visual and auditory modalities (i.e., if the participant sees or hears a "one" they click the mouse and if the participant sees or hears a "two" they do not click anything) . The number and latency of hits, misses, false alarms, and correct rejections for auditory and visual trials are recorded. A measure of auditory sustained attention (i.e., the Auditory Attention Quotient) is calculated by adding the Vigilance Auditory Scale Percent (i.e., VIA_perc) to the Mean Auditory Response Time (MNA) and the Focus Auditory Scale (i.e., FOCA):

[(100-((number of misses for auditory trials/45) * 100) + (mean latency of hits across all auditory test trials) + (1-((SD of auditory hit latencies /mean hit latency for auditory trials) *100)]

The Visual Attention Quotient is calculated with the same equation, but with values (e.g., latencies) from visual (not auditory) trials. A higher score indicates better sustained attention ability.

2.2.5.3. Processing Speed

The digit symbol substitution task (Wechsler, 1981) was used as a measure of processing speed. Participants were given a piece of paper and a pen. At the top of the page, the numbers 1-0 are listed with a symbol below each number. The rest of the page consists of a list of numbers with blank spaces. Participants were instructed to draw the symbol associated with each number in the blank spaces. The score is the number of correct symbols drawn in 90 seconds.

2.2.6 Life-Experience

2.2.6.1 Social Economic Status (SES)

SES was operationalized by the sum of scores from two likert scales concerning self and paternal/maternal education levels. The scales for each item ranged from 1 to 4 (highest education level achieved: middle school, high school/equivalent, college, or graduate/professional), so final scores ranged from 2-8.

2.2.6.2 Physical Activity

Physical activity was measured using the General Practice Physical Activity Questionnaire (National Health Services, 2013). Participants rated the amount of physical activity involved in their work and recreation. Scores ranged from 0 to 19, with 19 indicating the highest level of physical activity.

2.2.6.3 Intellectual Engagement

Intellectual engagement was measured by evaluating self-reported engagement in eight activities (cross-words, Sudoku, Scrabble™, "other word games," chess, reading, computer games, and "other"). Response options ranged from 0 (never) to 2 (regularly). The summed score was used, with scores ranging from 0 to 16 (Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013).

2.2.6.4 Musical Experience

Participants were asked to report if they played any musical instruments. For those who responded "yes" musical ability was assessed based on responses to the following questions: "How many instruments do you play?" "How many years of musical training do you have?" and "Do you still play music regularly?".

2.2.6.5 Hearing and Listening Experiences

A short form of the Speech, Spatial, and Qualities of Hearing Scale (i.e., the SSQ 12) was also included as a measure of participants' subjective hearing and listening experience (Gatehouse & Noble, 2004; Noble, Jensen, Naylor, Bhullar & Akeroyd, 2013). The SSQ12 involves rating perceived listening difficulty in real life situations on a scale from 0 to 10 for 12 items. Nine subscales have been identified within the SSQ12: Speech in Quiet, Speech in Noise, Speech in Speech Contexts, Multiple Speech Stream Listening, Localization, Distance and Movement, Segregation, Identification of Sound, Quality and Naturalness, and Listening Effort.

2.2.7 Testing Procedure

First, participants provided informed consent and completed a questionnaire booklet that asked questions about age, sex, native language and the various life-experience factors of interest. The researcher then proceeded to administer the paper based and computer-based tasks. The order of the type of tasks administered first (paper or computer based) was counterbalanced across participants; however, the administration order within each task type was not controlled. Audiograms were typically completed last. Participants were allowed to take a break at any point during the testing session. Sessions took approx. 2.5-3 hours.

2.3 Results and Discussion

The main aim of this study was to characterise younger and older adults' performance on a series of tasks measuring different hearing, cognitive, and life-experience factors that have been previously shown to be related to performance on speech perception in noise tasks. Descriptive statistics for younger and older adults for each task were calculated using JASP 0.10.2 (JASP Team, 2019) and are shown in Table 2.1. Age group differences for each task were also assessed using JASP 0.10.2. Classic independent t-tests (two-tailed) were used for each task; if the results indicated that there was no significant difference between age groups (i.e., $p > .05$) a Bayesian independent samples t-test was used quantify the evidence in support of the null hypothesis.

Table 2.1

Descriptive Statistics for Auditory, Cognitive, and Life-Experience Factors

		Younger (n=30)			Older (n=30)		
Skill or Experience	Task	M (SD)	Min.	Max.	M (SD)	Min.	Max.
Listening	BASTE	-1.35 (0.92)	-3.60	1.10	0.00 (1.03)	-1.90	2.21
Hearing	PTA_{MOD}	9.22 (4.41)	0.00	20	25.72 (8.97)	10.00	50.00
	PTA_{HIGH}	5.83 (5.14)	-2.50	20	40.33 (19.37)	10.00	72.50
Temporal Processing	Gap	2.53 (1.37)	1.30	7.21	5.93 (3.49)	2.55	13.69
	FM	6.05 (5.16)	1.76	23.58	12.59 (10.26)	2.24	38.11
Working Memory	LSPAN	25.37 (15.01)	2.00	61.00	11.27 (9.28)	0.00	31.00
	DS-Back	8.37 (2.24)	4.00	15.00	8.20 (1.95)	4.00	12.00
Short Term Memory	DS-Forward	10.20 (2.14)	8.00	6.00	10.27 (2.70)	16.00	16.00
Selective Attention	TEA 6 [1]	2.61 (0.43)	1.81	3.39	3.70 (1.54)	1.70	9.07
Sustained Attention	TEA 7 [1]	3.01 (0.63)	1.89	4.29	4.80 (1.66)	2.36	10.00
	AAQ	744.06 (80.98)	593.92	930.96	759.32 (81.60)	575.30	889.10
	VAQ	595.73 (51.57)	526.51	795.31	650.94 (51.83)	551.00	755.86
Processing Speed	DSST	46.47 (7.18)	30	64	32.27 (8.83)	14	50
Physical Activity	Self Report	-1.35 (0.92)	-3.60	1.10	0.00 (1.03)	-1.90	2.21
Intellectual Engage.	Self Report	9.22 (4.41)	0.00	20.00	25.72 (8.97)	10.00	50.00
SES	Self Report	5.83 (5.14)	-2.50	20.00	40.33 (19.37)	10.00	72.50

Note. [1] = the time/target score was used for the Test of Everyday Attention.

A secondary aim of this study was to test whether any of the hearing, cognitive, and life-experience factors measured were associated with older adults' performance on the BASTE speech perception in noise task. Two approaches were used to investigate these relationships. First, consistent with previous research, Pearson correlations between each factor and performance on the BASTE speech perception in noise task were calculated (See Table 2.2). Second, estimation statistics (Ho, Tumkaya, Aryal, Choi, Claridge-Chang, 2018) were computed to evaluate whether there was a meaningful difference in performance for any of the tested factors between older adults who performed best on the BASTE (i.e., 25th percentile) and older adults who performed worst on the BASTE (i.e., 75th percentile). It should be noted that the results obtained from both of these analyses should be interpreted with discretion as the relatively small sample sizes (i.e., 30 participants per age group) may not have accurately captured the populations' distribution (Efron, 2000).

Table 2.2

Pearson correlations with speech in noise (BASTE) performance for older adults

Skill or Experience	*Task*	*Pearson's r*	*p value*
Hearing	PTA_{MOD}	0.28	0.14
	PTA_{HIGH}	0.32	0.09
Temporal Processing	Gap	0.22	0.26
	FM	0.11	0.57
Working Memory	LSPAN	0.02	0.90
	DS-Back	0.20	0.28
Short Term Memory	DS-Forward	0.03	0.89
Selective Attention	TEA 6 [1]	-0.18	0.35
Sustained Attention	TEA 7 [1]	-0.27	0.15
	AAQ	**-0.37**	**0.04***
	VAQ	-0.31	0.09
Processing Speed	DSST	0.03	0.87
Physical Activity	Self Report	0.06	0.76
Intellectual Engage.	Self Report	0.11	0.56
SES	Self Report	-0.15	0.44

Note. [1] = a time/target score was used for the correlation, * = $p < .05$.

As can be seen in Table 2.2, the correlational analysis indicated that the only task significantly related to BASTE performance for older adults was the index of auditory sustained attention (i.e., the Auditory Attention Quotient of the Continuous Performance Task). Figure 2.3 shows this correlation, as well as the correlation between the Visual Attention Quotient and the BASTE, which approached significance ($p = 0.09$). The AAQ correlation for the current study is consistent with Anderson, White-Schwoch, Parbery-Clark, & Kraus' (2013) results which show a moderate correlation (0.265, $p < 0.01$) between older adults' AAQ scores and speech perception thresholds (Note: the difference in directionality between our study and Anderson et al. (2013) is due to transformations applied to Anderson et al.'s (2013) data; both studies suggest that as AAQ performance improves, speech perception in noise performance improves).

The absence of a significant correlation between the LSPAN and the BASTE is inconsistent with Gordon-Salant & Cole's (2016) results, which showed a strong relationship between sentence recognition in noise and LSPAN ($r = -.70$, $p < .01$). In contrast to Gordon-Salant & Cole (2016) (but consistent with the current study), Schuman, Brungart, & Gordon-Salant (2014) showed that the LSPAN did not accurately predict SRTs on a sentence recognition in noise task when the task involved immediately recalling the most recently spoken sentence; however, when the task required a previously presented sentence to be recalled (i.e., 1-back) the LSPAN was predictive of SRTs (Energetic Noise: $r = 0.707$, $p < 0.01$; Two-Talker Noise r = 0.690, $p < 0.01$; Two-Talker Spatialised Noise: $r= 0.642$, $p < 0.01$, 1 Talker Noise $r = 0.593$, $p < 0.01$). Schuman, Brungart, & Gordon-Salant (2014) suggested that simplified immediate recall type tasks that are typically used to assess speech perception in both research and clinical settings may not be sensitive to the contributions of working memory to speech perception in noise.

The correlational results for the current study are consistent with Gordon-Salant and Cole (2016) and Anderson et al. (2013) in that pure-tone audiometry from moderately-high frequencies (i.e., 0.5-4 kHz) was not significantly associated with speech recognition in noise performance (Anderson et al. (2013): $r = 0.11$ Gordon-Salant & Cole, 2016: $r = .05$; but see Heinrich, Henshaw, & Ferguson (2015) for a significant PTA correlation, $r = .39$, $p < .01$). As the high frequency PTA correlation approached significance, this result is consistent with studies suggesting that high frequency hearing loss is predictive of speech perception in noise ability (Besser, Festen, Goverts, Kramer, Pichora-Fuller, 2015).

The r values for both the digit span and TEA tasks are smaller than those reported by other studies that have tested the relationship between these tasks and sentence recognition in noise for older adults. Heinrich, Henshaw, and Ferguson (2015), for example, found moderately strong r values for the TEA (Subset Six: $r = -.38$, $p < .01$), Subset Seven: $r = -.29$, $p < .01$) and digit span (Forwards: $r = -.20$, $p > .05$, Backwards: $r = -.32$, $p < .05$). Additionally, Füllgrabe, Moore, & Stone (2015) found strong r values for the digit span (Forwards: $r = .76$, $p > .01$, Backwards: $r = .60$, $p < .01$) and moderately-strong r values for the TEA (Subset Six: $r = 0.42$, $p > .05$, Subset Seven: $r = 0.47$, $p < .05$). Several factors could be contributing to the differing correlations found across studies including noise type (informational vs. energetic) and level, task procedure (adaptive threshold vs. percent correct), and hearing ability. Regardless of why some correlations may not be secure, the significant correlation between the AAQ and the BASTE suggests that the role of attention in speech perception in noise deserves to be considered in future research.

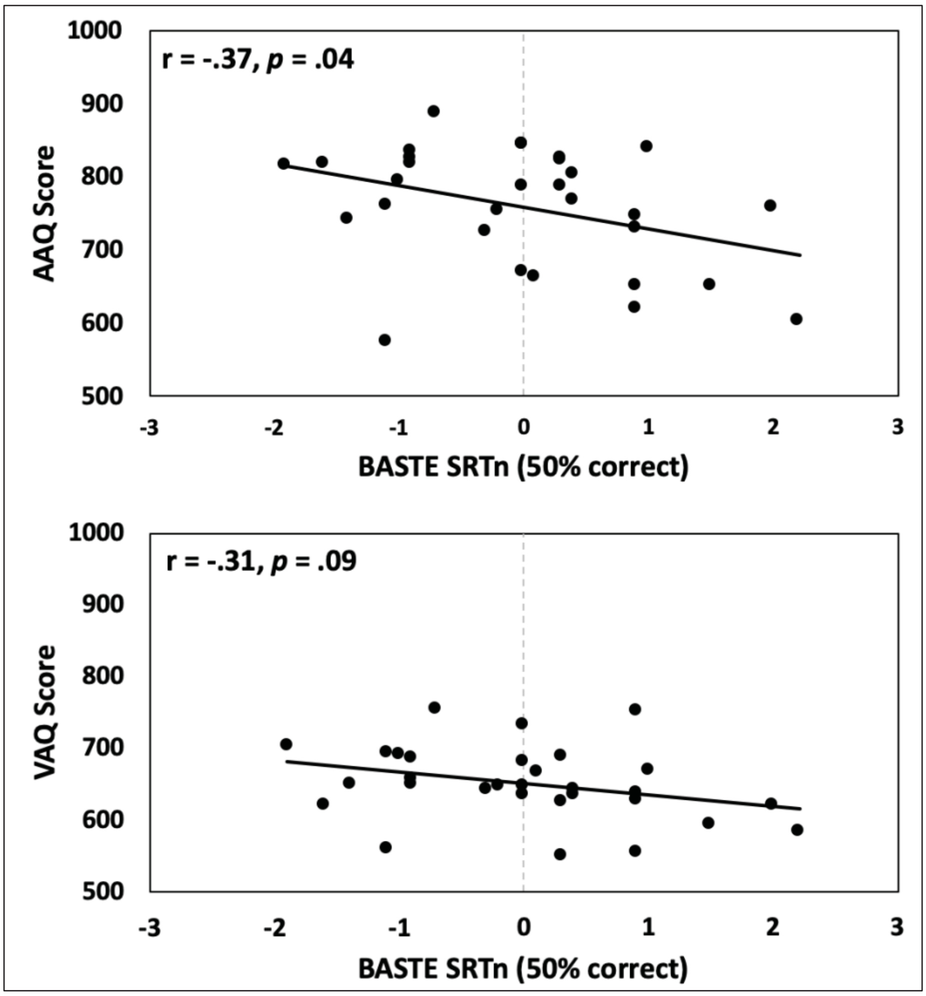

Figure 2.3. Pearson Correlations Between Sustained Attention Measures (Auditory and Visual) and Speech Reception in Noise Thresholds.

The second approach used to investigate the relationships between older adults' performance on hearing, cognitive, and life-experience factors and performance on the BASTE speech perception in noise task was to use the data analysis with bootstrap estimation (DABEST) package with R software to evaluate whether there was a meaningful difference in performance between older adults who performed best on the BASTE (i.e., 25th percentile) and older adults who performed worst on the BASTE (i.e., 75th percentile) for any of the tested factors. Estimation statistics represent a framework that avoids the major limitations of null hypothesis significance testing (e.g., an accept/reject dichotomy) by focusing on the size and precision of an effect. For this approach, bootstrap resampling creates multiple resamples of a given set of observations and computes the effect size for each resample. These effect sizes are then used to determine a 95% confidence interval. If the mean difference between a test population and control population is outside this 95% confidence interval, then a meaningful difference between groups is inferred (Ho, Tumkaya, Aryal, Choi, Claridge-Chang, 2018).

Using the DABEST R package, the mean unpaired difference between older high and low BASTE performers for each factor was computed and Gardner-Altman estimation plots were produced to depict the results (Ho, Tumkaya, Aryal, Choi, Claridge-Chang, 2018). This analysis computes 5000 bootstrap resamples and confidence intervals that are bias corrected and accelerated. The next section presents the results from the DABEST analysis and the between groups analysis (i.e., t-tests) for each task administered.

2.3.1 Speech Perception in Noise

As can be seen in Figure 2.4, younger adults had significantly lower speech reception thresholds (SRTs) at 50% keywords correct than older adults ($t(58) = -5.36$, $p < .001$, $d = -1.39$) with lower SRTs indicating better performance. Older adults' SRTs were variable (*M* = 0.00, *SD* = 1.03, *Min*. = -1.90, *Max.* = 2.21), with some older adults performing at a similar level to younger adults, and others having higher SRTs than younger adults. These results are consistent with research suggesting that although older adults do seem to have greater difficulty than younger adults at identifying speech in noise overall, there is also considerable variability in older adults' performance on standard speech perception in noise tasks (e.g., Füllgrabe, Moore & Stone, 2015; Schoof & Rosen, 2014).

There was also considerable variability in younger adults' performance on the BASTE task (*M* = -1.35, *SD* = 0.92, *Min*. = -3.60, *Max.* = 1.10). One younger participant performed almost two standard deviations higher than the mean SRT for younger adults, with an SRT of 1.10, and a different younger adult performed two standard deviations lower than the mean SRT for younger adults, with an SRT -3.60. The performance from these participants in particular (and younger adults overall) is consistent with studies showing that, even for healthy younger adults, there is variability in performance on standard speech perception in noise tasks (Goossens, Vercammen, Wouters, & van Wieringen, 2017; Schoof & Rosen, 2014). The auditory, cognitive, and life-experience factors that could be contributing to this variability for younger and older adults are explored in the next sections of this chapter.

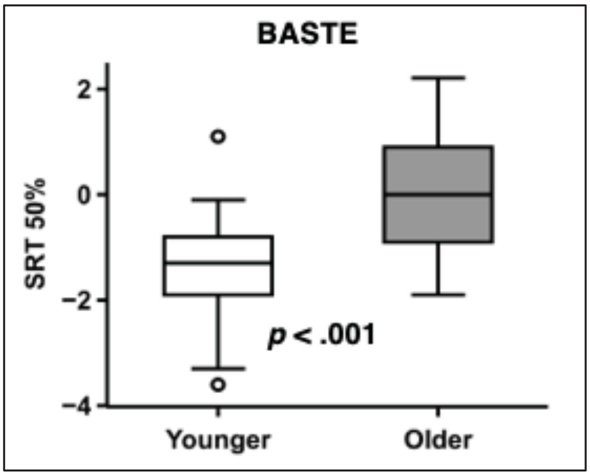

Figure 2.4. Speech Reception Thresholds (SRTs) for the BASTE Task. Tukey's box plots represent the median and interquartile range (Q3-Q1) of the SNR at which participants understood 50% of the keywords in the BASTE sentences (i.e., SRT 50%) as a function of Age. Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range.

2.3.2 Hearing

Table 2.2 summarises the hearing sensitivity levels for both younger and older adults (derived from pure-tone audiograms). All younger participants had normal hearing (i.e., ≤ 25dB HL at .25, .5, 1, 2, 4 kHz). Older adults' hearing levels were more diverse, ranging from normal to moderate-severe hearing loss (i.e., > 55dB ≤ 70dB HL at one frequency in the better ear), with the majority of older participants (i.e., 13) having only mild hearing loss (i.e., > 25dB and ≤ 40dB HL at .25, .5, 1, 2, 4 kHz in the better ear). Pure-tone hearing thresholds for each tested frequency are summarised in Figure 2.5. Younger adults had significantly lower thresholds than older adults for both ears at all tested frequencies (all *p* values ≤ .001).

Table 2.3
Hearing Sensitivity Levels Derived from Pure-Tone Audiogram Thresholds

Hearing Level	Definition		
		Younger (n=24)	Older (n=30)
Normal	≤ 25[1] at all frequencies[2]	30	6
Mild Loss	>25 ≤ 40 at one frequency	0	18
Moderate Loss	> 40 ≤ 55 at one frequency	0	4
Moderate-Severe Loss	> 55 ≤ 70 at one frequency	0	3

Note. Hearing level definitions adapted from Wayne et al., 2016 and are measured from the better ear.
[1]dB Hearing Loss
[2]All frequencies refers to .25, .5, 1, 2, 4 kHz

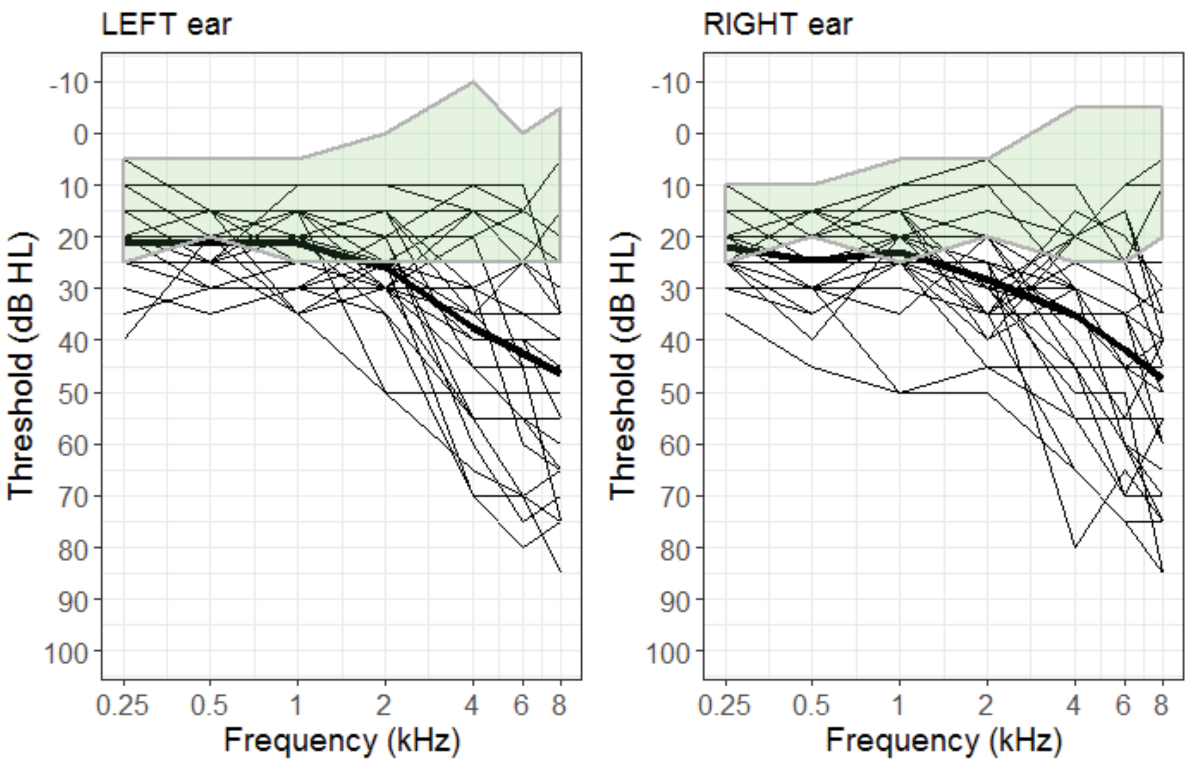

Figure 2.5. Audiogram Results for the Left and Right Ears. The bold black line represents the mean threshold for older adults as a function of frequency. The fine black lines represent individual audiograms for older adults as a function of frequency. The green shaded area represents the audiometric threshold range for younger adults.

Figure 2.6 (A & C) shows the mean thresholds (dB HL) for moderately-high frequencies (1, 2, and 4 kHz) and high frequencies (6 & 8 kHz), for younger and older adults.

As can be seen in Figure 2.6, younger adults had significantly less hearing loss (i.e., lower thresholds) than older adults for both moderately-high and high frequencies (Moderate: $t(42.28) = -9.04$, $p < .001$, $d = -2.35$, High: $t(33.07) = -9.43$, $p < .001$, $d = -2.53$).

The distribution of older adults' hearing thresholds was particularly wide-spread for the higher frequencies (i.e., 6 & 8 kHz, *Min.* = 10.00, *Max.*= 72.50). This suggests that although some of the other adults tested were experiencing hearing loss at high frequencies, others had very minimal high frequency loss. As discussed in the introduction, life-experience factors (e.g., occupational noise exposure, exercise, and smoking) contribute to the progression and severity of high frequency hearing loss across the lifespan (Gates & Mills, 2005; Yamasoba et al., 2013). These life-experience factors, in addition to age, are likely contributing to the variability in high frequency hearing loss for the older participants in the current study.

Figure 2.6 (B & D) shows mean hearing thresholds for moderately-high frequencies (1, 2, & 4 kHz) and high frequencies (6 & 8 kHz) for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation (i.e., DABEST) is also shown in Figure 2.6 (B & D). This analysis indicates that there was no meaningful difference in mean hearing thresholds between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group for both moderately-high frequency thresholds (95 CI: -13.9; 2.67) and high frequency thresholds (95 CI: -27.1; 4.69).

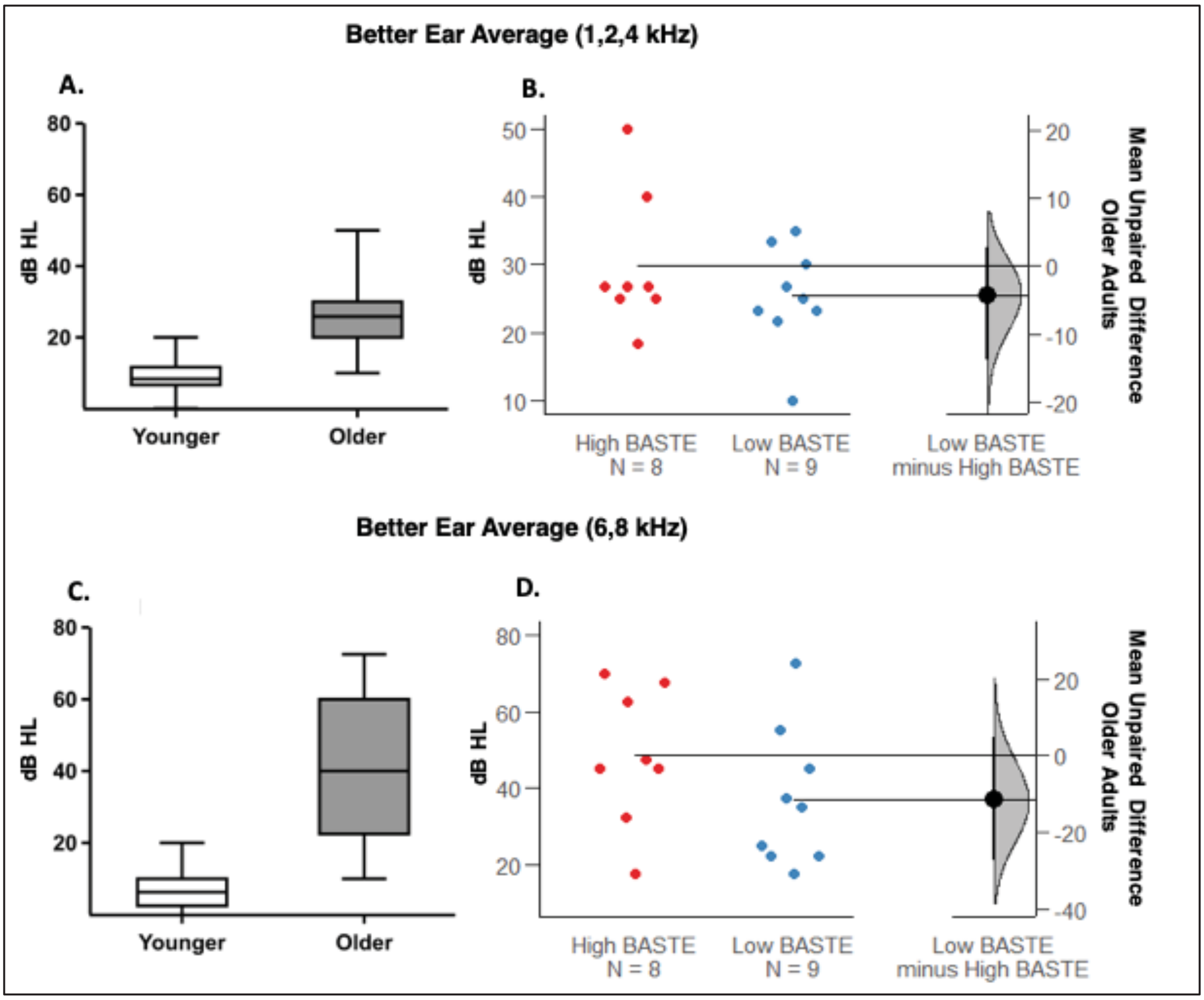

Figure 2.6. Mean Hearing Thresholds for Moderately-High and High Frequencies. Panels A and C show mean hearing thresholds (dB HL) for moderately high frequencies (A) and high frequencies (C) as a function of age group. Tukey's box plots represent the median and interquartile range (Q3-Q1). Panels B and D show Gardner-Altman estimation plots of moderately-high frequencies (B) and high frequencies (D) for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE). Individual scores are plotted in red for High BASTE older adults and blue for Low BASTE older adults. The complete distribution of the bootstrapped mean difference between the Low BASTE group and the High BASTE group is represented by the curve filled in grey; the horizontal black line passing through the circle filled in black represents the mean of this distribution and the vertical black line represents the 95% confidence interval of this distribution. The mean difference between the High BASTE group and the High BASTE group is represented by the horizontal black line at zero.

Figure 2.6 (B) shows that there was one older participant in the High BASTE group that had higher moderately-high mean hearing thresholds (i.e., 50) than the rest of the older adults in the High BASTE group. This same participant also had the highest high frequency hearing threshold for the High BASTE group (i.e., 70). However, these high hearing thresholds did not seem to have a large effect on this participant's performance overall, as this participant had average or better than average scores for all cognitive and temporal processing tasks. It is possible that this participant's hearing thresholds were high due to participant error (i.e., not understanding the task). However, it is also possible that these hearing thresholds are accurate, and despite these higher thresholds, this older adult was able to perform well on the other tasks involved in the study.

Together, the results from the DABEST analysis suggest that performance on a pure-tone audiometry assessment does not explain the variance in older adults' performance on a standard speech perception in noise test. That is, older adults who performed differently on the BASTE (i.e., 25th percentile vs. 75th percentile) did not have meaningfully different moderately-high or high frequency hearing thresholds. This result is consistent with other studies that have not found strong relationships between pure-tone audiometry and speech perception in noise performance for older adults (Dubno 1984; Duquesnoy 1983; Jerger 1992; Vermiglio, Soli, Freed, & Fisher, 2012). In contrast, this finding does not fit with studies which have suggested that high frequency hearing thresholds are predictive of older adults' speech perception in noise ability (Besser, Festen, Goverts, Kramer, Pichora-Fuller, 2015).

2.3.3 Temporal Processing

Figure 2.7 (A & C) shows gap detection (A) and frequency modulation (C) thresholds for younger and older adults. As can be seen in Figure 2.7 (A & C), there was greater variability in both gap detection and frequency modulation thresholds for older adults (Gap: *M* = 5.93, *SD* = 3.49, *Min*. = 2.55, *Max.* = 13.69, Frequency: *M* = 12.59, *SD* = 10.26, *Min*. = 2.24, *Max.* = 38.11) in comparison to younger adults (Gap: *M* = 2.53, *SD* = 1.37, *Min*. = 1.30, *Max.* = 7.21, Frequency: *M* = 6.05, *SD* = 5.16, *Min*. = 1.76, *Max.* = 23.58).

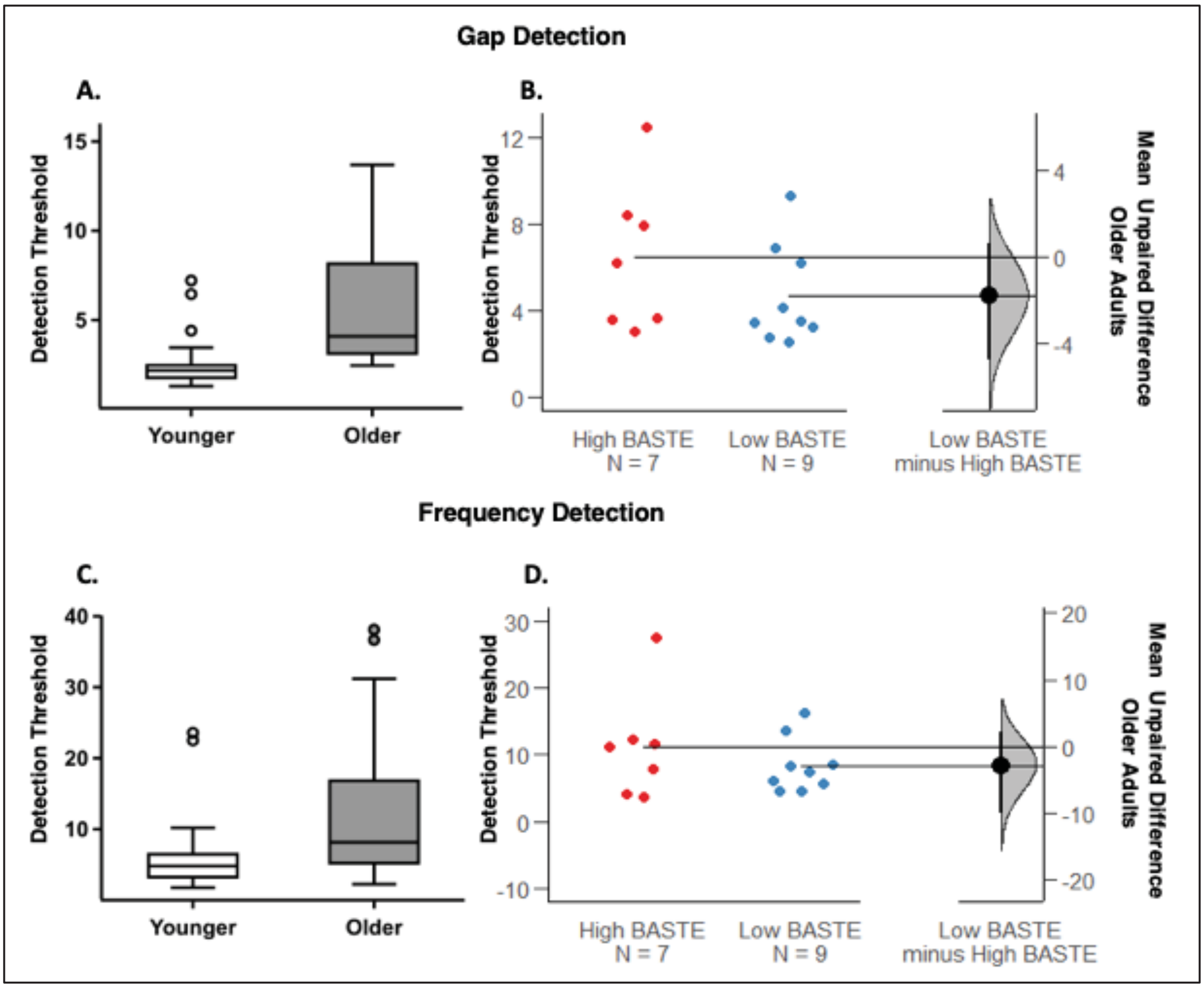

Figure 2.7. Panels A and C show mean detection thresholds for Gap Detection (A) and Frequency Modulation Detection (C) as a function of age group. Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panels B and D show Gardner-Altman estimation plots of Gap Detection (B) and Frequency Modulation Detection (D) for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE). Notably, data from one older adult in the High BASTE group is missing (N = 7) due to technology failure.

Independent samples t-tests indicated that younger adults had significantly lower detection thresholds than older adults for both the gap detection and frequency modulation detection tasks, where lower detection thresholds indicate better detection abilities (Gap: $t(35.06) = -4.80$, $p < .001$, $d = -1.28$, Frequency: $t(39.85) = -2.97$, $p = .01$, $d = -0.79$). These results are consistent with research that suggests temporal processing abilities decline in old age (Pichora-Fuller & MacDonald 2007; Vermeire, 2016).

Figure 2.7 (B & D) shows the mean detection thresholds for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.7 (B & D). This analysis indicates that there was no meaningful difference in detection thresholds between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group for both gap detection (95 CI: –4.73; 0.653) and frequency modulation detection thresholds (95 CI: -10; 2.18).

For both gap detection and frequency modulation detection tasks, there was one older adult in the High BASTE group that had a higher detection threshold than the other members of the High BASTE group. Notably, this was a different participant for the gap detection and frequency modulation detection tasks respectively. It is possible that for these participants, poorer temporal processing abilities contributed to their poorer performance on the BASTE task (and thus their inclusion in the High BASTE group). When these outliers are not considered, the range of detection thresholds for High and Low BASTE older adults are similar within the gap detection task (i.e., High BASTE: *Min.* = 3.02. *Max.* = 8.41, low BASTE: *Min.* = 2.53, *Max.* = 9.30) and within the frequency modulation task (i.e., High BASTE: *Min.* = 3.68, *Max.* = 12.39, Low BASTE: *Min.* = 4.59, *Max.* = 6.35). That is, performance on neither of the temporal processing tasks seemed to account for the variance in BASTE performance for the older adults included in this study.

2.3.4 Processing Speed

Figure 2.8 (A) illustrates younger and older adults' performance on the digit symbol substitution task (DSST). As displayed in Figure 2.8 (A), performance on the DSST varied for both age groups, with scores ranging from 30.00-64.00 items correct for younger adults and 14.00-50.00 items correct for older adults, where more items correct indicates better performance. One younger adult outperformed all other younger and older adults with a score of 64.00 items correct. An independent samples t-test indicated that younger adults performed significantly better than older adults ($t(58) = 6.84$, $p < .001$, $d = 1.77$). This supports the proposal that information processing is slower in old age (Salthouse, 1996).

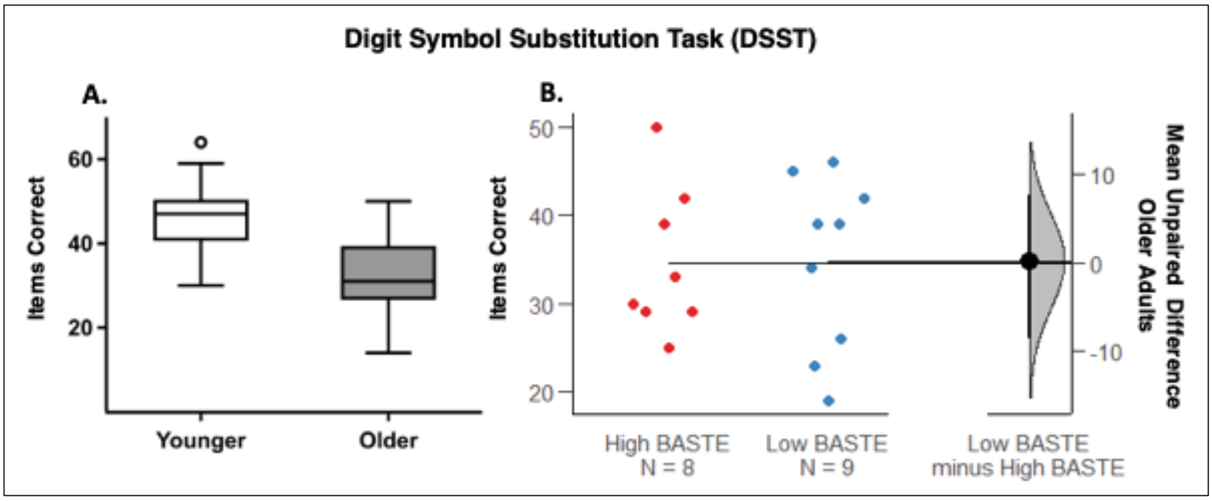

Figure 2.8. Accuracy Scores for the Digit Symbol Substitution Task (DSST). Panel A shows accuracy scores for the DSST as a function of age group. Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panel B shows Gardner-Altman estimation plots of the DSST for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) to older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE).

Figure 2.8 (B) shows DSST performance for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). There were two older adults in the Low BASTE group (i.e., the group who performed best on the BASTE) who performed the poorest on the DSST. This is inconsistent with Processing Speed Theory, which would predict that individuals with better processing speed (i.e., higher performance on the DSST) would perform better on a standard speech perception in noise task (i.e., have lower SRTs; Helfer, 2014; Salthouse, 1996).

A summary of the data analysis with bootstrapped coupled estimation is also shown in Figure 2.8 (B). This analysis indicates that there was no meaningful difference in DSST performance between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group (95 CI: -8.60; 7.60). Indeed, DSST scores for High BASTE older adults (*Min*. = 25.00 *Max.* = 50.00) and Low BASTE older adults (*Min*. = 19.00, *Max.* = 45.00) were similarly distributed across a large range of possible scores. This suggests that the DSST may not be a very precise measure of processing speed for older adults.

2.3.5 Memory

2.3.5.1 Listening Span

Figure 2.9 (A) shows younger and older adults' performance on the listening span (i.e., the LSPAN). As depicted in Figure 2.9 (A), younger adults' performance on the LSPAN was higher overall yet more variable than older adults (Younger: M = 25.37, SD = 15.01, *Min*. = 2.00, *Max.* = 61.00, Older: M = 11.27, SD = 9.28, *Min*. = 0.00, *Max.* = 31.00; $t(48.35) = 4.38$, $p < .001$, $d = 1.13$).

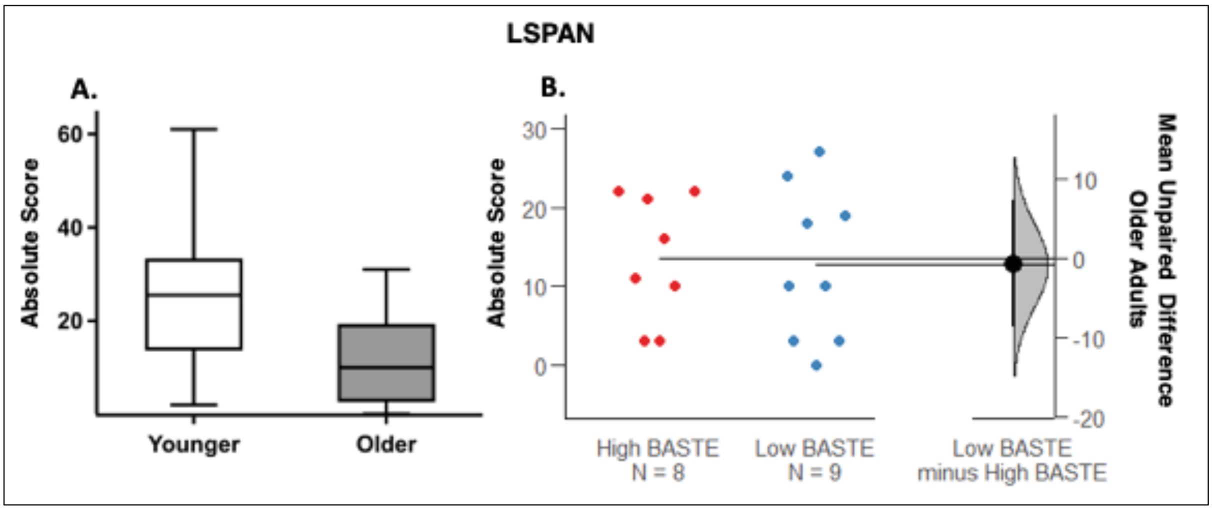

Figure 2.9. Listening Span (LSPAN) Results. Panel A shows LSPAN scores for younger and older adults. Tukey's box plots represent the median and interquartile range (Q3-Q1). Panel B shows Gardner-Altman estimation plots of LSPAN scores for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE).

Figure 2.9 (B) shows LSPAN performance for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.9 (B). This analysis indicates that there was no meaningful difference in LSPAN performance between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group (95 CI: -8.68; 7.36). Indeed, the distributions of LSPAN scores for High BASTE older adults (*Min*. = 3.00, *Max.* = 22.00) and Low BASTE older adults (*Min*. = 0.00, *Max.* = 27.00) were very similar.

Overall, performance on the LSPAN is consistent with research suggesting that working memory capacity reduces in old age and that there is considerable variability in performance on working memory tasks for both younger and older adults (Souza, Arehart, & Neher, 2015). However, the results do not support the proposal that working memory capacity, particularly when indexed by complex span tasks, explains the variance in older adults' performance on speech perception in noise tasks (Akeroyd, 2008; Gordon-Salant & Cole, 2016; Rönnberg et al. 2008, 2013, 2019).

2.3.5.2 Digit Span

Performance on the digit span forward and digit span backward, for both age groups, is depicted in Figure 2.10 (A & C). Independent samples t-tests indicated that there was no significant difference between younger and older adults' performance on the digit span forward ($t(58) = -0.12$, $p = .92$, $d = -0.03$; Younger: $M = 10.20$, $SD = 2.14$, *Min*. = 8.00, *Max.* = 16.00; Older: $M = 10.27$, $SD = 2.70$, *Min*. = 6.00, *Max.* = 16.00) or the digit span backward ($t(58) = 0.31$, $p = .76$, $d = 0.08$; Younger: $M = 8.37$, $SD = 2.24$, *Min*. = 4.00, *Max.* = 15.00; Older: $M = 8.20$, $SD = 1.95$, *Min*. = 4.00, *Max.* = 12.00). Bayesian independent samples t-tests suggested that, for the digit span forward, the data was 3.79 times more likely to occur under the null hypothesis (i.e., that there was no difference between age groups) than the alternative hypothesis (i.e., that there was a difference between age groups), and for the digit span backward, the data was 3.66 times more likely to occur under the null hypothesis (i.e., that there was no difference between age groups) than the alternative hypothesis (i.e., that there was a difference between age groups). For the digit span forward, approximately 70% of younger adults and 70% of older adults scored higher than eight (i.e., 50% correct), and for the digit span backwards, approximately 50% of younger adults and 50% of older adults scored higher than eight.

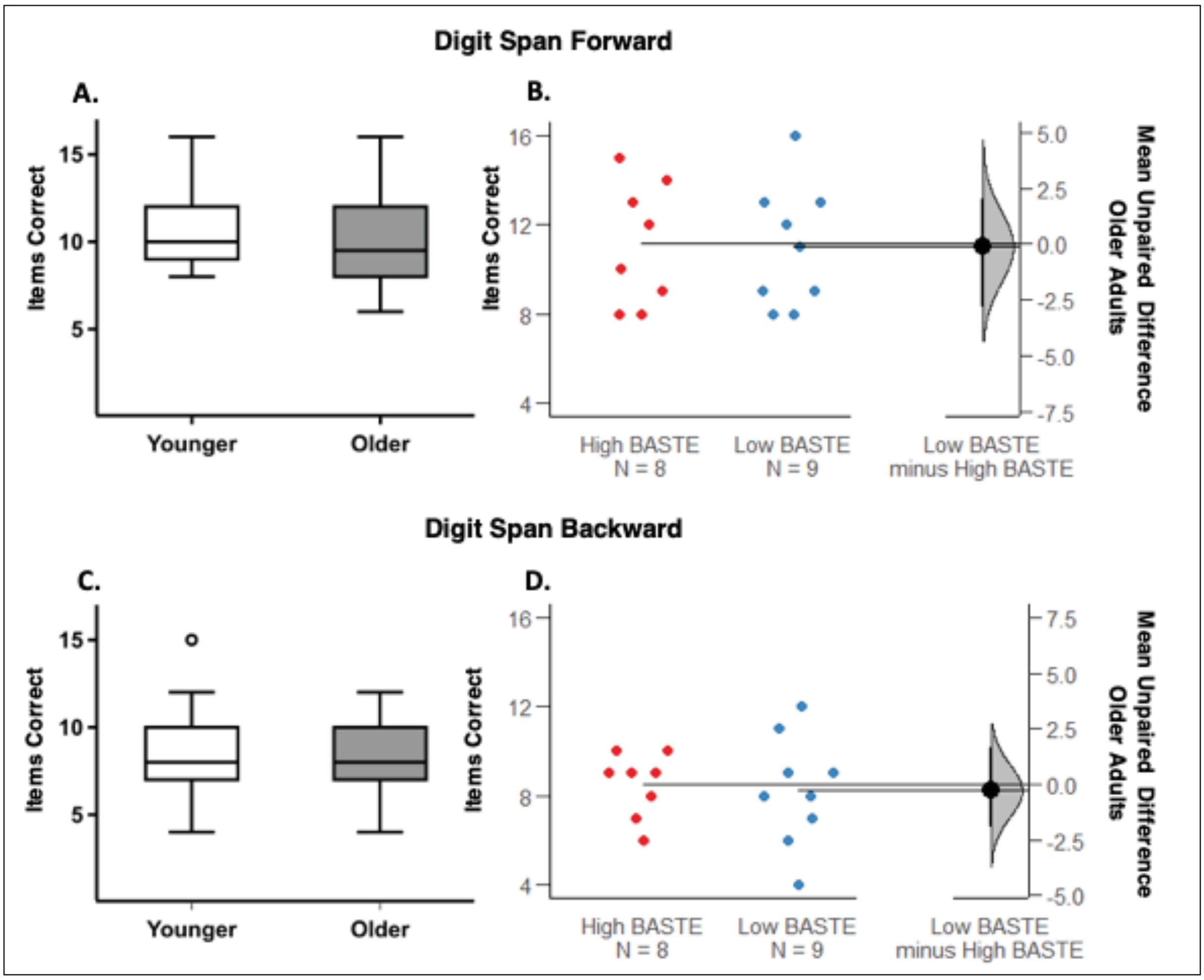

Figure 2.10. Results from the Digit Span Forward and Backward. Panels A and C show total scores for the Digit Span Forward (A) and the Digit Span Backward (C) as a function of age group. Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panels B and D show Gardner-Altman estimation plots of the Digit Span Forward (B) and Digit Span Backward (D) for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE).

Figure 2.10 (B & D) shows the total span scores for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.9 (B & D). This analysis indicates that there was no meaningful difference in total span scores between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95%

confidence interval of the mean difference between the Low BASTE group and the High BASTE group for both digit span forward (95 CI: -2.83; 2.04) and digit span backward (95 CI: -1.90; 1.65) tasks. The results from the DABEST analysis suggest that when operationalised with the digit span forward and backward, short-term memory and working memory may not explain the variance in older participants' performance on the BASTE task.

2.3.6 Attention

2.3.6.1 Test of Everyday Attention (TEA)

Subset Six of the TEA is a measure of selective attention; smaller amounts of time taken per target indicates better selective attention. See Figure 2.11 (A) for a summary of younger and older adults' performance on Subset Six. Younger adults performed faster per target than older adults (Younger: M = 2.61, SD = 0.43, *Min*. = 1.81, *Max*. = 3.39; Older: M = 3.60, SD = 1.25, *Min*. = 1.70, *Max*. = 6.70; $t(36.34) = -4.11$, $p < .001$, $d = -1.06$). This is consistent with other studies that have shown that ageing may affect the way selective attention is deployed (Chapter 3 this thesis; Madden, & Monge, 2019; Zanto & Gazzaley). According to the standard percentiles outlined in the TEA manual, approximately 70% of younger adults and 70% of older adults performed above average for their age group for Subset Six (i.e., a score < 2.60 for younger adults and < 3.60 for older adults).2014). That is, the majority of younger and older adults who participated in the current study had better than average (for their respective age group) selective attention ability (as measured by Subset Six of the TEA).

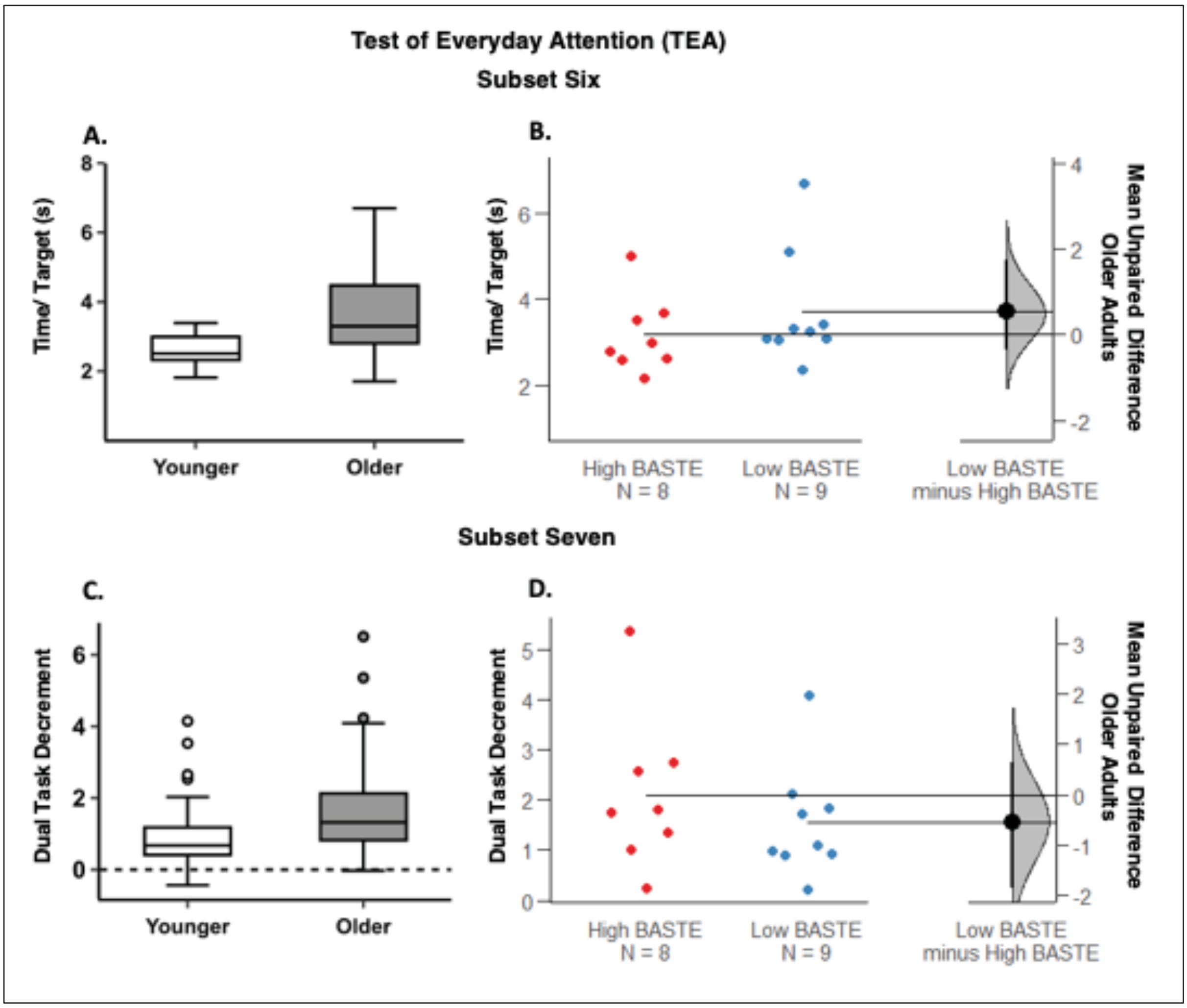

Figure 2.11. Results from The Test of Everyday Attention. Panels A and C depict performance on Subset Six of the TEA (A) and Subset Seven of the TEA (C) as a function of age group. Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panels B and D show Gardner-Altman estimation plots of Subset Six (B) and Subset Seven (D) for older adults who had BASTE scores at or above the 75_{th} percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25_{th} percentile (i.e., Low BASTE).

Subset Seven of the TEA is a measure of sustained attention; a smaller dual task decrement indicates better sustained attention ability. See Figure 2.11 (C) for a summary of younger and older adults' Subset Seven scores. An independent samples t-test suggested that younger adults (Younger: *M* = 0.97, *SD* = 1.06, *Min.* = -0.44, *Max.* = 4.15) had significantly smaller dual task decrements than older adults (*M* = 1.79, *SD* = 1.55, *Min.* = -0.03, *Max.* = 6.51; $t(58) = -2.39$, $p = 0.02$, $d = -0.62$). However, younger and older adults' average performance on Subset Seven for the current study (Younger: 0.97, Older: 1.79) was higher (i.e., poorer) than the standard 50th percentiles outlined in the TEA manual (Younger: 0.5, Older: 1.5). That is, 70% of younger adults in the current study performed below average for their age group (i.e., > 0.5) and 50% of older adults performed below average for their age group (i.e., > 1.5).

Figure 2.11 (B & D) shows the TEA scores for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.11 (B & D). This analysis indicates that there was no meaningful difference in total span scores between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group for both Subset Six (95 CI: 0.343; 1.73) and Subset Seven (95 CI: -0.25; 2.05) of the TEA. The results from the DABEST analysis suggest that for the older adults tested for this study, performance on Subsets Six and Subsets Seven of the TEA (i.e., measures of selective and sustained attention) do not explain the variance in performance on the BASTE speech perception in noise task.

2.3.6.2 IVA + Continuous Performance Task

2.3.6.2.1 Auditory Attention Quotient (AAQ)

The AAQ is a measure of auditory sustained attention based on the IVA+ Continuous Performance Task; higher scores indicate better auditory sustained attention. Younger and older adults' AAQ scores are presented in Figure 2.12 (A). An independent samples t-test indicated that there was no significant difference in AAQ scores between younger adults (M = 774.06, SD = 80.98, *Min*. = 593.92, *Max.* = 930.96) and older adults (M = 759.32, SD = 81.60, *Min*. = 575.30, *Max.* = 889.10); $t(58) = -0.73$, $p = .47$, $d = -0.19$). A Bayesian independent samples t-test suggested that the data was 3.05 times more likely to occur under the null hypothesis (i.e., that there was no difference between age groups) than the alternative hypothesis (i.e., that there was a difference between age groups). As can been seen in Figure 2.12 (A) the majority of participants from both age groups had AAQ scores between 700-800. However, one older adult did have a particularly low AAQ score (i.e., 575.30).

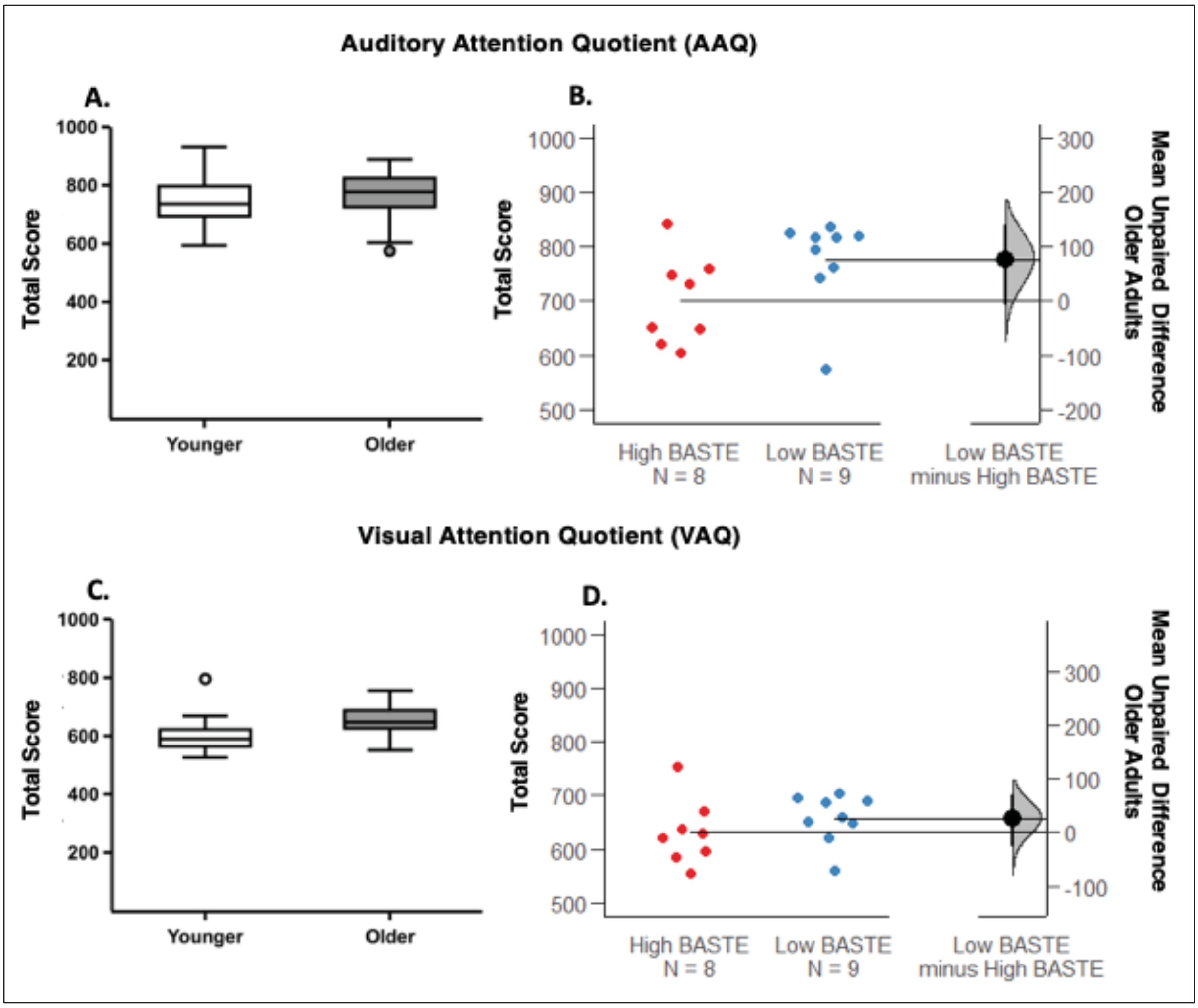

Figure 2.12. Result Summaries for the Auditory Attention Quotient (AAQ) and the Visual Attention Quotient (VAQ). Panels A and C show the AAQ (A) and the VAQ (C) as a function of age group. Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panels B and D show Gardner-Altman estimation plots of the AAQ (B) and the VAQ (D) for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE).

Figure 2.12 (B) shows AAQ scores for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The older adult with the lowest AAQ score is included the Low BASTE group, however, the remainder of the older adults in the Low BASTE group had AAQ scores that were greater than the mean AAQ score of the High BASTE group (i.e., 700.22). Figure 2.12 (B) indicates that the mean difference between the High BASTE group and the High BASTE group (i.e., 0) is approaching the edge of the lower bound of the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group (i.e., -8.39). If the mean difference between the High BASTE group and the High BASTE group was less than the lower bound of the 95% confidence interval, we could have been confident that older adults who performed better on the BASTE had better auditory attention than older adults who performed poorer on the BASTE. This finding would have been consistent with the FUEL (Pichora-Fuller et al., 2016), which would predict that attention moderates speech perception in noise ability.

2.3.6.2.2 Visual Attention Quotient (VAQ)

The VAQ is a measure of visual sustained attention based on the IVA+ Continuous Performance Task; higher scores indicate better visual sustained attention. Younger and older adults' VAQ scores are presented in Figure 2.12 (C). An independent samples t-test indicated that there was a significant difference in VAQ scores between younger adults (M = 595.73, SD = 51.57, *Min*. = 526.21, *Max.* = 795.31 and older adults (M = 650.94, SD = 51.83, *Min*. = 551.00, *Max.* = 755.86); $t(58) = -4.14$, $p < .001$, $d = -1.07$). That is, contrary to the cognitive ageing literature, the VAQ results suggest that older adults had better visual sustained attention than younger adults.

As can be seen in Figure 2.12 (B & D), there was less variation for both High and Low BASTE groups for the VAQ (High: *Min*. = 555.82, *Max.* = 753.88, Low: *Min*. = 559.77, *Max.* = 703.36) in comparison to the AAQ (High: *Min*. = 620.20, *Max.* = 841.11, Low: *Min*. = 575.30, *Max.* = 835.38). As auditory trials for the IVA+ were presented at 440Hz, it is possible that low frequency hearing sensitivity could have contributed to the variance in older adults AAQ scores.

Figure 2.12 (D) also illustrates that there was no meaningful difference between the older adults in the Low BASTE and High BASTE groups for the VAQ. That is, the data analysis with bootstrapped coupled estimation shown in Figure 2.12 (D) indicates that the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group for the VAQ task (95 CI: –25.7; 69.7). This suggests that for this group of older adults, sustained attention, as measured by the VAQ, is not able to explain the variation in speech perception in noise abilities.

2.3.7 Life-Experience Factors

The younger and older adults who participated in this study did not significantly differ on any of the ratings of life-experience factors that have been shown to influence speech perception in noise ability (i.e., physical activity, intellectual engagement, SES, or musical experience). That is, the majority of participants in the study were moderately physically active, moderately intellectually engaged, and had a moderate to high SES. The number of participants who reported that they still played a musical instrument regularly was also not significantly different between groups (i.e., three for younger and four older). This suggests that any age group differences in performance on the BASTE, auditory tasks, or cognitive tasks included in the study are not due to differences in life-experience.

2.3.7.1 Physical Activity

Total scores for the physical activity questionnaire are depicted in Figure 2.13 (A). An independent samples t-test indicated that there was not a significant difference in physical activity levels between younger adults (M = 8.67, SD = 2.67, *Min*. = 4.00, *Max.* = 16.00) and older adults (M = 7.60, SD = 2.28, *Min*. = 3.00, *Max.* = 12.00); $t(58) = 1.66$, $p = .10$, d =

0.43). A Bayesian independent samples t-test suggested that the data was 1.21 times more likely to occur under the null hypothesis (i.e., that there was no difference between age groups) than the alternative hypothesis (i.e., that there was a difference between age groups). The majority of participants were moderately physically active, with scores ranging from 6-10, where a score of 20 indicates the highest level of physical activity. The most physically active participant was a younger adult, who scored 16 on the physical activity questionnaire.

Figure 2.13 (B) shows the physical activity scores for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.13 (B) and indicates that there was a meaningful difference in physical activity scores between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE. The mean difference between the High BASTE group and the High BASTE group (i.e., 0) was less than the lower bound of the 95% confidence interval (i.e., 0.28) suggesting that older adults who performed better on the BASTE speech in noise task reported higher levels of physical activity than older adults who performed worse on the BASTE speech in noise task. This suggests that physical activity may be able to offset difficulties understanding speech in noise. Older adults with higher fitness levels likely have better vascular health in the cochlea and brain, which could help to facilitate the sensory and cognitive processes involved in speech perception in noise (Anderson, White-Schwoch, & Parbery-Clark, 2013; Yamasoba et al., 2013).

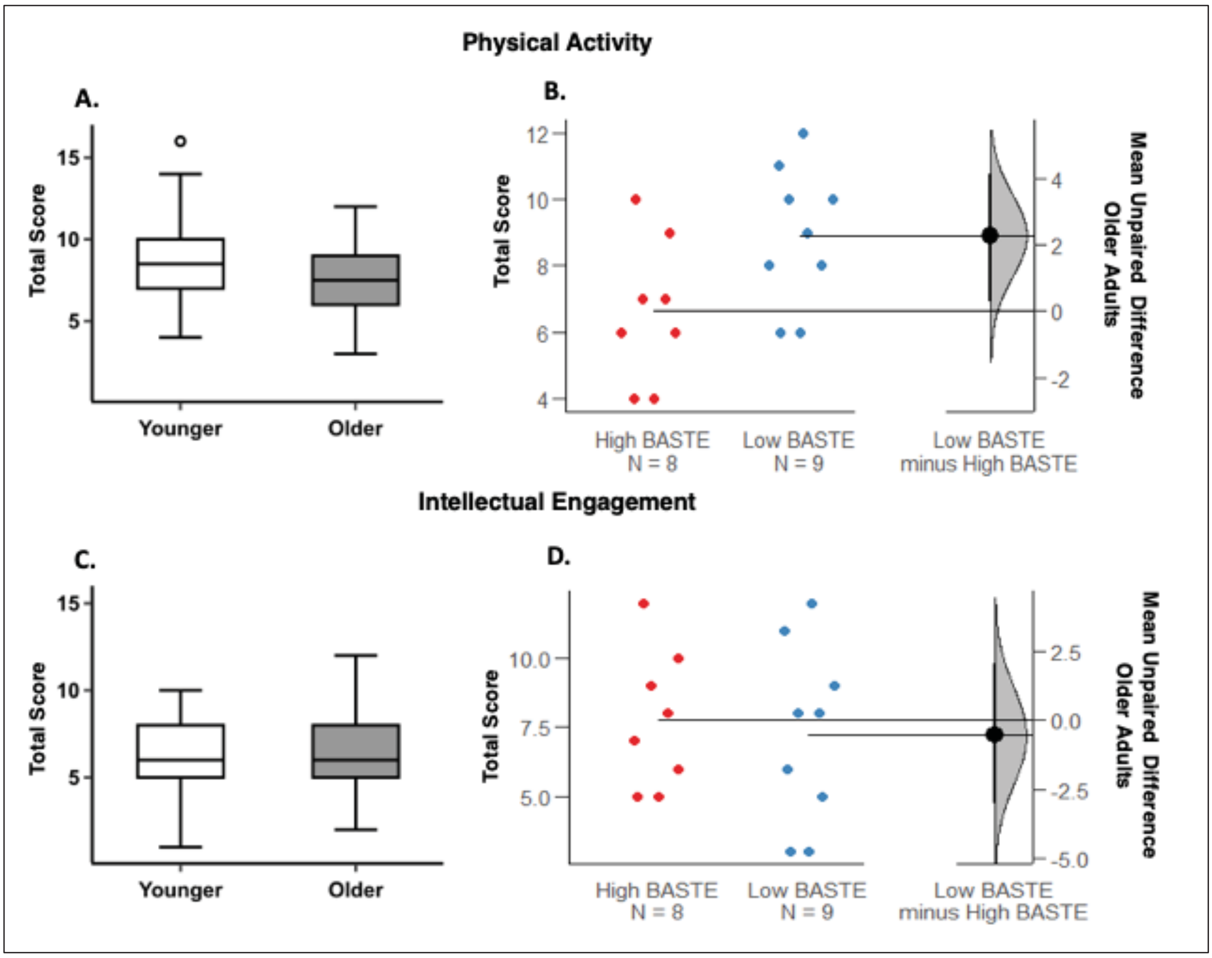

Figure 2.13. Physical Activity and Intellectual Engagement Results. Panels A and C shows younger and older adults' responses to the physical activity questionnaire (A) and the intellectual engagement questionnaire (C). Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panels B and D show Gardner-Altman estimation plots of the index of physical activity (B) and the index of intellectual engagement (D) for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE).

2.3.7.2 Intellectual Engagement

Total scores for the intellectual engagement questionnaire are depicted in Figure 2.13 (C). An independent samples t-test indicated that there was not a significant difference in intellectual engagement levels between younger adults (*M* = 5.73, *SD* = 2.57, *Min*. = 1.00, *Max.* = 10.00) and older adults (*M* = 6.73, *SD* = 2.61, *Min*. = 2.00, *Max.* = 12.00); *t*(58) = -1.49, *p* = .14, *d* = -0.39). A Bayesian independent samples t-test suggested that the data was 1.50 times more likely to occur under the null hypothesis (i.e., that there was no difference between age groups) than the alternative hypothesis (i.e., that there was a difference between age groups).The majority of participants were moderately intellectually engaged, with scores ranging from 5-8, where a score of 16 indicates the highest level of intellectual engagement.

Figure 2.13 (D) shows the intellectual engagement scores for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.13 (D). This analysis indicates that there was no meaningful difference in intellectual engagement scores between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE (95 CI: -3.04; 2.04). Indeed, as Low BASTE and High BASTE older adults both ranged from not very engaged (< 5) to engaged (≥ 12), this particular measure of intellectual engagement does not seem to be sensitive to older adults' speech perception in noise abilities.

2.3.7.3 Social Economic Status (SES)

Figure 2.14 (A) shows younger and older adults total scores for the index of SES. An independent samples t-test indicated that there was not a significant difference in SES between younger adults (*M* = 5.70, *SD* = 1.39, *Min*. = 2.00, *Max.* = 8.00) and older adults (*M* = 5.17, *SD* = 1.37, *Min*. = 3.00, *Max.* = 8.00); *t*(58) = 1.50, *p* = .14, *d* = 0.39. A Bayesian independent samples t-test suggested that the data was 1.50 times more likely to occur under the null hypothesis (i.e., that there was no difference between age groups) than the alternative hypothesis (i.e., that there was a difference between age groups). The majority of participants had a moderate (4-6) or high (6-8) SES.

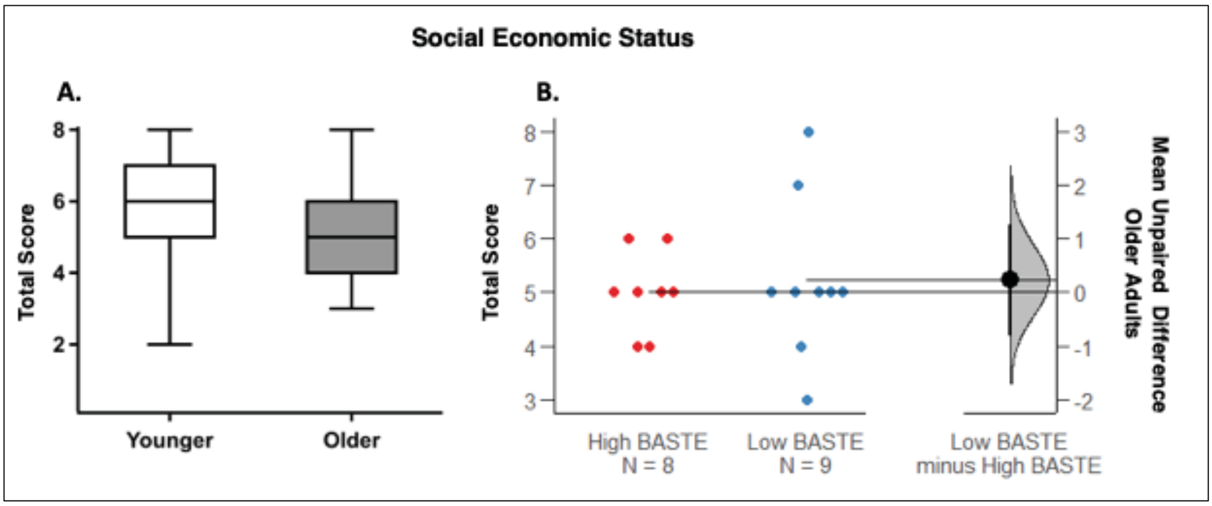

Figure 2.14. Social Economic Status (SES) Results. Panel A illustrates younger and older adults' responses to the SES questionnaire. Tukey's box plots represent the median and interquartile range (Q3-Q1). Circles represent data points that are at least one and a half times greater than (or less than) the interquartile range. Panel B shows a Gardner-Altman estimation plot of the SES measure for older adults who had BASTE scores at or above the 75th percentile (i.e., High BASTE) and older adults who had BASTE scores at or below the 25th percentile (i.e., Low BASTE).

Figure 2.14 (B) shows the SES scores for older adults who performed the best on the BASTE (i.e., low BASTE scores) and older adults who performed the worst on the BASTE (i.e., high BASTE scores). The data analysis with bootstrapped coupled estimation is also shown in Figure 2.14 (B). This analysis indicates that there was no meaningful difference in SES scores between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE, as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence interval of the mean difference between the Low BASTE group and the High BASTE group (95 CI: -0.81;1.26). As can be seen in Figure 2.14 (B), SES scores were more diverse overall for the Low BASTE group, however, both groups had multiple participants with a moderate SES score of five.

2.3.7.4 Musical Experience

As outlined in Table 2.3, approximately 30% of older participants reported playing a musical instrument and approximately 60% of younger participants reported playing a musical instrument. Very few participants reported that they had more than six years of formal training (i.e., four younger and four older) and that they still play music regularly (i.e., three younger and four older). The four older adults with the most musical experience in the current study all had BASTE scores greater than or equal to zero. Thus, none of these participants were included in the Low BASTE group; one was even included in the High BASTE group. These results (although based off a small sample size) are inconsistent with research suggesting that musicians have superior speech perception in noise abilities than non-musicians (Anderson, White-Schwoch, & Parbery-Clark, 2013; Parbery-Clark, Skoe, Lam, & Kraus, 2009) and aligned with studies showing that speech perception in noise ability is similar for samples with different levels of musical experience (Madsen, Marschall, Dau, & Oxenham, 2019).

Table 2.4
Musical Experience Questionnaire Results

Question	Response		
		Younger (n=30)	Older (n=30)
Do you play any instruments?	Yes	19	8
	No	11	22
If yes, how many instruments do you play?	1	8	4
	2+	4	4
Number of years of musical training?	0/n.a.	19	23
	0.5-5 yrs	7	3
	6-10 yrs	3	3
	10+ yrs	1	1
Do you still play music regularly?	Yes	3	4
	No	4	10
	Sometimes	2	0
	N/A	18	15
	Other	3	0

2.3.7.5. Hearing and Listening Experience (SSQ12)

The SSQ was administered to measure younger and older adults' perception of communication problems that they experience day-to-day. For display purposes, each SSQ item is listed in Table 2.5. Participants' ratings were regrouped into two negative categories (ratings < 3 = "Severe", ratings 4-6 = "Moderate") and two positive categories (ratings 7-9 =

“Mild”, and a rating of 10 = “None”). The percentage of responses for each category are displayed in Figure 2.15 (younger adults) and Figure 2.16 (older adults). As can been seen in the figures, there were two categories where older adults rated their listening much worse than the younger adults (Item 2, 36.5% negative responses compared to 30% for younger adults; and Item 4, 30% negative response for older adults compared to 10% for younger adults.

Table 2.5

SSQ 12 Items

Item	Question
1	You are talking with another person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?
2	You are listening to someone talking to you, while at the same time trying to follow the news on the TV. Can you follow what both people are saying?
3	You are in a conversation with one person in a room where there are many others talking. Can you follow what the person you are talking to is saying?
4	You are in a group of about five people in a busy restaurant, you can see everyone else in the group. Can you follow the conversation?
5	You are in a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?
6	You are outside. A dog barks loudly. Can you tell immediately where it is, without having to look?
7	Can you tell how far away a bus or truck is, from the sound?
8	Can you tell from the sound whether a bus or truck is coming towards you or going away?
9	When you hear more than one sound at a time, do you have the impression that it seems like a single jumbled sound?[1]
10	When you listen to music, can you make what which instruments are playing?
11	Do everyday sounds that you can hear easily seem clear to you (not blurred)?
12	Do you have to concentrate very much when listening to someone or something? [2]

Note. [1] = Likert response options range from 1: Jumbled – 10: Not Jumbled, [2] = Likert response options range from 1: Concentrate hard to 10: No need to concentrate. For all other items, response options range from 1: Not at all – 10: Perfectly

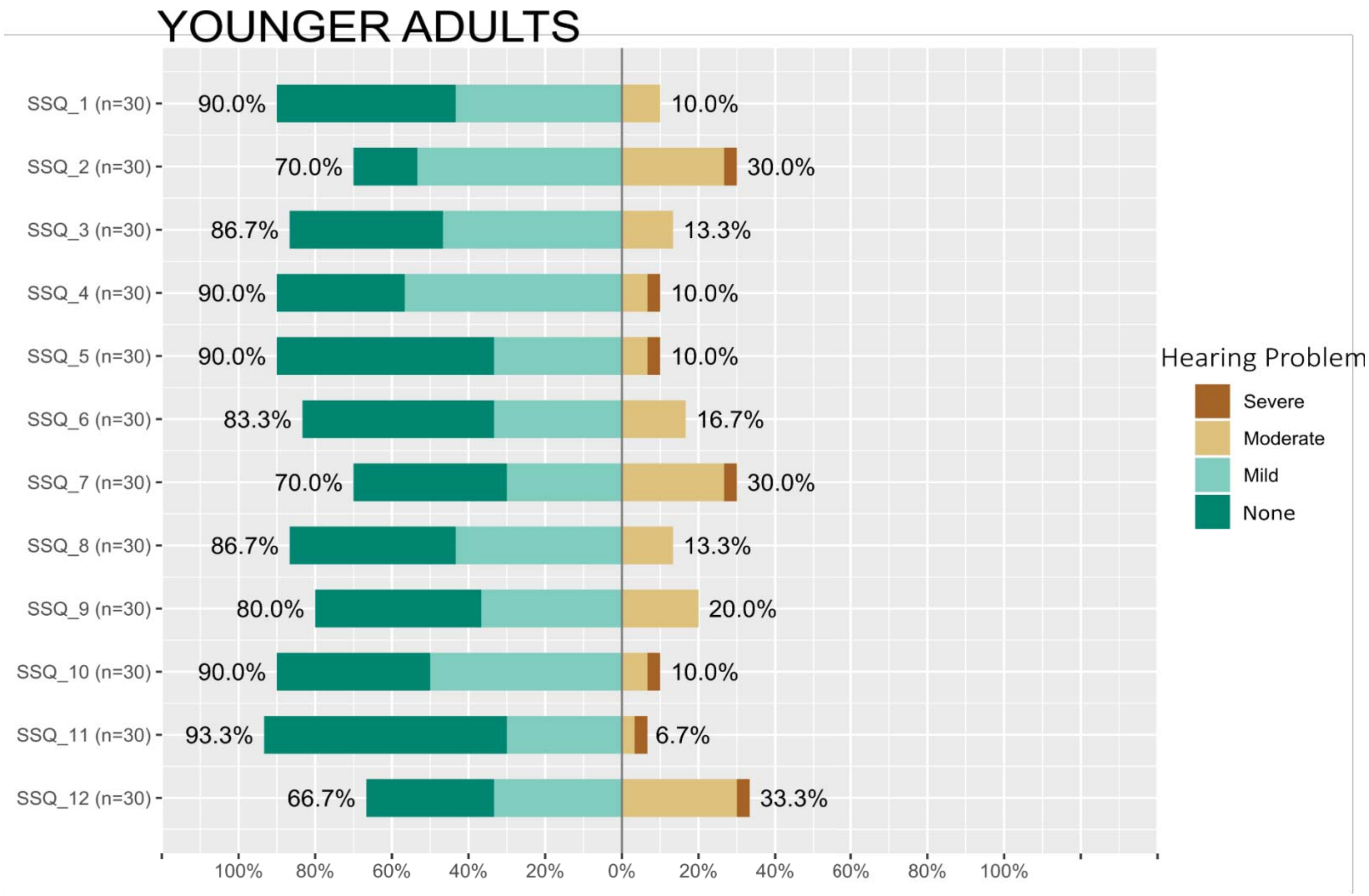

Figure 2.15. Younger Adults' Speech, Spatial and Qualities of Hearing Scale Likert Ratings.

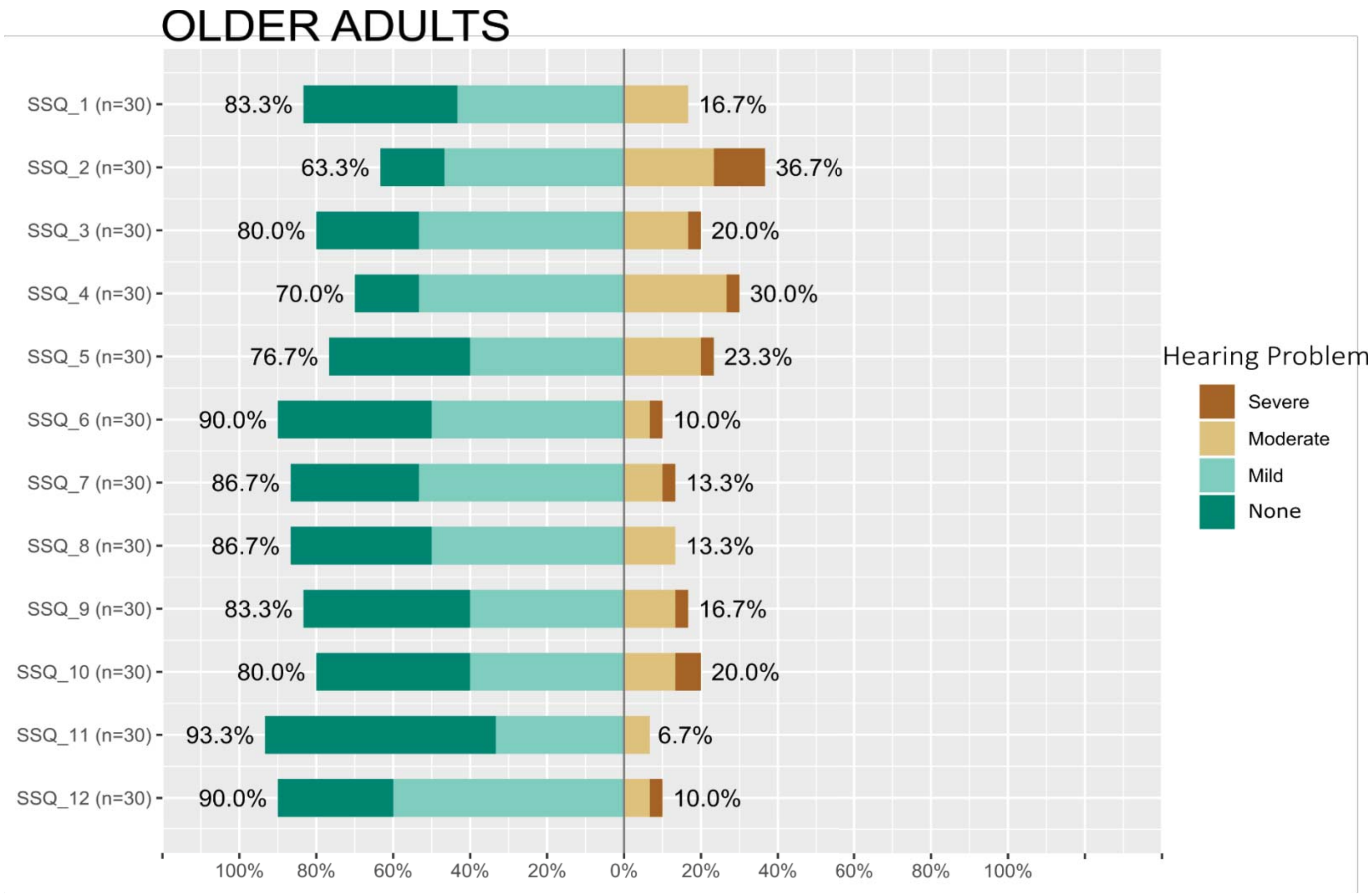

Figure 2.16. Older Adults' Speech, Spatial and Qualities of Hearing Scale Likert Ratings.

To test whether there was a difference in the ratings of these items as a function of age, bootstrapped coupled estimation was conducted with younger adults' ratings as the control group and older adults' ratings as the test group (See Figure 2.17). This analysis indicated that there was a meaningful difference in ratings between age groups for Item 4, but not for Item 2. That is, for Item 4, the mean difference between the younger adult group and the younger adult group (i.e., 0) was greater than the upper bound of the 95% confidence interval of the mean difference between the younger group and the older group (i.e., -0.13). This indicates that at least some older adults are aware of having more difficulty listening to and following conversations in a noisy restaurant than younger adults.

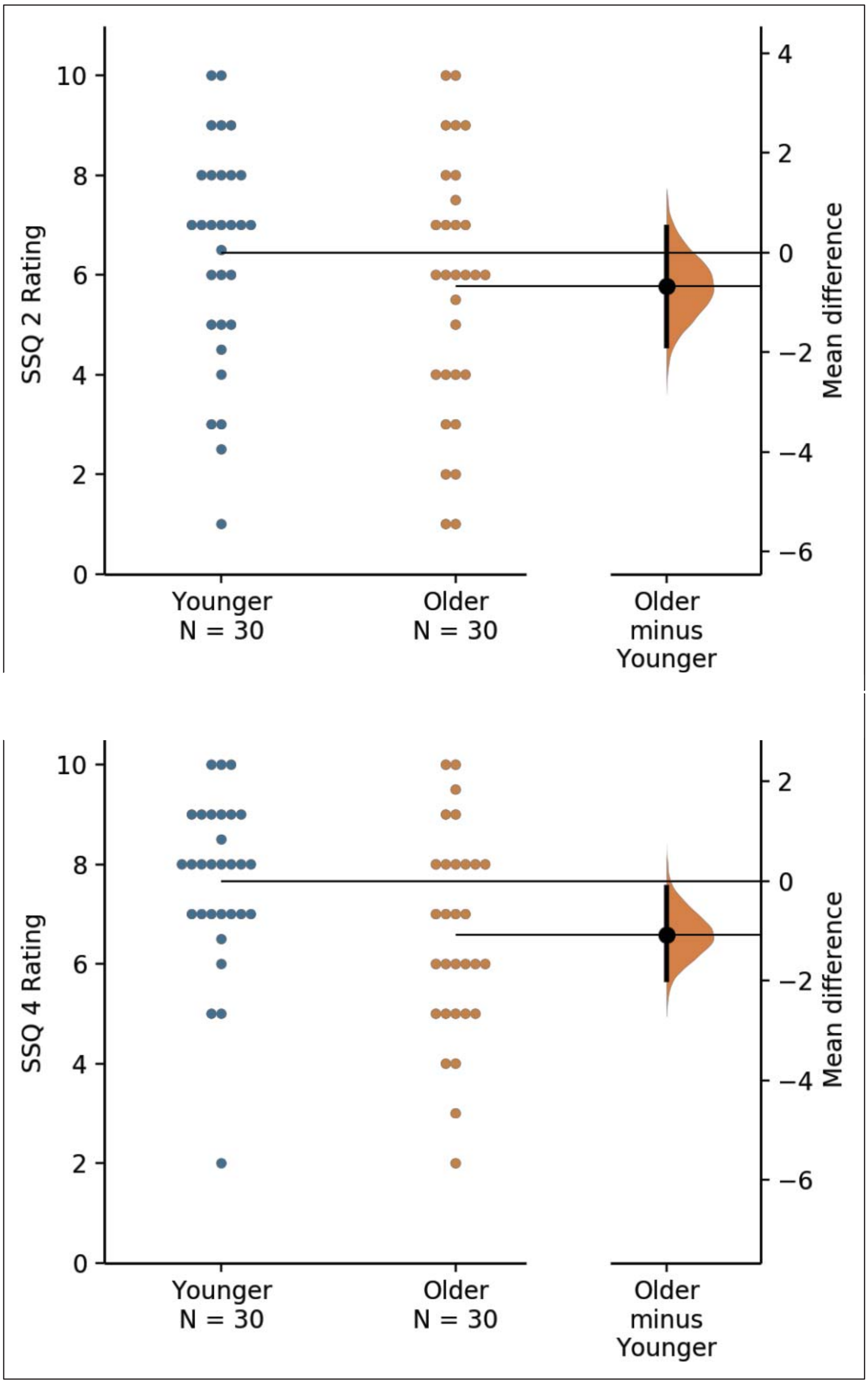

Figure 2.17. Gardner-Altman Plot of Younger and Older Adults Ratings for Speech, Spatial and Qualities of Hearing Scale Items 2 & 4.

To determine whether responses to these two SSQ items were related to older adults' performance on the BASTE speech perception in noise task a correlation over the data of the 30 older adults was performed along with a targeted analysis of the High and Low BASTE performers. The correlational analysis indicated that there were no secure correlations with BASTE performance for Item 2 (r = .09, p = .62) or Item 4 (r = -.17, p = .37). The data analysis with bootstrapped coupled estimation shown in Figure 2.18 was consistent with the correlational analysis. That is, the estimation suggested that there was no meaningful difference in SSQ ratings between older adults who performed the best on the BASTE and older adults who performed the worst on the BASTE for Item 2 or Item 4 as the mean difference between the High BASTE group and the High BASTE group (i.e., 0) did not exceed the 95% confidence intervals of the mean difference between the Low BASTE group and the High BASTE group for either item.

In summary, ratings for SSQ items that were the most sensitive to age did not explain the variance in BASTE performance for the older adult sample. This is consistent with studies showing that self-report measures of listening do not always strongly correlate with objective listening tests (Heinrich, Henshaw, & Ferguson, 2016). One reason for the inconsistency between older adults' performance on objective and subjective listening measures is that visual speech could help some older adults overcome poor listening performance during real-life listening, but not during the auditory-only speech recognition in noise tasks typically used to objectively measure listening ability.

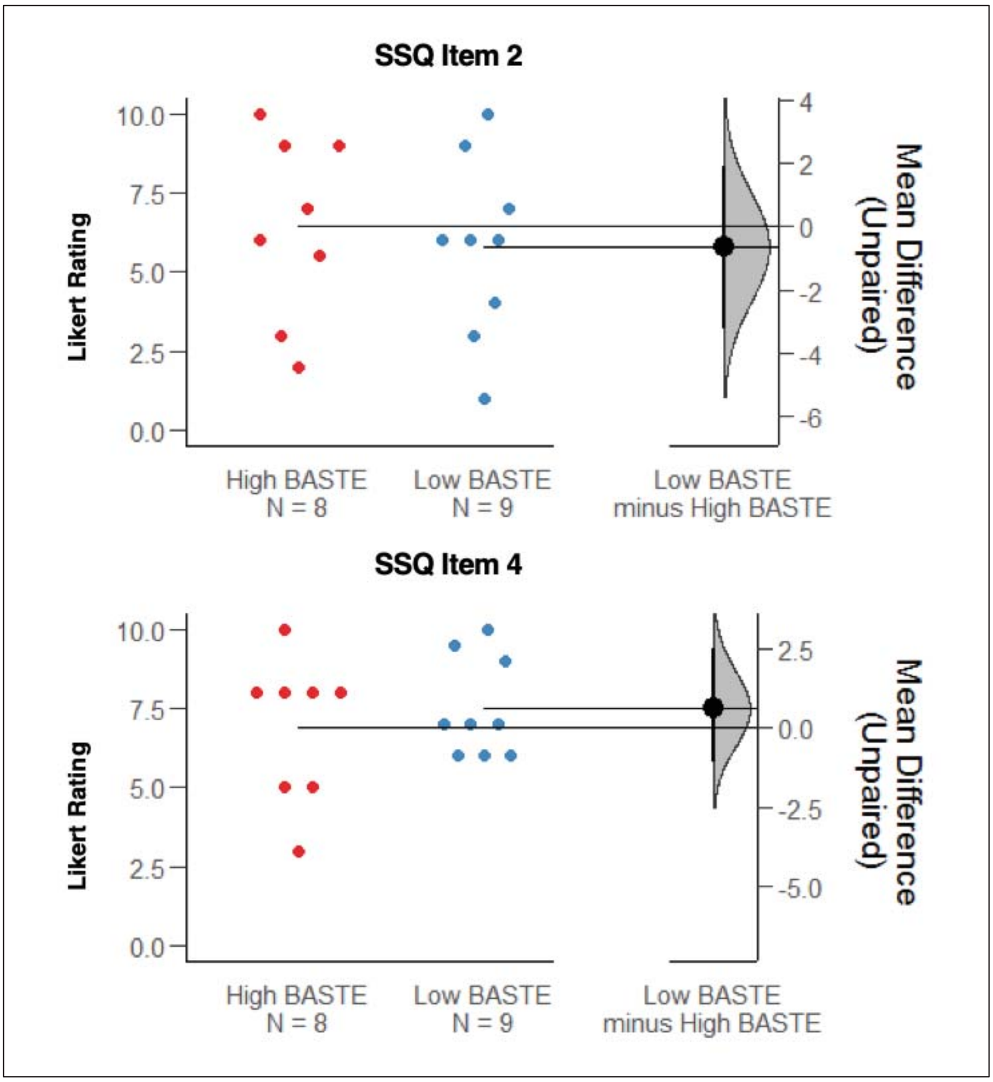

Figure 2.18. Gardner-Altman estimation plots for Older Adults SSQ Ratings

2.4 Summary and Future Directions

Consistent with previous research, the current study found that older adults had higher SRTs on a standard speech perception in noise task than younger adults. Older adults who participated in the current study also performed more poorly than younger adults on several auditory and cognitive tasks that have been related to speech perception in noise ability. Furthermore, the results from the correlational analysis and the DABEST analysis suggest that auditory attention is important for older adults' speech recognition in noise performance. These results are consistent with the FUEL and suggest that the role of attention should be carefully considered in future investigations of ageing and speech perception in noise.

Contrary to the predictions outlined in the ELU model, working memory did not seem to explain the variance in performance on the BASTE task for the older adults included in this study. That is, the results from the LSPAN and digit span tasks were either precise and the High and Low BASTE older adults performed within a similar precise range or the measure was imprecise and the variance in performance between High and Low BASTE groups was similar (e.g., the LSPAN). It is possible that the sample sizes used for the DABEST analysis may have contributed to the lack of meaningful difference between High and Low BASTE older adults on the cognitive assessments. It is also possible that as the memory tasks used in the current study were rather general, these tasks may not be sensitive to specific cognitive skills that are involved in speech perception in noise.

An interesting finding from the DABEST analysis was that older adults who had lower SRTs had higher levels of physical activity than older adults who had higher SRTs. Physical activity likely supports the integrity of different structures within the auditory-cognitive system and thus in turn supports speech perception in noise. Future research should examine the relationship between physical activity and speech perception in noise in more depth, using both subjective (e.g., questionnaire) and objective (e.g., stress test) measures.

As the results from the current study indicated that older adults' difficulties perceiving speech in noise may be modulated by attention, the experiments in Chapters 3 and 4 of this thesis take a more direct approach to investigating the role of attention in speech perception in noise by manipulating the auditory and visual processing demands of the speech perception task itself. These studies are also more ecologically valid than the current study as auditory-visual speech stimuli (rather than auditory-only speech stimuli) are used to investigate the effects of cognitive ageing on speech perception in noise.

Chapter 3

Effects of age and multiple talking faces on the visual speech benefit in noise

3.1 Introduction

Understanding speech in noise is challenging, especially for older adults (Pichora-Fuller, Alain, & Schneider, 2017). Seeing a talker's face facilitates speech perception in noise as the temporal and segmental speech information provided by mouth and lip movements compliments the auditory speech signal (i.e., the visual speech benefit in noise; Aubanel, Davis, & Kim, 2015; Kim & Davis, 2014; Munhall, Jones, Callan et al., 2004). The visual speech benefit has been well replicated and typically yields an 11dB improvement in speech recognition when compared to an auditory only baseline (Gordon & Allen, 2009, MacLeod & Summerfield, 1987; Sumby & Pollack, 1954). Notably, even though older adults understand less speech in noise than younger adults overall, the magnitude of the visual speech benefit in noise is typically the same for both age groups (Sommers, Tye-Murray, & Spehar, 2005; Tye-Murray, Spehar, Myerson, Hale & Sommers, 2016).

One limitation of virtually all investigations of aging and the visual speech benefit, however, is that only one video of a person producing visual speech that matches the auditory signal is presented during auditory-visual trials. These trials place little demand on visual-spatial selective attention, as there is only one face to look at (i.e., foveate) and/or focus attention on. As research using paradigms other than the visual speech benefit have found that the attentional demands of a task can modulate auditory-visual processing effects, it is possible that attention also plays a role in gaining a visual speech benefit. If this is the case, then age-related declines in attentional capacity may affect older adults' ability to gain a visual speech benefit when auditory and/or visual processing are attentionally demanding.

Although older adults have not been specifically tested, experiments using paradigms other than the visual speech benefit have shown that auditory-visual processing effects are limited by the availability of attentional resources. For example, the McGurk Effect, a classic example of a visual influence on auditory speech perception in which the auditory speech token /ba/ is perceived as /da/ when watching a face saying /ga/ (McGurk & MacDonald, 1976) is reduced under attentionally demanding conditions (Alsius, Navarra, Campbell & Soto-Faraco, 2005; Alsius, Navarra & Soto-Faraco, 2007). Given that the McGurk effect is reduced when attentional resources are depleted, it seems plausible that other auditory-visual processing effects involving speech (i.e., the visual speech benefit) will also be limited by the availability of attentional resources.

The way that Alsius and colleagues manipulated attention was by requiring participants to engage in a secondary task (i.e., to detect a repeated auditory or visual stimulus or a specific tactile pattern). However, not only can a secondary task interact with the primary one in difficult to interpret ways, but also the precise cognitive mechanisms involved can be unclear (Damos, 1991). For the current experiments, we took inspiration from visual search paradigms where the extent of the involvement of visual-spatial selective attention in locating a target can be deduced by what happens to search time when the number of distractors is varied.

That is, for visual search tasks where participants are charged with locating a particular type of visual target amongst a varying number of distractors, it is generally agreed that when search time increases as a linear function of the search set size (i.e., the number of distractors), each visual stimulus needs to be processed individually, and that top-down, visual-spatial selective attention is involved in this serial search for the target (Wolfe, 2010). Alternatively, finding that the time taken to locate the target does not vary as the number of distractors is increased indicates that a property of the target stimulus can guide the search process directly to it so that a serial search involving top-down, visual-spatial selective attention is not necessary, or that the participant is able to process the distractors and target together at the same time (Wolfe, 2010).

With the visual search paradigm in mind, we adapted the standard speech in noise paradigm, where participants are required to identify speech in noise for Auditory Only (AO) and Auditory-Visual (AV) presentations, and manipulated the set size by adding one, three or five talking faces silently uttering irrelevant sentences (i.e., distractors). The same person was used in both target and distractor videos to control for any individual differences between talkers that might influence the way visual-spatial selective attention is deployed. If participants need to engage in a serial visual search in order to process the relevant talking face, then there should be a linear decrease in the size of the visual speech benefit proportional to the number of distractor videos presented. When two talking faces are presented, for example, participants have a 50% chance of processing the irrelevant video first. Presuming they are not able to shift in time to the alternate (matching) talking face video, they would only gain half of the benefit that was gained during a trial with only one relevant talking face. Under this simple model, the participant would have an even smaller chance (i.e., 25%) of initially processing the matching talking face when four talking faces are presented. This pattern of results would be consistent with the proposal that visual-spatial selective attention is involved in gaining a visual speech benefit.

It should be emphasised that participants in our study were not explicitly asked to locate the talker that matched the auditory signal (although this is likely a strategy adopted by participants). Rather, participants were only instructed to report the speech that they heard on each trial. As the current study is a speech recognition in noise task, it is important to note that in addition to the number of visual distractors present, degrading the quality of the auditory signal also can influence how visual-spatial selective attention may be deployed. To illustrate this effect, consider the study by Stacey, Murphy, Sumner, Kitterick, and Roberts (2014). In Stacey et al.'s (2014) study, participants were shown videos of two, three or four people each uttering a different sentence and were asked to locate the face that matched an auditory target. The attentional demands on auditory processing were manipulated by presenting target sentences as natural or vocoded speech, presented in multi-talker babble or in quiet. The results showed that when the auditory target was presented clearly, there was no significant increase in search time as the set size increased (i.e., response times were approximately 2000 ms for each condition). However, when the auditory target was degraded, response times increased by approximately 200 ms as the number of faces increased.

Stacey et al. (2014) interpreted their results as showing that, when the speech was natural, the perceiver had sufficient attentional resources to combine the auditory and visual speech information from several faces at once (regardless of where they were foveating) and thus they were able to identify the talker producing visual speech that matched the auditory target independent of the number of faces in the stimulus set. However, when attentional resources had to be expended just to process the auditory speech information (i.e., when the speech was vocoded and/or mixed with noise), then participants could only process the visual speech information from one talking face at a time, which, as Stacey et al. (2014) suggest, is indicative that visual-spatial selective attention was involved. As the auditory targets in our study were also masked, it seems likely that talking face videos will need to be processed serially, which would be indicated by a decrease in the magnitude of the visual speech benefit as the set size increases.

In summary, existing research on auditory-visual processing and attention suggests that gaining a visual speech benefit is likely limited by the availability of attentional resources. If this is the case, then older adults' ability to gain a visual speech benefit could be impaired relative to younger adults, especially when the attentional demands of auditory and visual processing are high, i.e., when there are many visual distractors, and when the auditory task is difficult. That is, research suggests that older adults have a smaller maximum attentional resource capacity than younger adults (Bialystok & Craik, 2006; Craik & Byrd, 1982; Craik & Salthouse, 2011; Heinrich, Gagné, Viljanen, Levy, Ben-David, & Schneider, 2016). Furthermore, when hearing acuity is reduced, as is the case for many older adults, the perceptual effort needed for speech recognition could deplete attentional resources (Pichora-Fuller, Kramer, Eckert, et al., 2016; Wingfield, Amichetti, & Lash, 2015). Finally, even when attentional resources are available, older adults appear to be poorer at controlling them. For example, older adults are less proficient than younger adults at shifting focused attention from one object or location to another (i.e., orienting) and at selectively focusing on a stimulus while inhibiting distractors (Hasher & Zacks, 1988; Greenwood & Parasurman, 2004; Lustig, Hasher, & Zacks, 2007; Madden, Connelly & Pierce, 1994; Zanto & Gazzaley, 2014). Thus, if the number of talking faces that can be processed at once is limited by the availability of attentional resources, then older adults should have particular difficulty gaining a visual speech benefit when the attentional demands of auditory and visual processing are high.

For the present study, younger and older adults were presented with spoken sentences mixed with speech shaped noise in five visual display conditions: Static (a static image of a face or faces), One Talking Face (one visual speech video relevant to the auditory signal), and three conditions with multiple (two, four or six) visual speech videos. The conditions with multiple visual speech videos always included one video that matched the auditory signal (i.e., the relevant video); the other video(s) showed irrelevant visual speech. To control for any individual differences between talkers that might influence the way that attention may be deployed in any visual search, all speech videos (relevant and irrelevant) consisted of the same female talker.

As it was not possible to know the exact amount of attentional resources that would be required for auditory processing at different SNRs, two experiments (with different SNRs and thus different auditory processing demands) were run. For Experiment 1, a SNR that would be difficult for both younger and older adults was selected (i.e., -6dB) in order to prevent ceiling performance on the Static Condition. In case the auditory processing demands of -6dB SNR completely deplete the attentional resource capacities of both age groups (thus minimising any differences between groups), Experiment 2 used a less demanding SNR (i.e., -1dB).

For both experiments, it was predicted that both age groups would get a standard visual speech benefit (i.e., speech recognition would be better during the One Talking Face Condition than the Static Condition), but that older adults would perform worse overall than younger adults on the speech recognition task due to age-related declines in hearing sensitivity (Tun, Williams, Small, & Hafter, 2012). For Experiment 1, (i.e., when auditory processing demands were high; -6dB SNR), it was expected that if participants need to engage in a serial visual search in order to process the relevant talking face, then the magnitude of the visual speech benefit should reduce for both age groups as more irrelevant talking faces are presented. However, older adults should have greater difficulty in comparison to younger adults when irrelevant talking faces are presented, due to an age-related decline in the maximum capacity of top-down attentional resources (Craik & Byrd, 1982; Craik & Salthouse, 2011).

For Experiment 2 (i.e., -1dB SNR) it was expected that due to age-related declines in hearing sensitivity, the majority of older adults' attentional resources would be devoted to auditory processing, and thus, they would only be able to process one talking face a time (i.e., their speech recognition would reduce as more talking faces were presented). However, it was expected when the SNR was -1dB, younger adults with good hearing would have sufficient attentional resources to process more than one talking face at a time and would therefore be able gain a full visual speech advantage when more than one talking face was presented.

To control for any potential effects of peripheral vision, the visual angle (Swearer, 2011) of the Six Talking Faces Condition was approximately 20°. That is, for a trial from the Six Talking Face Condition, if a participant was foveating on the last face on the right-hand side of the monitor (i.e., face number six), and the relevant talking face was the first face presented on the left-hand side (i.e., face number one; the face furthest away from face number six), then there would only be 20° visual angle between fixation and the relevant talking face. As research suggests that the McGurk Effect only significantly reduces when an individual's gaze is displaced beyond 20° (Paré, Richler, ten Hove, & Munhall, 2003), and an auditory-visual benefit for syllable detection in noise can be accrued from visual speech presented at

23.60° eccentricity (Kim & Davis, 2013), any reduction in the visual speech benefit when multiple talking faces are presented for the current study should be due to attention (which is not necessarily locked to eye movement; Posner, 1980) rather than limitations of peripheral vision.

3.2 Experiment 1

3.2.1 Method

3.2.1.1 Participants

Twenty-four younger adults (19 Females, M_{Age} = 22) and 24 older adults (14 Females, M_{Age} = 69) participated in this study. Younger adults were undergraduates at Western Sydney University and participated for course credit. Older adults were recruited from the Australian Seniors Computer Club Association (ASCCA) and were given a monetary reimbursement. All participants reported English as their first language and passed a screening test for mild cognitive impairment (The Clock Test; Nishiwaki et al., 2004). None of the participants were hearing aid users.

3.2.1.2 Stimuli

The stimuli consisted of 140 auditory and visual recordings (120 test trials, 12 catch trials, and eight practice trials) of a native Australian-English female talker uttering Harvard IEEE sentences. The recordings were selected from the MAVA database (i.e., MARCS Auditory –Visual Australian recordings of IEEE sentences, Aubanel, Davis, Kim, 2017). Each individual video was cropped to show only the lower portion of the face and measured 4.5cm (height) x 8cm (width) with a total visual angle of 28°**.**

3.2.1.2.1 Summary of Experimental Conditions

Video recordings were manipulated (using FFmpeg) to produce five experimental conditions which are summarised in Table 3.1.

Table 3.1

Visual Stimuli Presented for Each Experimental Condition

Condition	Visual Stimuli
One Talking Face	One silent video showing a person uttering a sentence that matches the auditory signal.
Two Talking Faces	Two silent videos side-by-side. Each video shows the same person uttering a different sentence; visual speech from one video matches the auditory signal. See Figure 3.1.
Four Talking Faces	Four silent videos side-by-side. Each video shows the same person uttering a different sentence; visual speech from one video matches the auditory signal.
Six Talking Faces	Six silent videos side-by-side. Each video shows the same person uttering a different sentence; visual speech from one video matches the auditory signal.
Static	A static black and white image of a single face or multiple faces.

3.2.1.2.2 Test Trials

Five versions of the experiment were created so that each item could appear in all conditions without being repeated to a participant. For each version of the experiment, 24 IEEE sentences were assigned to each condition.

Each version of the One Talking Face Condition consisted of 24 videos of one female uttering a single sentence. Each video recording was combined with the auditory signal (which had been mixed with speech shaped noise derived from the long-term average spectrum of the 140 sentences used) that matched the visual utterance.

Each version of the Two Talking Faces condition consisted of 12 sets of video pairs (i.e., a single video file with two silent visual speech videos, side-by-side, each simultaneously uttering a different IEEE sentence). Each video pair was presented twice: once with the auditory signal (with noise) matching the visual speech video on the left, and again so that the auditory signal (with noise) matched the visual speech video on the right, producing a total of 24 trials. See Figure 3.1.

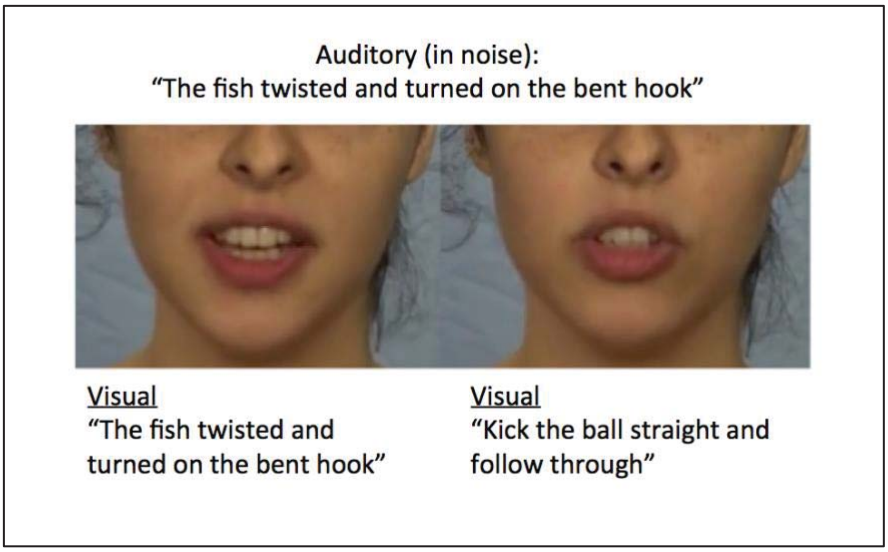

Figure 3.1. Example Trial from the Two Talking Faces Condition.

The same procedure was applied to stimuli production for the Four and Six Talking Face conditions. For example, for the Four Talking Faces Condition, six sets of side-by-side videos with four talking faces were used to make 24 auditory-visual stimuli, with auditory recordings added to match each of the visual speech videos. This procedure ensured that the spatial location of the matching talking face appeared at each possible location an equal amount of times throughout the experiment.

For the Static Condition, one still frame of each type of visual speech video (i.e., one, two, four, or six faces) was taken and edited to be black and white with Adobe Illustrator CS6. Each picture type was presented six times (six presentations x four picture types = 24 trials). Images were presented simultaneously with an IEEE sentence mixed with noise at -6dB.

3.2.1.2.3 Catch Trials

Catch trial sentences were different from sentences used for practice trials and test trials. Twelve catch trials were included in all versions of the experiment to ensure that participants attended to the visual stimuli. Catch trials were identical in appearance to test trials, except an image with a red cross placed over the mouth region(s) was displayed for 200ms at the end of each video.

3.2.1.3 Apparatus

Stimuli were presented using DMDX software (Forster & Forster, 2003) on a Dell T7810 computer with Windows 7 software. Visual stimuli were presented on a monitor measuring 30cm (height) x by 53cm (width). Auditory stimuli were presented through Sennheiser HD280pro headphones.

3.2.1.4 Procedure

Participants were tested individually. They first completed a questionnaire that asked about their age, sex, and native language. Next, participants were seated approximately 70 cm from a computer monitor in a sound attenuating booth. Participants were told that they would see and hear a person uttering a sentence. They were instructed to attend to the videos (or picture) that appeared on the screen, listen carefully to the speech presented in noise, and type out what they heard once the word "respond" appeared on the screen.

Participants then completed one version (out of five possible versions) of the experiment. All versions began with the same two-phase practice session. Phase one consisted of six trials (two trials from the single talking face condition, and one trial from each of the other conditions). Phase two of the practice session consisted of two practice catch trials. Participants were instructed to look for a red cross that appeared over the persons mouth at the very end of each video, and to type "999" whenever they saw a red cross. The researcher told each participant that red crosses would appear randomly throughout the experiment. Sentences used for the practice session were not included in the test trials. Participants were invited to repeat both phases of the practice session as many times as they felt necessary. Two of the older participants requested to repeat the practice session. Participants were invited to adjust the volume to a comfortable level.

After the practice session, participants completed 132 test trials (i.e., 24 trials from each condition and 12 catch trials) presented in a pseudo-random order. Participants were encouraged to take a short break after completing 72 trials. The total listening time for each participant was approximately 40 minutes.

At the conclusion of the experiment, participants completed a visual acuity test, (the Freiburg Visual Acuity and Contrast Test [FrACT]; Bach, 2007) and a screening measure for mild cognitive impairment (i.e., The Clock Test). Hearing sensitivity was assessed by measuring pure-tone thresholds at five different frequencies (0.25,0.5, 1, 2, and 4 kHz) with a Diagnostic Audiometer (AD229e).

3.2.2 Results

Mean correct keywords recognised for younger and older adults as a function of display condition is shown in Figure 3.2. As can be seen, speech recognition for both younger and older adults improved from the Static Condition to the One Talking Face Condition, and then reduced as more talking faces were presented.

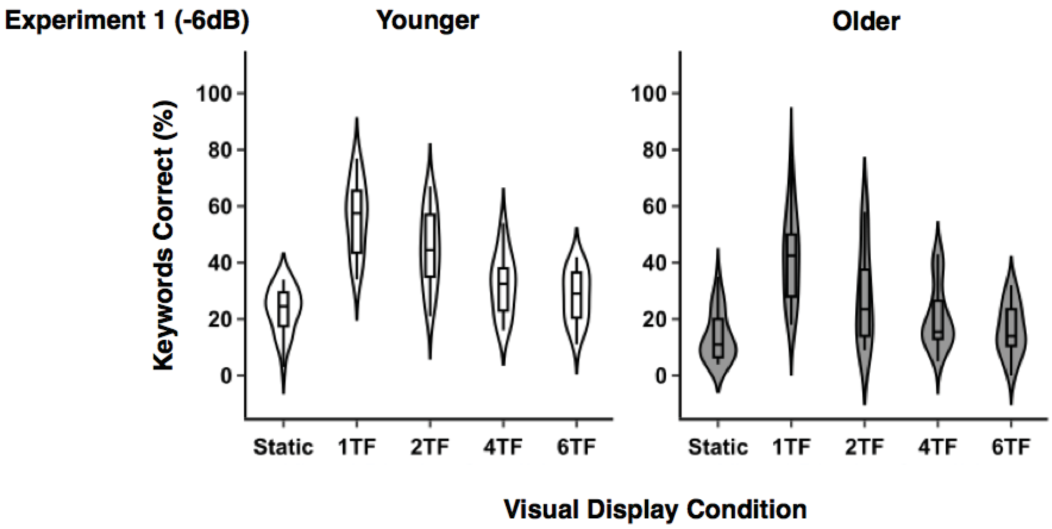

Figure 3.2. Results from Experiment 1. Tukey's box plots represent the median and interquartile range (Q3-Q1) of the percentage of keywords correctly reported as a function of visual display condition for younger and older adults. Violin plots represent the probability density of the data across the distribution.

A mixed factorial ANOVA was conducted with Age (Younger vs. Older) as the between participants factor and Face Condition (Static vs. One Talking Face vs. Two Talking Faces vs. Four Talking Faces vs. Six Talking Faces) as the within participants factor. A significant main effect of Age $F(1, 46) = 19.47$, p=.000, η^2 =. 30 was found. Younger adults (M = 36.87, SE = 2.02) correctly recognized more key words than Older Adults (M = 24.24, SE = 2.02). There was also a significant main effect of Condition, $F(3.20, 145.357) = 135.89$, $p < .001$, $\eta^2 = .75$ (Greenhouse-Geisser used for degrees of freedom). Bonferroni corrected post hoc tests revealed that participants (both younger and older adults) recognised significantly more words during the One Talking Face Condition (M = 49.02, SE = 2.05) than the Static Condition (M = 18.52, SE = 1.22). There was no significant difference between performance on the Static Condition and the Six Talking Face Condition (M = 22.10, SE = 1.30), however, all other conditions were significantly different from each other (2TF, M = 36.56, SE = 2.20 4TF, M = 26.56, SE = 1.59). There was no significant interaction between Age and Condition, $F(4,46) = 1.47$, $p = .215$, $\eta^2 = .03$.

As the results from existing research (including the current study) indicate that that there is approximately a 30% improvement in listeners' speech recognition in noise ability when listeners can see one talker's face in comparison to an auditory-only condition, I calculated how much of this standard percentage benefit participants gained when multiple talking faces were presented. To do this, I multiplied the mean percentage of keywords correctly identified for a condition with multiple faces (e.g., the Two Talking Faces Condition) by 100, and divided this value by the mean percentage of keywords correctly identified for the One Talking Face Condition. The resulting values represent the percentage of the standard visual speech benefit that was gained for each condition with multiple talking faces, which are illustrated in

Figure 3.3. As can be seen in the figure, the percentage of the standard visual speech advantage gained decreased as the number of talking faces increased. For the Two Talking Faces Condition, younger adults gained approximately 80% of the standard visual speech benefit, whereas older adults gained approximately 70% of the standard visual speech benefit (Younger: *M* = 79.37, *SE* = 4.35, Older: *M* = 66.83, *SE* = 4.35). In comparison to the Two Talking Faces Condition, the percentage of the standard visual speech benefit gained by younger and older adults reduced by approximately 20% when four talking faces were presented (Younger: *M* = 57.69, *SE* = 3.05, Older: *M* = 48.88, *SE* = 3.05) and by an additional (approximate) 10% when six talking faces were presented (Younger: *M* = 50.42, *SE* = 3.07, Older: *M* = 37.10, *SE* = 3.07).

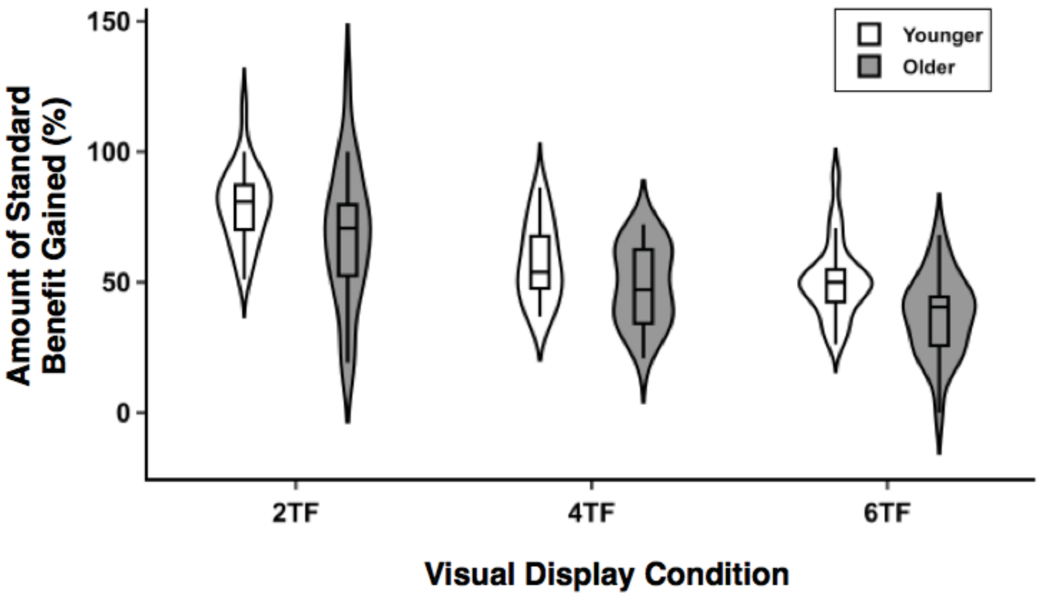

Figure 3.3. The Percentage of the Standard Visual Speech Benefit Gained as a Function of Display Condition for Experiment 1. Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution.

3.2.2.1 Visual acuity

3.2.2.1.1 Younger Adults

All younger participants had normal or corrected to normal vision (i.e., ≥ 1.0 on the FrACT visual acuity measure; Bach, 2007). Younger adults' visual acuity scores ranged from .99 to the maximum score of 2.0 (*M* = 1.41, *SD* = .30).

3.2.2.1.2 Older Adults

Eight older adults had worse than normal vision (i.e., < 1.0 on the FrACT visual acuity measure) with visual acuity scores ranging from 0.71 to the maximum score of 2.0 (*M*= 1.21, *SD*= .37). Pearson product-moment correlation coefficients were computed to test whether visual acuity was related to performance on the speech recognition task. The results indicated that older adults' visual acuity scores were not significantly related to performance on any of the conditions (all p values ≥ .50; r values: Static = -.12, 1TF = -.06, 2TF = -.20, 4TF = -.02, 6TF = -.04).

3.2.2.2 Hearing Sensitivity

Table 3.2 (Experiment 1) summarises the hearing sensitivity levels for both younger and older adults. All younger participants had normal hearing (i.e., ≤ 25dB HL at .25, .5, 1, 2, 4 kHz). Older adults' hearing levels were more diverse, ranging from normal to moderate-severe hearing loss (i.e., ≥ 40dB and ≤ 70dB HL at one frequency), with the majority of older participants (i.e., 14) having only mild hearing loss (i.e., ≥ 25dB and ≤ 40dB HL for all tested frequencies). Mean pure-tone hearing thresholds for each tested frequency are shown in Table 3.3 (Experiment 1). As can be seen, younger adults had significantly lower thresholds than older adults for both ears at all tested frequencies (all p values ≤ .03), except for .25 kHz (p = .25).

Better Ear Average scores were calculated by averaging hearing thresholds across all tested frequencies for each ear and selecting the lower average threshold.

The within group variation for the Better Ear Average was greater for older adults (*Min.* = 13.00, *Max.* = 34.00, *M* = 23.00, *SD* = 5.79) than younger adults (*Min.* = 8.00, *Max.* = 20.00, *M* = 14.00, *SD* = 3.38). The Better Ear Average was not significantly related to performance on any of the display conditions for either age group (all p values ≥ 0.16; younger adult r values: Static = -.30, 1TF = -.17, 2TF = -.06, 4TF=-.23, 6TF = -.07; older adult r values: Static = -.40, 1TF = -.25, 2TF = -.13, 4TF = -.26, 6TF = -.22).

Table 3.2

Hearing Levels for Younger and Older Adults

Hearing Level	Definition	Experiment 1		Experiment 2	
		Younger (n=24)	Older (n=24)	Younger (n=20)	Older (n=20)
Normal	≤ 25[1] at all frequencies[2]	24	7	20	4
Mild Loss	>25 - ≤ 40 at one frequency	0	13	0	9
Moderate Loss	> 40 - ≤ 55 at one frequency	0	2	0	4
Moderate-Severe Loss	> 55 - ≤ 70 at one frequency	0	2	0	3

Note. Hearing level definitions adapted from Wayne et al., 2016 and are measured from the better ear.

[1]dB Hearing Loss

[2]All frequencies refers to .25, .5, 1, 2, 4 kHz

Table 3.3

Mean Pure-Tone Hearing Thresholds for Experiments 1 and 2

		Experiment 1		Experiment 2	
Ear	Frequency (kHz)	Mean dB HL (SD)		Mean dB HL (SD)	
		Younger (n=24)	Older (n=24)	Younger (n=20)	Older (n=20)
Right	0.25	18.33 (4.34)	22.08 (4.40)	16.75 (2.94)	21.25 (5.82)
	0.50	17.50 (5.52)	23.54 (6.51)	14.00 (3.48)	23.50 (7.45)
	1.00	16.04 (4.66)	22.29 (5.89)	13.50 (3.66)	23.75 (9.58)
	2.00	12.71 (4.89)	23.33 (8.68)	13.25 (4.95)	28.00 (10.31)
	4.00	9.58 (4.87)	32.50 (14.45)	9.00 (6.20)	37.25 (15.09)
Left	0.25	18.54 (4.54)	20.42 (6.41)	16.50 (2.86)	21.75 (6.13)
	0.50	17.71 (4.42)	21.67 (7.32)	14.25 (4.38)	23.00 (9.65)
	1.00	14.38 (4.96)	22.08 (5.50)	12.25 (3.02)	21.50 (10.27)
	2.00	13.33 (6.54)	22.92 (7.65)	10.75 (5.91)	27.25 (13.13)
	4.00	11.46 (6.51)	34.17 (16.98)	9.00 (7.71)	44.25 (17.72)

3.2.3 Discussion

Experiment 1 tested whether the magnitude of the standard visual speech benefit, found when only a single talker's face is shown, is reduced when additional talking faces are presented. The results confirmed that the standard visual speech benefit was found for both younger and older adults, i.e., speech recognition in noise was better for the One Talking Face Condition compared to the auditory-only Static Condition. Consistent with previous research, the magnitude of the standard visual speech benefit (i.e., approx. 30%) was not significantly different for younger (M = 32.75) and older adults (M = 28.25; $t(46)$= 1.27, p = 0.212, BF_{01} = 1.82; Sommers, Tye-Murray, & Spehar, 2005; Tye-Murray, Spehar, Myerson, Hale & Sommers, 2016).

As expected, the size of the visual speech benefit became smaller as additional talking faces were presented. That is, each time the number of faces increased by two the visual speech benefit reduced by approximately 50%. This pattern suggests that younger and older adults conducted a serial visual search for the relevant talking face and is consistent with the proposal that visual speech information can only be combined with auditory information for a single face at a time. That is, the chance of attending to the matching face would reduce as the number of faces increased. Given that the visual speech information from even six talking faces would be available (Paré, Richler, ten Hove, & Munhall, 2003), this finding suggests that in order to gain a visual speech benefit a person must direct visual-spatial selective attention to relevant visual speech, and that the ability to do this decreases as the stimulus set-size increases.

As expected, older adults understood significantly less speech in noise than younger adults. This suggests that the task was more difficult for older adults and likely reflects age-related hearing problems, as indicated by older listener's higher pure-tone thresholds. Unexpectedly, the impact of additional irrelevant talking faces was similar for both age groups. That is, there was no differential effect of age on the visual speech benefit when multiple talking faces were presented. One explanation for the lack of difference between age groups is that the older adults who participated in this study had similar attentional resource levels as the younger adult group. As measures of cognitive skills like attention and working memory were not collected in Experiment 1, it is not possible to evaluate this proposal. This limitation is considered in Experiment 2.

Another explanation for why there was no differential effect of age when additional talking faces were presented is that the overall difficulty of the speech recognition task was too high and this depleted the attentional resource capacities of both age groups. That is, a SNR of -6dB was selected to prevent a ceiling effect for the younger listeners. However, this resulted in a task with very high auditory processing demands for both younger and older adults. Indeed, even younger adults only correctly reported 23% of the keywords for the Static Condition.

High auditory processing demands may have depleted the attentional resources for both younger and older adults, thus reducing the chance to observe differences due to attentional capacity between age groups (i.e., a floor effect). If this was the case, then any between age group effect of presenting additional talking faces should be clearer when a less adverse SNR (e.g., -1dB) is used. That is, if the SNR is -1dB younger adults would need to devote less attentional resources to processing the auditory speech than when the SNR was -6dB. Thus, it is possible that younger adults would have sufficient attentional resources to combine the auditory and visual speech information from several faces at once, gaining a visual benefit that is no different from when a single face is presented. As older adults may still need to devote a significant portion of their resources to auditory processing when the SNR is -1dB (due to age-related hearing loss), then older adults may only have sufficient attentional resources to process one talking face at a time. This was tested in Experiment 2.

3.3 Experiment 2

3.3.1 Method

3.3.1.1 Participants

Twenty younger adults (13 Females, $M_{\mathrm{Age}} = 21$) and 20 older adults (12 Females, $M_{\mathrm{Age}} = 72$) participated in the experiment. Ten of the older adults who participated in Experiment 2 had previously participated in Experiment 1. Practice effects were not expected as there was approximately a two-year gap between experiments, and as different sentences were used for each experiment. None of the younger adults who participated in Experiment 2 had participated in Experiment 1. All participants reported English as their first language and passed a screening test for mild cognitive impairment (Nishiwaki et al., 2004).

3.3.1.2 Stimuli & Procedure

The methods and procedure used for Experiment 2 were the same as Experiment 1 except for a few differences. First, as 10 of the older participants had previously been exposed to the MAVA sentences during Experiment 1, different IEEE sentences were recorded and used as stimuli. Second, the SNR was set at -1dB (instead of -6dB) to reduce the attentional demands of auditory processing. Third, the Listening Span (Conway et al., 2005) and the Trail Making Task (Reitan, 1992) were administered to examine working memory capacity and executive function, respectively. These cognitive tasks were administered for two reasons. First, as attentional resources are important for both working memory and executive function, any age differences in performance on these tasks would support the claim that older adults have smaller attentional resource capacities that younger adults (Craik & Byrd, 1982; Heinrich, Gagné, Viljanen, Levy, Ben-David, & Schneider, 2016). Second, including cognitive measures provides the opportunity to test whether the results from the current study are consistent with existing models of the role of attention and working memory in difficult listening situations (Pichora-Fuller, Kramer, Eckert, et al., 2016; Rönnberg, Holmer & Rudner, 2019).

3.3.2 Results

3.3.2.1 Speech Perception Task

Mean correct keywords recognised for younger and older adults as a function of display condition is shown in Figure 3.4. As can be seen, speech recognition for both younger and older adults improved from the Static Condition (Younger: *M* = 66.10, *SE* = 3.58, Older: *M* = 46.10, *SE* = 3.58) to the One Talking Face Condition (Younger: *M* = 81.55, *SE* = 2.38, Older: *M* = 72.75, *SE* = 2.38). When two talking faces were presented, younger adults speech recognition was no different from the One Talking Face Condition (Younger 2TF: *M* = 79.00, *SE* = 2.78). However, older adults' speech recognition significantly reduced from the One Talking Face Condition to the Two Talking Faces Condition (Older 2TF: *M* = 60.25, *SE* = 2.78). Notably, there was no significant difference in speech recognition performance as a function of display condition for older adults who had previously participated in Experiment 1 and older adults who only participated in Experiment 2, ($F(4, 88) = 2.97$, $p = .10$, $\eta_2 = .12$).

A mixed factorial ANOVA with Face Condition (Static vs. One Talking Face vs. Two Talking Faces vs. Four Talking Faces vs. Six Talking Faces) as the within participants factor and Age (Younger vs. Older) as the between participants factor was conducted. There was a significant main effect of Age $F(1, 38) = 19.35$, $p < .001$, $\eta_2 = .34$. Younger adults ($M = 73.85$, $SE = 2.83$) recognized more key words than Older Adults ($M = 56.27$, $SE = 2.83$). There was also a significant main effect of Condition, $F(4, 152) = 63.88$, $p < .001$, $\eta_2 = .63$. Bonferroni corrected post hoc tests revealed that when the data was collapsed across age groups, there was no significant difference in speech recognition on the Four Talking Faces Condition ($M = 62.35$, $SE = 2.15$) and the Six Talking Faces Condition ($M = 60.08$, $SE = 2.57$), and neither of these conditions were significantly different than the Static Condition ($M = 56.10$, $SE = 2.53$). However, all other conditions were significantly different from each other (1TF, $M = 77.15$, $SE = 1.68$, 2TF, $M = 69.63$, $SE =1.96$).

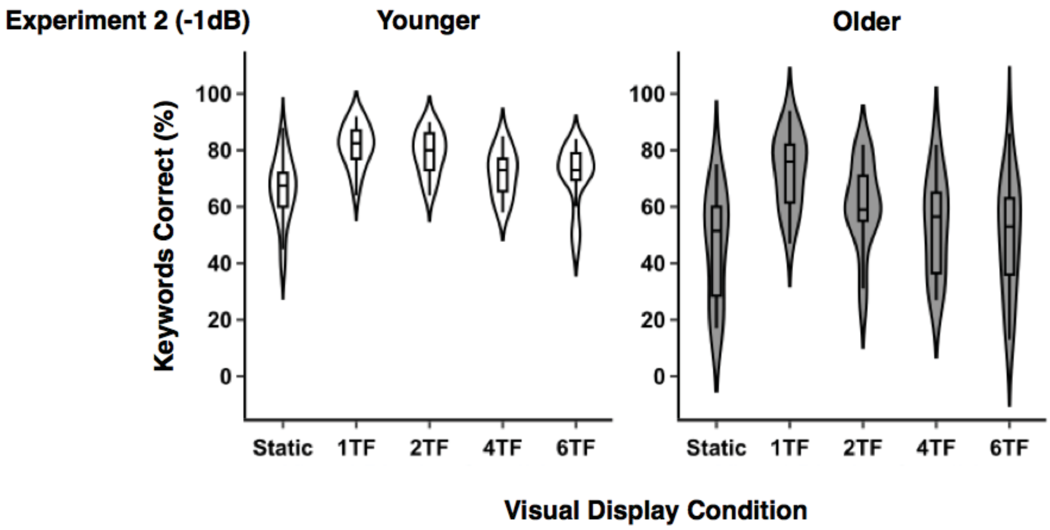

Figure 3.4. Results from Experiment 2. Tukey's box plots represent the median and interquartile range (Q3-Q1) of the percentage of keywords correctly reported as a function of visual display condition for younger and older adults. Violin plots represent the probability density of the data across the distribution.

A significant interaction between Face Condition and Age was found, $F(4, 152) = 5.84$, $p < .001$, $\eta_2 = .13$, suggesting that the effect of visual speech on speech recognition in noise was different for younger and older adults. Bonferroni corrected two-way contrasts showed that for younger adults, speech recognition during the One Talking Face Condition ($M = 81.55$, $SE = 3.54$) was significantly better than all conditions (Static, *M* = 66.10, *SE* = 3.58, 4TF, *M* = 71.70, *SE* = 3.04, 6TF, *M* = 71.00, *SE* = 3.64) except the Two Talking Faces Condition (*M* = 79.00, *SE* = 2.78). Younger adults also recognised significantly more keywords during the Two Talking Faces than the Four Talking Faces conditions, $F(4, 76) = 25.70$, $p < .001$, $\eta^2 = .58$. For older adults, speech recognition during the One Talking Face Condition (M = 72.75, SE = 2.38) was significantly better than all other conditions (Static, *M* = 46.10, *SE* = 3.58, 2TF, *M* = 60.25, *SE* = 2.78, 4TF, *M* = 53.00, *SE* = 3.04, 6TF, *M* = 49.25, *SE* = 3.64). Further, there was no significant difference between the Two Talking Faces and the Four Talking Faces conditions, $F(4, 76) = 39.88$, $p < .001$, $\eta^2 = .68$. For both age groups, there was no significant difference between the Four Talking Faces and the Six Talking Faces conditions, and neither of these conditions were significantly different from the Static Condition.

Figure 3.5 shows the percentage of the standard visual speech benefit gained for the conditions with multiple talking faces (e.g., (percentage correct on the Two Talking Faces Condition*100)/ percentage correct on the One Talking Face Condition) for younger and older adults. For the Two Talking Faces Condition, younger adults gained approximately 100% of the standard visual speech benefit, whereas older adults only gained approximately 80% of the standard visual speech benefit (Younger: *M* = 97.18, *SE* = 2.72, Older: *M* = 82.15, *SE* = 2.72). The percentage of the standard visual speech benefit gained reduced by approximately 10% when four talking faces were presented for younger and older adults (Younger: *M* = 88.19, *SE* = 2.73, Older: *M* = 71.51, *SE* = 2.73). Both age groups gained a similar percentage of the standard visual speech benefit for the Four Talking Faces Condition and the Six Talking Faces Condition (Younger: *M* = 86.85, *SE* = 3.67, Older: *M* = 65.51, *SE* = 3.67).

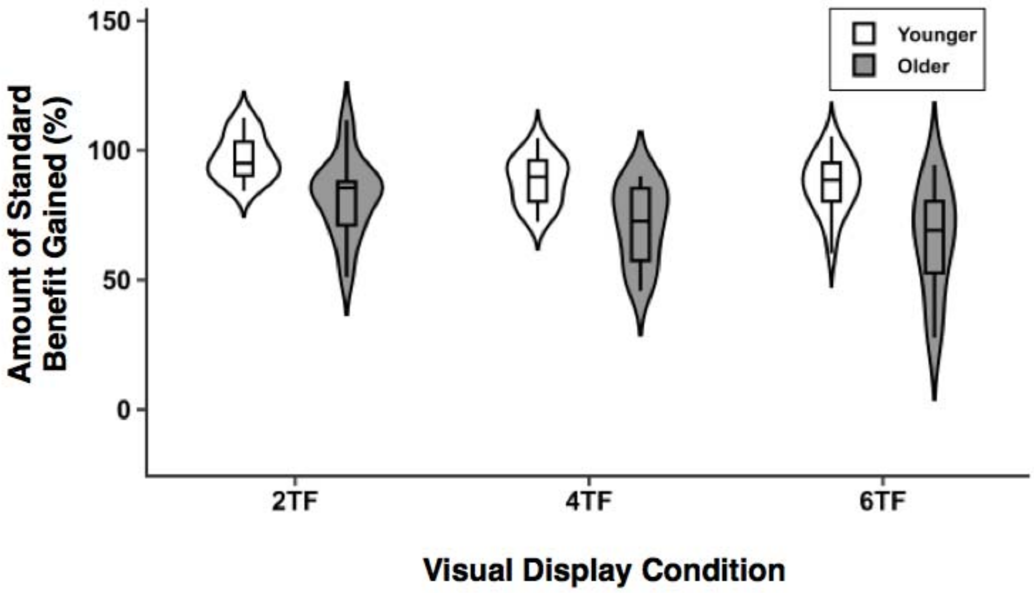

Figure 3.5. The Percentage of the Standard Visual Speech Benefit Gained as a Function of Display Condition for Experiment 2. Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution.

3.3.2.2 Visual acuity

3.3.2.2.1 Younger Adults

All younger participants had normal or corrected to normal vision (i.e., ≥1.0 on the FRACT visual acuity measure; Bach, 2007). Younger adults' visual acuity scores ranged from 1.22 to the maximum score of 2.0 (M = 1.64, SD = .25).

3.3.2.2.2 Older Adults

Six older adults had worse than normal vision (i.e., a score < 1.0 on the FrACT visual acuity measure), with visual acuity scores ranging from 0.83 to the maximum score of 2.0 (M = 1.12, SD = .22). Pearson product-moment correlation coefficients were computed to test whether visual acuity was related to performance on the speech recognition task. As found in Experiment 1, older adults' visual acuity was not related to performance on any display condition (all p values ≥ .35; r values: Static = -.10, 1TF = -.16, 2TF = -.11, 4TF = -.02, 6TF = -.23).

3.3.2.3 Hearing Sensitivity

Table 3.2 (Experiment 2) summarises hearing sensitivity levels for both younger and older adults. All younger participants had normal hearing (i.e., ≤ 25dB HL at .25, .5, 1, 2, 4 kHz). As in Experiment 1, older adults' hearing levels ranged from normal to moderately-severe hearing loss, with the majority of older adults (i.e., 9) having only mild hearing loss. Mean pure-tone hearing thresholds for each tested frequency are shown in Table 3.3 (Experiment 2). Younger adults had significantly lower thresholds than older adults at all frequencies for both ears (all p values were ≤ .01).

3.3.2.3.1 Younger Adults

Better Ear Average scores (*Min.* = 5.71, *Max.* = 17.86, M = 10.04, SD = 2.71) were not significantly related to speech recognition scores for any of the visual display conditions (all p values ≥ .18; r values: Static = -.12, 1TF = .16, 2TF =.12, 4TF = -.32, 6TF = -.16).

3.3.2.3.2 Older Adults

The Better Ear Average (*Min.* = 18.57, *Max.* = 47.14, M = 31.43, SD = 9.74) was strongly negatively correlated with performance on the Static Condition ($p <$ 0.01, r = -.73). The SNR for Experiment 2 (i.e., -1dB) likely contributed to this relationship by facilitating a wide-spread distribution for the Static Condition. That is, when the SNR was -6dB (i.e., Expeirment1), older adults' performance for the Static Condition was consistently poor, and there was no significant correlation with BEA. When performance on the Static Condition was partialled out, older adults' Better Ear Average scores were not significantly related to performance on any other condition (all p values > .45; r values: 1TF = -.03, 2TF =.16, 4TF = .06, 6TF = -.01) for Experiment 2.

3.3.2.4 Cognitive Tasks

Younger adults (*Min.* = 7.00, *Max.* = 57.00, *M* = 25.55, *SE* = 3.47) scored significantly higher on the listening span (i.e., LSPAN) than older adults (*Min.* = 0.00, *Max.* = 23.00, *M* = 9.30, *SE* = 1.77); $t(38) = 4.17$, $p < 0.01$. As can be seen in Figure 3.6, younger adults' LSPAN scores were strongly positively correlated with all conditions of the speech recognition task ($p < 0.01$; *r* values: Static = .75, 1TF = .59, 2TF = .59, 4TF = .66, 6TF = .60). That is, younger adults with larger working memory capacities recognised more keywords in noise. Younger adults' LSPAN scores were not significantly related to the percentage of the standard visual speech benefit gained for any of the conditions with multiple talking faces (all *p* values ≥ .18, *r* values: 2TF = .01, 4TF = .17, 6TF = -.31).

For older adults, LSPAN scores were moderately positively correlated with the Static ($r = .45$) and Two Talking Faces ($r = .51$) conditions ($p < 0.05$; See Figure 3.6). As can be seen in Figure 3.7, older adults' LSPAN scores were also moderately positively correlated with the percentage of the standard visual speech benefit gained for the Two Talking Faces Condition ($p < 0.05$, $r = .50$). Older adults' LSPAN scores were not significantly related to the percentage of the standard visual speech benefit gained for the Four Talking Faces Condition or the Six Talking Faces Condition (*p* values ≥ .11, *r* values: 4TF = .14, 6TF = -.37).

Scores from parts A and B of the Trail Making Test were computed to assess age differences in executive control [(Part B-PartA)/PartA]. There was no significant difference between younger and older adults' computed scores ($p = .73$) and this measure did not significantly correlate with performance on the speech recognition task for either age group (all *p* values ≥ .20; younger adult *r* values: Static = -.10,

1TF= .13, 2TF = -.23, 4TF = -.25, 6TF = -.57; older adult *r* values: Static = -.00, 1TF = .31, 2TF = .24, 4TF = .10, 6TF = .14).

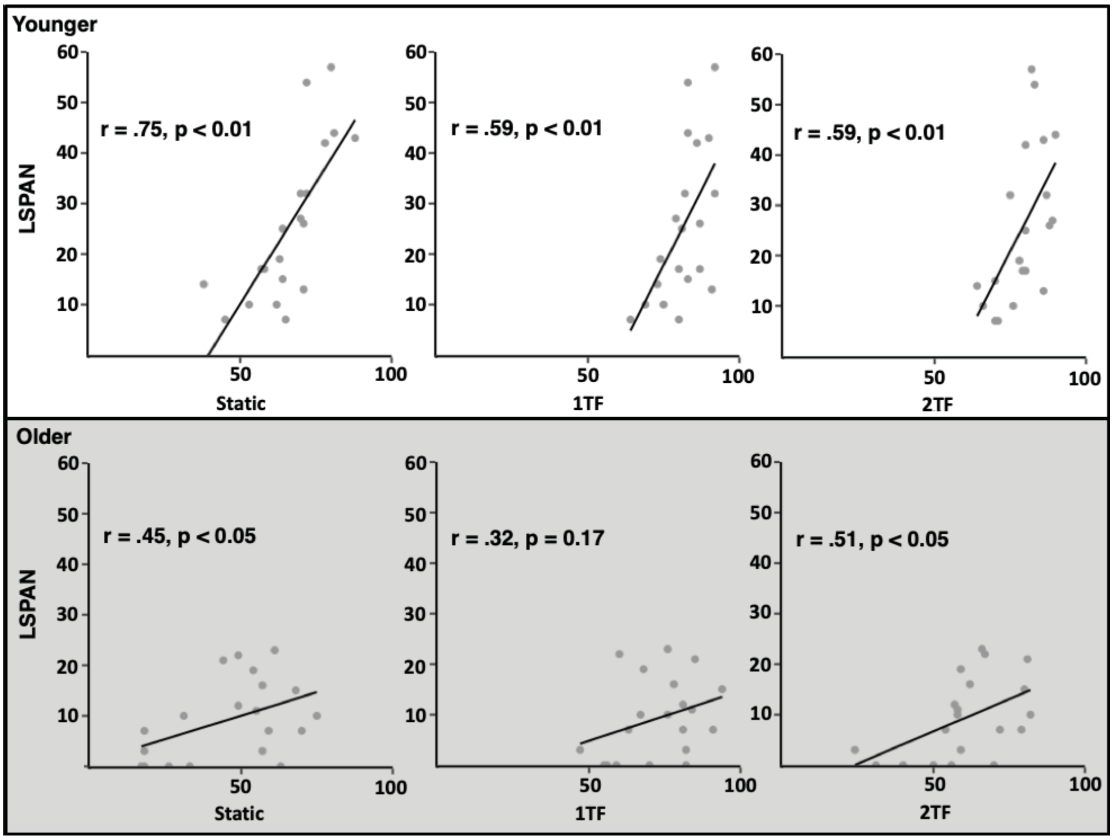

Figure 3.6. Pearson Correlations between the Listening Span and Performance on the Static, One Talking Face, and Two Talking Faces Conditions (% Correct) for Younger and Older Adults.

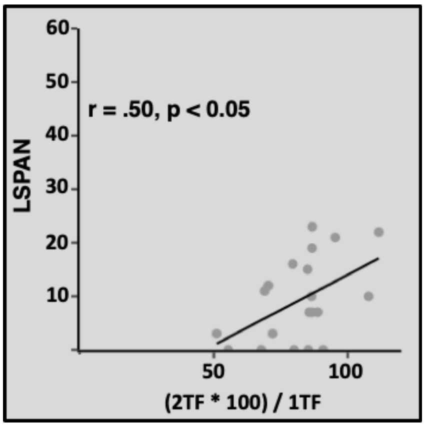

Figure 3.7. Pearson Correlation between the Listening Span and the Percentage of the Standard Visual Benefit Gained for the Two Talking Faces Condition for Older Adults.

3.3.3 Discussion

Experiment 2 tested whether the impact of presenting additional faces on the visual benefit would be different for younger and older adults when there was less masking of the auditory speech (an SNR of -1dB) than Experiment 1 (i.e., an SNR of -6dB). Both age groups had better speech recognition performance in the current experiment (-1dB SNR) compared to Experiment 1 (-6dB SNR). For older adults, the overall pattern across conditions was the same as Experiment 1. That is, speech recognition in noise improved from the Static Condition to the One Talking Face Condition (i.e., a standard visual speech benefit); and this visual speech benefit was reduced as additional talking faces were presented. As in Experiment 1, the decline in the effect was proportional to the number of talking faces presented. This result suggests that even when the auditory processing demand was reduced (i.e., Experiment 2 had a more favourable SNR than Experiment 1) older adults needed to perform a serial search in order to process the relevant visual speech.

Younger adults also showed the standard visual speech benefit (more words recognized in the One Talking Face Condition compared to the Static Condition), but unlike Experiment 1, they showed an equal visual speech benefit for the Two Talking Faces Condition. That is, there was no significant decline from the One Talking Face Condition to the Two Talking Faces Condition. It possible that a ceiling effect contributed to this pattern of results, as many younger participants' speech recognition performance was at or near 100% for both the One Talking Face Condition and the Two Talking Faces Condition. However, considering the interpretation of Stacey et al.'s (2014) results, it is also possible that younger adults had sufficient attentional resources available to process visual speech from two taking faces at once.

The difference between younger and older adults shown in the current experiment could be interpreted as being due to younger adults having a greater attentional resource capacity than older adults. An alternative (not necessarily exclusive) interpretation is that the group difference was related to age-related changes in hearing sensitivity. Given that older adults' speech recognition was relatively poorer than younger adults' in the control condition, it could be argued that -1dB was a more attentionally demanding SNR for older adults than for younger adults. As such, older adults might not have had spare attentional resources to devote to visual processing or gaining a visual speech benefit when two talking faces were presented.

The claim that younger adults have more attentional resources than older adults is supported by the results from the LSPAN (i.e., younger adults performed significantly better on the LSPAN than older adults). Moderate and strong positive correlations were also found between the LSPAN and performance on the speech recognition in noise task for both age groups. These correlations are consistent with models of speech understanding in difficult listening situations, which suggest that when speech perception becomes difficult (due to background noise or hearing loss), cognitive resources (e.g., working memory) are recruited to help resolve perceptual ambiguity (Pichora-Fuller, Alain, & Schneider, 2017; Rönnberg, Holmer, & Rudner, 2019; Wingfield, Tun, & McCoy, 2005).

A novel finding from Experiment 2 is that older adults' LSPAN scores were moderately-positively correlated with the percentage of the standard visual speech benefit gained for the Two Talking Faces Condition. That is, older adults who performed better on the LSPAN gained a larger visual speech benefit for the Two Talking Faces Condition (i.e., a benefit closer to the standard benefit) than older adults who performed poorer on the LSPAN. One interpretation of this finding is that, as attentional resource capacity likely influences performance on the LSPAN (Engle, 2002; Cowan,1999; Wingfield, Amichetti, & Lash, 2015), older adults who performed better on the LSPAN may have had sufficient attentional resources to process two talking faces at once for at least some trials, and thus gained a visual speech benefit closer to that of the standard benefit for the Two Talking Faces Condition. However, within-group differences in auditory processing abilities that are independent of hearing acuity (e.g., temporal processing; Pichora-Fuller & MacDonald, 2007) could have influenced older adults LSPAN performance in addition to (or instead of) attentional resources, and in turn older adults' performance on the Two Talking Faces Condition. Future research should examine the relationship between measures of working memory capacity and the visual speech benefit using an experimental design that allows for direct comparison between low and high working memory capacity groups of normal hearing, younger and older adults (e.g., Gordon-Salant & Cole, 2016).

3.4 General Discussion

Seeing a talker provides a sizeable benefit when recognizing speech in noise. The current study investigated what happens to this benefit, for both younger and older adults, when the number of talking faces presented and the SNR are manipulated. In Experiment 1, both younger and older adults gained the largest visual speech benefit from a single talking face and smaller benefits were gained for the Two Talking Faces and Four Talking Faces conditions. The results from Experiment 1 clearly show that neither younger or older adults gained a full visual speech benefit when multiple talking faces were presented, even though the visual speech from one talking face always matched the auditory signal. A likely explanation for this pattern of results is that combining auditory and visual speech information requires attentional resources, and when these resources need to be devoted to auditory processing, combing auditory and visual information is done in a serial fashion by directing visual-spatial selective attention to only one talker's face.

The manipulation used in Experiment 2 (i.e., making the auditory speech signal clearer) was based on the findings of Stacey et al., (2014) that suggest that clearer auditory speech releases attentional resources, allowing for auditory and visual information from more than a single face to be combined. The results from Experiment 2 showed that when the SNR was -1dB, the magnitude of the visual speech benefit for the One Talking Face Condition and the Two Talking Faces Condition was the same for younger adults. However, as in Experiment 1, the visual speech benefit reduced by approximately 50% for older adults when two talking faces were presented. This is consistent with Stacey et al. (2014) and suggests that when the SNR was less adverse, younger adults had sufficient attentional resources to combine auditory and visual information from two talking faces at once (i.e., the scope of visual-spatial selective attention encompassed both faces regardless of where the perceiver was foveating), whereas older adults only had sufficient resources to attend to one talking face at a time (i.e., the scope of visual-spatial selective attention only encompassed the area of one face). However, it is also possible that younger adults' performance on the One Talking Face Condition and the Two Talking Faces Condition was the same for Experiment 2 because of a ceiling effect.

So far, the way we have interpreted our results is consistent with the interpretations of studies that have used visual search paradigms to test the involvement of selective attention in auditory-visual processing effects (e.g., Alsius & Soto-Faraco, 2011; Fujisaki, Koene, Arnold, Johnston, & Nishida, 2005; Stacey et al., 2014; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008). That is, we have taken evidence of a serial search (i.e., that the visual speech benefit declines as additional talking faces are presented) as evidence for the deployment of visual-spatial selective attention. However, if in a serial visual search attention is constrained to where a person is looking (i.e., foveating) so that in order for each stimulus to be processed it needs to be foveated, then it is possible that age-related declines in processing-speed could have contributed to older adults' inability to gain a standard visual speech benefit for the Two Talking Faces Condition in Experiment 2.

For example, if older adults started the serial visual search with foveating (and thus processing) a face that did not match the auditory signal then, due to age-related declines in processing-speed or ocular-motor control, they would likely be slower than younger adults to switch to the relevant talking face. Thus, it is possible that younger adults were able to switch quickly enough from looking at the irrelevant face to looking at the relevant face to gain a standard visual speech benefit. On the other hand, by the time older adults were able to switch, any benefit was significantly reduced.

Hearing loss could have exacerbated reduced processing-speed for older adults in that if the auditory signal was unclear, then it may have been difficult for older adults to decide if an auditory signal matched a talking face or not. In contrast, the absence of hearing loss would have facilitated younger adults' ability to hear the auditory speech signal, which could have facilitated younger adults' ability to quickly decide if they needed to switch to processing an alternate face (i.e., if they were looking a talking face that was irrelevant to the auditory signal).

There are, however, several potential problems with a processing-speed focused interpretation of our results. First, since for the younger adults in Experiment

2 there was no apparent switching cost for the Two Talking Faces condition, younger adults would need to carry out an extremely rapid rejection (of the irrelevant talking face) and switch (to the relevant talking face). Such cost-free switching seems inconsistent with the result that there was a switching cost for the Four Talking Faces Condition (i.e., speech recognition declined for the Four Talking Faces Condition in comparison to the One Talking Face Condition).

Second, our results do not support the suggestion that hearing loss prevented older adults from being able to rapidly relate auditory and visual speech information (and thus to decide if a talking face matched the auditory signal or not) for the Two Talking Faces Condition in Experiment 2. Rather, as older adults gained the same amount of benefit for the One Talking Face Condition in comparison to younger adults for both experiments 1 and 2, our results suggest that, despite poorer hearing sensitivity, older adults are able to rapidly use visual speech information to assist with auditory speech processing. This interpretation is consistent with studies that have shown that people can relate auditory and visual speech information even when the speech signal is masked so that speech is barely detectable (e.g., at SNR levels from -20 to -30dB; Grant & Seitz, 2000, Kim & Davis, 2004). Furthermore, there was not a significant correlational relationship between Better Ear Average scores and older adults' performance on the Two Talking Faces Condition (when performance on the One Talking Face Condition was partialled out), which would be expected if hearing loss was underlying the differential effect of age that was found for Experiment 2.

Third, a processing-speed focused interpretation of our results relies on the assumption that perceivers can only gain a visual speech benefit when they are foveating visual speech that matches the auditory signal (i.e., that attention is only deployed to the location or object that an individual is foveating). This assumption goes against studies that show that attention can be directed independent of eye-gaze (Posner, 1980). Indeed, studies showing that an auditory-visual speech benefit can be accrued when a talking face is presented in the visual periphery (e.g., Kim & Davis, 2013; Paré et al., 2003), and when the quality of a face's spatial frequency is low (e.g., Munhall, Kroos, Jozan, & Vatikiotis-Bateson, 2004) support the suggestion that attention is not necessarily deployed to the location or object that an individual is foveating.

Taken together, the reduction in the visual speech benefit for the Two Taking Faces Condition that was observed for older adults but not for younger adults (i.e., Experiment 2) was likely due, at least in part, to cognitive ageing rather than auditory ageing alone. Although visual-spatial selective attention may be important for gaining a visual speech advantage, processing-speed and/or oculo-motor functioning may also contribute. Future research investigating ageing and the visual speech benefit should present both relevant and irrelevant visual speech information and collect eye movement data to confirm any differences in visual search strategies between younger and older adults, with and without hearing loss.

Chapter 4

The visual speech benefit in noise: Effects of listener age, seeing two talkers and spatial cueing

4.1 Introduction

Older adults with and without hearing loss have difficulty understanding speech in noise (Bernstein & Grant, 2009; Füllgrabe, Moore, & Stone, 2015). Although research suggests that this difficulty can be offset when a talker's face can be seen (i.e., the visual speech benefit in noise; Tye-Murray, Sommers & Spehar, 2007), the auditory-visual trials used for most research on ageing and the visual speech benefit only display a video of a single person producing visual speech that matches the auditory signal. That is, auditory-visual trials for speech recognition in noise tasks are typically not demanding on visual-spatial selective attention, since there is only one face to look at (i.e., foveate) and attract attention. Given research indicates that older adults are more susceptible to visual distraction than younger adults (e.g., Madden, Connelly & Pierce, 1994; Guerreiro, Murphy, & Van Gerven, 2013), it seems possible that age-related changes in how visual attention is allocated and directed could reduce older adults' ability to gain a visual speech benefit when a visual scene includes visual distractors. However, this potential reduction in the visual speech benefit for older listeners would not have been observed by speech perception studies that have only presented a single talking face.

To investigate whether the visual speech benefit is reduced when additional visual elements are included in a visual display, the experiments in the previous chapter tested the effect of presenting multiple talking faces to both older and younger adults. As reported in Chapter 3, when an auditory signal mixed with speech shaped noise at -1dB was presented with two talking faces (one that matched the auditory signal and one that did not), younger adults gained a full visual speech benefit, whereas older adults' visual speech benefit reduced by approximately 50% in comparison to a standard one talking face condition.

One possible explanation for this differential effect of age is that combining auditory and visual speech requires attentional resources. That is, older adults (with some degree of age-related hearing loss) may have allocated the majority of their attentional resource capacity to auditory processing rather than visual processing, which could have narrowed the scope of visual-spatial attention so that only one talking face could be processed (and combined with the auditory signal) at a time. Younger adults with good hearing, however, may have had sufficient attentional resources to allocate to visual processing so that the scope of visual-spatial attention was broad enough to process (and combine the auditory signal with) two talking faces at once.

An alternative explanation for the differential effect of age found when two talking faces were presented is that older adults were slower than younger adults at switching their eye-gaze between the two talkers. That is, if a participant started a trial from the two talking faces condition by looking at the face that did not match the auditory signal, a younger adult may have been able to decide that they were looking at the incorrect face and switch to the alternate face fast enough to gain a full visual speech benefit. An older adult, on the other hand, may have been slower to decide that they were looking at the wrong face (and slower to switch), and would therefore gain less of a visual speech benefit.

One problem with this explanation is that it relies on the assumption that the scope of visual-spatial attention is constrained to where a person is looking (i.e., foveating). This assumption is inconsistent with studies showing that visual speech can enhance speech perception in noise when it is presented in the visual periphery (e.g., Kim & Davis, 2013; Paré et al., 2003), and when the quality of the spatial frequency information is low (e.g., Munhall, Kroos, Jozan, & Vatikiotis-Bateson, 2004). Of course, it could be that rather than older adults being relatively slower in shifting eye-gaze, they are slower in shifting attention. This avoids having to claim that attention is constrained by eye-gaze; however, the proposal becomes very similar to the attention one (above), except rather than the difference between older and younger adults being attentional resource capacity, the difference would lie in the ability to either trigger rapid shifts in attention or in the speed of the shifts themselves.

To gain further insight into the underlying cause of the difference between younger and older adults' performance on the Two Talking Faces Condition, the current study tested whether visually cueing the location of a talker that matches the auditory signal would enable older adults to gain a standard visual speech benefit when two talking faces (i.e., one that matches the auditory signal and one that does not) are presented. If older adults are able to gain a standard visual speech benefit when two talking faces and a visual cue are presented, then this would suggest that a talking face needs to be foveated in order to be processed and that the older adults in Experiment 2 (Chapter 3) were poorer than the younger adults at switching their gaze/attention when they initially looked at the non-matching face. This pattern of results would also suggest that when older adults know where to look (i.e., who the target talker is), they should be able to gain a visual speech benefit even when a source of visual speech that does not match the auditory signal is within their visual field.

However, if older adults are not able to gain a standard visual speech benefit when two talking faces are presented with a salient visual cue indicating the target talker, then this would suggest that the presence of a talking face that does not match the auditory signal interfered with older adults' ability to gain a standard visual speech benefit, regardless of whether they knew where the target talker was located (i.e., where to look) or not. This pattern of results would be consistent with the inhibitory deficit hypothesis (Hasher & Zacks, 1988; Lustig, Hasher, & Zacks, 2007), which suggests that old age is accompanied by a reduced ability to ignore irrelevant stimuli.

The current study tested younger and older adults on a speech recognition in noise task. Spoken sentences mixed with speech-shaped noise were presented when there was a static image of a face (i.e., the Static Condition), one talking face relevant to the auditory signal, and two talking faces (one target and one distractor) on the screen. In addition, a visual cue (i.e., a white box) indicating where participants should look/attend was always presented one second prior to the talking face video(s) or image(s) and remained visible for the duration of each trial. When two talking faces were presented, the visual cue was either ambiguous (i.e., surrounding both target and distractor videos) or valid (i.e., surrounding only the target video). The location (i.e., right or left) of the valid cue and target talking face changed randomly throughout the experiment.

As Stacey al.'s (2014) study and the experiments in Chapter 3 suggest that the auditory processing demands of a speech recognition in noise task can differentially affect how younger and older adults deploy visual-spatial attention, stimuli for the current study were presented at two SNRs (i.e., -1dB and -4dB). Note, an SNR of -4dB was selected for the "more demanding" condition for the current experiment as speech recognition was very poor (younger adults only correctly reported 23% of the keywords for the Static Condition) when the SNR was previously set at -6dB (Experiment 1, Chapter 3).

For the current study, it was predicted that for both SNRs younger and older adults would gain a standard visual speech benefit (i.e., speech recognition would be better during the One Talking Face Condition than the auditory-only Static Condition). When the SNR was -1dB, it was predicted that younger adults' speech recognition performance on the conditions with two talking faces would not be significantly different from the One Talking Face Condition, regardless of the type of cue presented. When the SNR was more attentionally demanding (i.e., -4dB), it was predicted that younger adults would gain a standard visual speech benefit for the valid cue condition; but, as in Chapter 3 (Experiment 1), they would not be able to gain a standard benefit when two talking faces were presented and it was not clear which face was the target (i.e., the Ambiguous Cue Condition).

For older adults, it was predicted that speech recognition would be poorer when the ambiguous cue was presented in comparison to the One Talking Face Condition for both SNRs (i.e., -1dB and -4dB). Additionally, as older adults are more susceptible to visual distraction than younger adults, particularly when the primary task involves auditory processing (Guerreiro, Murphy, & Van Gerven, 2013), it was predicted that when two talking faces were presented, older adults would not be able to gain a full visual speech benefit, regardless of the cue presented.

Lastly, as research suggests that attentional resource capacity moderates performance on complex span tasks (Cowan et al., 2005), the Listening Span (i.e., LSPAN) and the Symmetry Span (i.e., SSPAN) were administered. It was predicted that if older adults have smaller attentional resource capacities than younger adults, then they should have significantly lower LSPAN and SSPAN scores than younger adults. It was also predicted that if attentional resource capacity affects the way visual-spatial selective attention is deployed, then, for both SNRs, younger and older adults with higher LSPAN and SSPAN scores would perform better on both the valid and ambiguous cue conditions than younger and older adults with lower LSPAN and SSPAN scores.

4.2 Method

4.2.1 Participants

Twenty-four younger adults (11 Females, $M_{\text{Age}} = 23$) and 24 older adults (14 Females, $M_{\text{Age}} = 70$ participated in this study. Younger adults were undergraduates at Western Sydney University and participated for course credit. Older adults were recruited from the community and were reimbursed $40. All participants reported English as their first language and passed a screening test for cognitive impairment (Addenbrook's Cognitive Assessment; Mioshi, Dawson, Mitchell, Arnold & Hodges, 2006).

4.2.2 Stimuli

The stimuli consisted of 147 auditory and visual recordings (128 test trials, 12 catch trials, and seven practice trials) of a native Australian-English female talker uttering Harvard IEEE sentences. These recordings were manipulated to create 4 visual display conditions (Static vs. One Talking Face vs. Ambiguous Cue vs. Valid Cue) each presented at two different SNRs (-1dB vs. -4dB) for a total of eight experimental conditions. Participants were always presented with a visual cue (i.e., a white rectangle) followed by an auditory-visual stimulus. The number of silent visual speech videos (1 or 2) included in the auditory-visual stimulus as well as the size and location of the cue varied across visual display conditions. See Figure 4.1 for a summary of the visual display conditions.

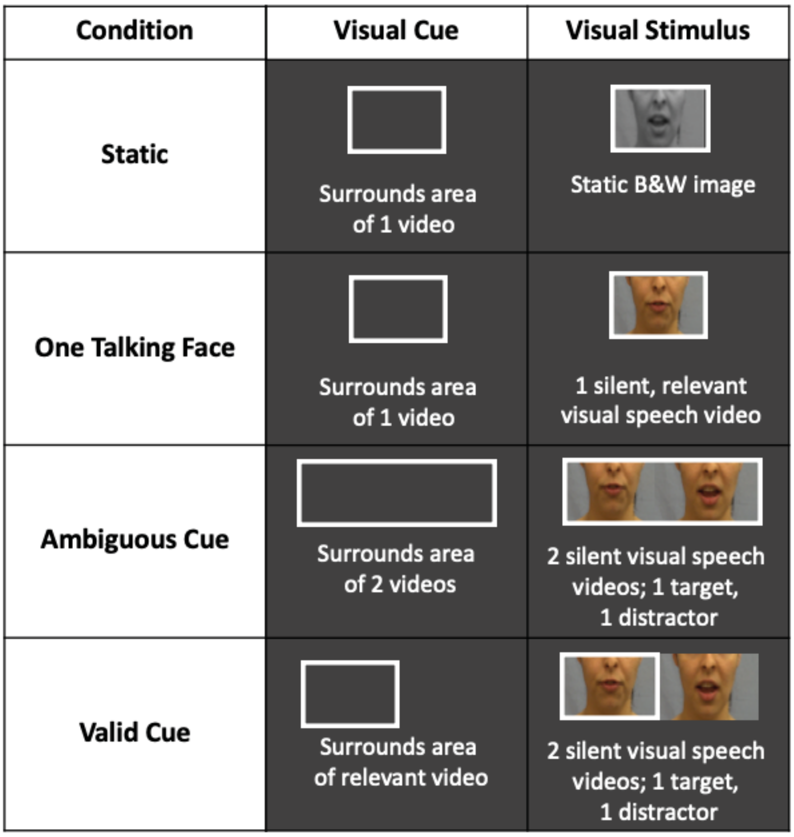

Figure 4.1. Visual Cues and Stimuli for Each Visual Display Condition. Visual Cues were always preceded by a 200ms white fixation cross. The visual cue was presented one second before the visual stimulus and remained visible until the end of each trial. The duration of each visual stimulus was approximately three seconds, but this varied depending on the content of each sentence.

The stimuli for each condition were edited using FFmpeg. All video recordings were scaled and cropped to show only the lower portion of the face, with each individual video measuring 5cm (height) x 8cm (width; visual angle 31°). Videos for the One Talking Face Condition consisted of one female uttering a single sentence. For the Static Condition, the video was always a black and white image of the female talker with her mouth closed. For both cueing conditions, sets of video pairs (i.e., a single video file with two silent visual speech videos, side-by-side, each simultaneously uttering a different IEEE sentence) were used to create two items: one with the auditory signal matching the visual speech video on the left, and one with the auditory signal matching the visual speech video on the right. For catch trials, an image of the talker's face(s) with red crosses was concatenated to the end the videos. The image with red crosses appeared for 500ms.

Two versions of the auditory recordings were created. One mixed with noise (speech shaped derived from the long-term average spectrum of the 146 sentences used) at -1dB and one mixed with noise at -4dB. The auditory sentences (mixed with noise) were mapped onto the appropriate videos, creating two sets of auditory-visual stimuli (-1 dB and -4 dB).

To create the cuing effect, a one second black video of the same dimension was concatenated to the beginning of each auditory-visual stimulus. A white border was then added to each stimulus for its total duration. The dimensions of the border varied according to the visual display condition (1TF = W: 8cm x H: 5cm x D: 0.3cm, Static = W: 8cm x H: 5cm x D: 0.3cm, Valid Cue = W: 8cm x H: 5cm x D: 0.3cm, Ambiguous Cue = W: 16cm x H: 5cm x D: 0.3cm).

4.2.3 Cognitive Tasks

4.2.3.1 The Listening Span

The Listening Span was used to measure auditory verbal working memory (Conway et al., 2005). For this task, participants listened to letter sequences ranging from 3-7 letters. Each letter in a sequence was preceded by an auditory semantic categorization task in which a sentence was presented (e.g. the train sang a song) and the participant judged whether the sentence made sense or not. At the end of each sequence, participants were instructed to recall each letter from that sequence using a provided letter matrix. The researcher performed all of the mouse clicking during the task while the participant provided oral responses (i.e., true, false, and letter sequences). Participants were instructed to adjust the volume to a comfortable level during the practice session. The LSPAN was calculated as the sum of all perfectly recalled sequences (i.e., the absolute scoring method). For example, if an individual recalled two letters in a set of two, three letters in a set of three, and four in a set of five, their absolute score would be five (i.e., 2 + 3 + 0 = 5).

4.2.3.2 The Symmetry Span

The Symmetry Span was used to measure visual spatial working memory (Conway et al., 2005). For this task, participants were shown a series of 4 x 4 grids with one square from each grid highlighted in red. Grid series ranged from 2-5 grids. Each 4 x 4 grid was preceded by an 8 x 8 grid that had a number of squares shaded in black to create a pattern. Participants were instructed to judge whether each pattern was vertically symmetrical or not. At the end of each series, participants were instructed to recall the order and location of each highlighted red square. The researcher performed all mouse clicking during the task while the participant provided oral responses related to the symmetry task and pointed to show which sections of the grid were red. The final score was calculated as the sum of all perfectly recalled sequences (i.e., the absolute scoring method).

4.2.4 Apparatus

Stimuli were presented using DMDX software (Forster & Forster, 2003) on a Dell T7810 computer with Windows 7 software. Visual stimuli were presented on a monitor measuring 30cm (height) x by 53cm (width). Auditory stimuli were presented through Sennheiser HD280pro headphones.

4.2.5 Procedure

Participants were tested individually. First, they completed a questionnaire that asked about their age, gender, and native language. Next, participants were seated approximately 70cm away from a computer monitor in a sound attenuating booth. Participants were told that for each trial they would see a white box followed by a video of a person uttering a sentence; they were told that the size and location of the box would change slightly, but that their task was to focus on the video that appeared inside the white box for each trial, listen carefully to the sentence, and type out what they heard once “respond” appeared on the screen. These same instructions appeared on the screen for the participant to read. Participants were then assigned to one of eight versions of the experiment and completed a practice session. The practice session consisted of five practice test trials (two One Talking Face, two Valid Cue, one Static) and two practice catch trials. Before the catch trials were presented, participants were instructed to look for a red cross that appeared over the talker’s mouth at the very end of each video, and to type “999” whenever they saw a red cross (instead of typing what they heard). They were told that the videos with red crosses would show up randomly throughout the experiment.

After the practice session, participants completed 128 test trials, (i.e., 16 trials from each condition and 12 catch trials) presented in a pseudo-random order. Participants were encouraged to take a short break after completing 64 trials. The total listening time for each participant was approximately 45 minutes.

After the speech perception task, participants completed a series of cognitive tasks: Addenbrooke’s Cognitive Examination (Mioshi, Dawson, Mitchell, Arnold & Hodges, 2006), The Listening Span and The Symmetry Span (Conway et al., 2005). Addenbrooke’s Cognitive Examination was always presented first, however, the presentation order of the working memory tasks was counterbalanced. Using a Diagnostic Audiometer (AD229e), hearing sensitivity was measured by obtaining pure-tone thresholds at seven frequencies (0.25,0.5, 1, 2, 4, 6, 8 kHz). Finally, visual acuity was measured using The Freiburg Visual Acuity and Contrast Test ([FrACT]; Bach, 2007).

4.3 Results

4.3.1 Speech Perception Task

Mean correct keywords recognised for younger and older adults as a function of SNR and Display Condition is shown in Figure 4.2. As can be seen in the figure, and as expected, older adults' speech recognition improved from the Static Condition to the One Talking Face Condition, declined from the One Talking Face Condition to the Ambiguous Condition, and did not improve when the valid cue was presented for both SNRs. Also consistent with our expectations, younger adults gained a standard visual speech benefit and, although less clearly represented in the figure, were able to benefit from the valid cue (i.e., speech recognition performance for the One Talking Face and the Valid Cue Condition was not significantly different) for both SNRs. However, in contrast to the results from Chapter 3 (and thus the hypotheses for the current study), younger adults' speech recognition for the Ambiguous Cue Condition was reduced in comparison to the One Talking Face Condition for both SNRs.

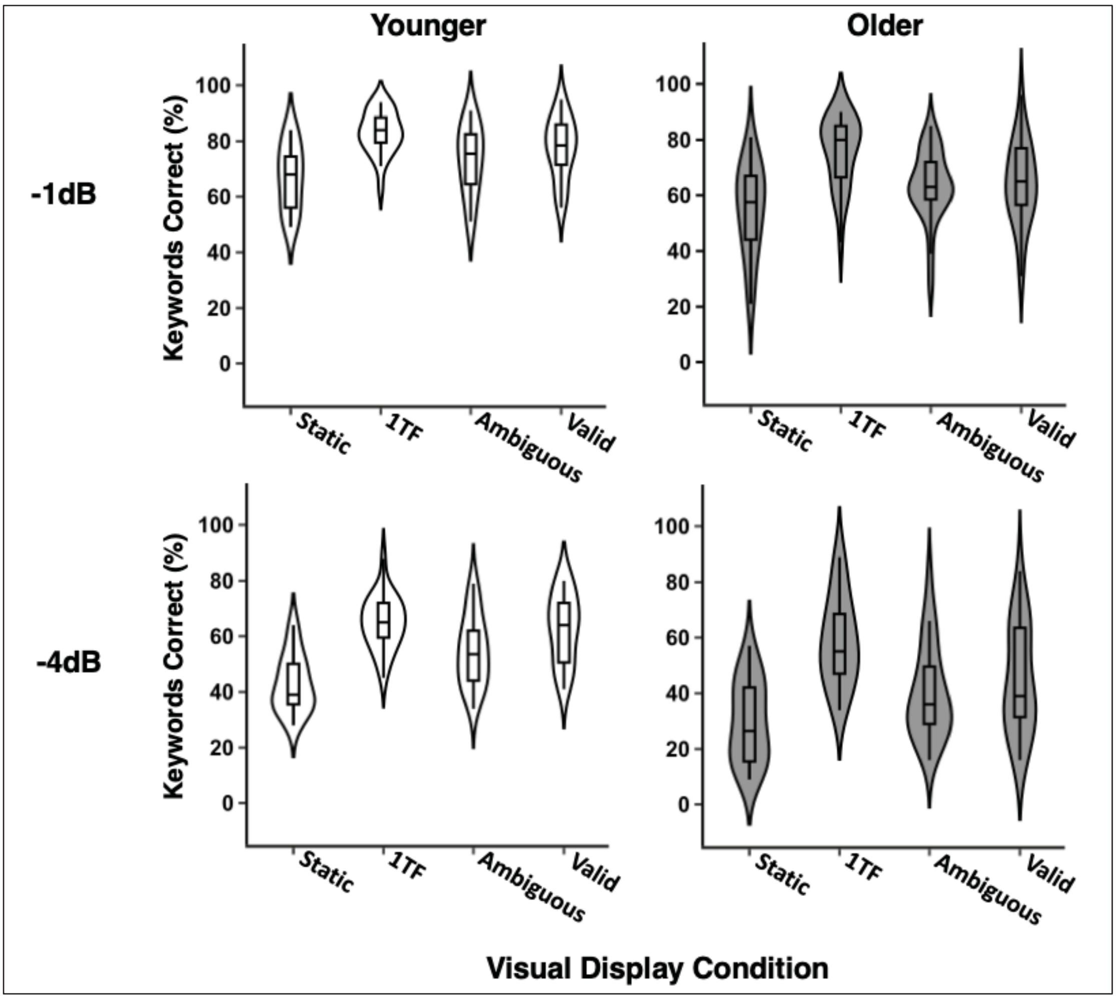

Figure 4.2. Cueing Experiment Results. Tukey's box plots represent the median and interquartile range (Q3-Q1) of the percentage of keywords correctly reported as a function of SNR and visual display condition for younger and older adults. Violin plots represent the probability density of the data across the distribution.

To test the effect of cueing the target talker on speech recognition in noise for younger and older adults, a mixed repeated measures ANOVA with Display Condition (Static vs. One Talking Face vs. Ambiguous Cue vs. Valid Cue) and SNR (-1dB vs. -4dB) as within participants factors and Age (Younger vs. Older) as the between participants factor was conducted. Follow-up analyses with Bonferroni adjusted alphas were conducted for significant interactions.

Significant main effects of Age and SNR were found. That is, Younger Adults (M = 65.73, SE = 2.14) recognized more keywords than Older Adults (M = 53.92, SE = 2.14), $F(1, 46) = 15.31$, $p < .001$, $\eta = 0.25$, and when the data was collapsed across age groups, participants recognised more keywords when the SNR was -1dB (M = 69.84, SE = 1.48) than when the SNR was -4dB (M = 49.81, SE = 1.77), $F(1, 46)$ = 266.08, $p < .001$, η = 0.85.

A significant main effect of Display Condition was also found, $F(3, 138)$ = 79.04, $p < .001$, η = 0.63. Performance on the Static Condition (M = 48.66, SE = 1.73) was significantly poorer than performance on all other conditions (1TF: M = 70.57, SE = 1.52, $p < .001$; Ambiguous Cue: M = 57.80, SE = 1.78, $p < .001$; Valid Cue: M = 62.27, SE = 1.96, $p < .001$) and performance on the One Talking Face Condition (M = 70.57, SE = 1.52) was significantly greater than all other conditions (Static: M = 48.66, SE = 1.73, $p < .001$; Ambiguous Cue: M = 57.80, SE = 1.78, $p < .001$; Valid Cue: M = 62.27, SE = 1.96, $p < .001$). Further, performance on the Ambiguous Cue Condition (M = 57.80, SE = 1.78) was significantly poorer than both the One Talking Face (M = 70.57, SE = 1.52, $p < .001$) and the Valid Cue (M = 62.27, SE = 1.96, p = .003) conditions.

A significant interaction between Display Condition and Age was found, $F(3, 138)$ = 3.16, $p < .05$, η = 0.64, suggesting that speech recognition in noise across conditions was different for younger and older adults. Younger adults showed a significant main effect of Display Condition, $F(3, 69)$ = 39.04, $p < .001$, η = 0.63. That is, speech recognition for the Static Condition (M = 54.96 , SE = 1.97) was significantly poorer than speech recognition for the One Talking Face Condition (M = 73.98, SE = 1.51, $p < .001$), Ambiguous Cue Condition (M = 63.94, SE = 2.35, $p <$ .001), and the Valid Cue Condition (M = 70.06, SE = 2.15, $p < .001$). Performance on the Ambiguous Cue Condition (M = 63.94, SE = 2.35) was significantly poorer than performance on the Valid Cue Condition (M = 70.06, SE = 2.15, p = .007) and the

One Talking Face Condition (*M* = 73.98, *SE* = 1.51, *p* = .001), which were not significantly different from each other (*p* = .227).

Older adults also showed a significant main effect of Display Condition, *F*(2.15, 49.37) = 42.56, *p* =.000, η = 0.65 (Greenhouse-Geisser correction was used as Mauchly's test of sphericity was violated, X_2(2) = 11.97, *p* = .04). That is, speech recognition performance during the Ambiguous Cue Condition (*M* = 51.48, *SE* = 3.27) and the Valid Cue Condition (*M* = 54.45, *SE* = 3.27) was not significantly different (*p* = .685); however, performance on both of these conditions was significantly poorer in comparison to the One Talking Face Condition (*M* = 67.17, *SE* = 2.63; Ambiguous-One Talking Face: *p* = .007; Valid-One Talking Face: *p* = .001). Similarly to younger adults, performance on the Static Condition (*M* = 42.35, *SE* = 2.85) was significantly poorer than performance on the One Talking Face (*M* = 67.17, *SE* = 2.63, *p* =.000, Ambiguous Cue (*M* = 51.48, *SE* = 3.27, *p* = .007) and Valid Cue (*M* = 54.45, *SE* = 3.27, *p* = .001) conditions.

A significant interaction between Display Condition and SNR was found, suggesting that, when the data was collapsed across age groups, speech recognition in noise performance across conditions was different for each SNR, *F*(3, 138) = 8.17, *p* < .001, η = 0.15. Data from trials presented at -1dB showed a significant main effect of Condition, *F*(3, 141) = 39.46, *p* < .001, η = 0.46. That is, there was no significant difference between performance on the Ambiguous Cue Condition (*M*=68.23, *SE*= 1.94) and the Valid Cue Condition (*M*=70.77, *SE*= 2.09, *p* = .578), and performance for both of these conditions was significantly poorer than performance for the One Talking Face Condition (*M* = 79.34, *SE* = 1.57, *p* < .001). Performance on the Static Condition (*M* = 79.34, *SE* = 1.57) was significantly poorer than performance on all other conditions (all *p* values ≤ .001).

Data from trials presented at -4dB also showed a significant main effect of Condition, $F(3, 141) = 76.67$, $p < .001$, $\eta = 0.62$. That is, when the SNR was -4dB, participants understood significantly more keywords during the Valid Cue Condition (*M* = 53.77, *SE* = 2.65) than the Ambiguous Cue Condition (*M* = 47.38, *SE* = 2.37; $p = .001$), and performance on both of these conditions was significantly greater than the Static Condition (*M* = 36.33, *SE* = 2.03; $p < .001$). Performance on the One Talking Face Condition (*M* = 61.76, *SE* = 1.92), was significantly greater than all other conditions (all *p* values $\leq .001$).

In contrast to the hypotheses, the repeated measures ANOVA indicated that a three-way interaction between Age, SNR, and Display Condition was not significant, $F(3, 138) = 1.42$, $p = .238$, $\eta = 0.03$).

4.3.2 Visual Acuity

4.3.2.1 Younger Adults

All younger participants had normal or corrected to normal vision (i.e., ≥1.0 on the FrACT visual acuity measure). Younger adults' visual acuity scores ranged from 1.10 to 1.68, where the maximum score is 2.0 (*M* = 1.45, *SD* = .20).

4.3.2.2 Older Adults

Nine older adults had worse than normal vision (i.e., < 1.0 on the FrACT visual acuity measure) with visual acuity scores ranging from 0.61 to 1.61 where the maximum score is 2.0 (*M* = 1.08, *SD* = .23). Pearson product-moment correlation coefficients were computed to test whether visual acuity was related to performance on the speech recognition task. The results indicated that older adults' visual acuity scores were not significantly related to performance on any of the conditions (all *p* values ≥ .29; *r* values -1dB: Static = .01 1TF = .14, Ambiguous Cue = .00, Valid Cue

= .05, r values -4dB: Static = -.07, 1TF = .19, Ambiguous Cue = .22, Valid Cue = .19).

4.3.3 Hearing Sensitivity

Table 4.1 summarises the hearing sensitivity levels for both younger and older adults. All younger participants had normal hearing (i.e., ≤ 25dB HL at .25, .5, 1, 2, 4 kHz). Older adults' hearing levels were more diverse, ranging from normal to moderate-severe hearing loss (i.e., > 40dB and ≤ 70dB HL at .25, .5, 1, 2, or 4 kHz in the better ear), with the majority of older participants having normal hearing (9 participants) or mild hearing loss (8 participants). Mean pure-tone hearing thresholds for each tested frequency are shown in Figure 4.3. Younger adults had significantly lower thresholds than older adults for both ears at all tested frequencies (all *p* values ≤ .05) except for .25 kHz for the left ear ($t(46) = -1.91$, $p = .06$).

Table 4.1
Hearing Levels for Younger and Older Adults

Hearing Level	Definition	Younger (n=24)	Older (n=24)
Normal	≤ 25[1] at all frequencies[2]	24	9
Mild Loss	>25 - ≤ 40 at one frequency	0	8
Moderate Loss	> 40 - ≤ 55 at one frequency	0	5
Moderate-Severe Loss	> 55 - ≤ 70 at one frequency	0	2

Note. Hearing level definitions adapted from Wayne et al., 2016 and are measured from the better ear.
[1]dB Hearing Loss
[2].25, .5, 1, 2, 4 kHz

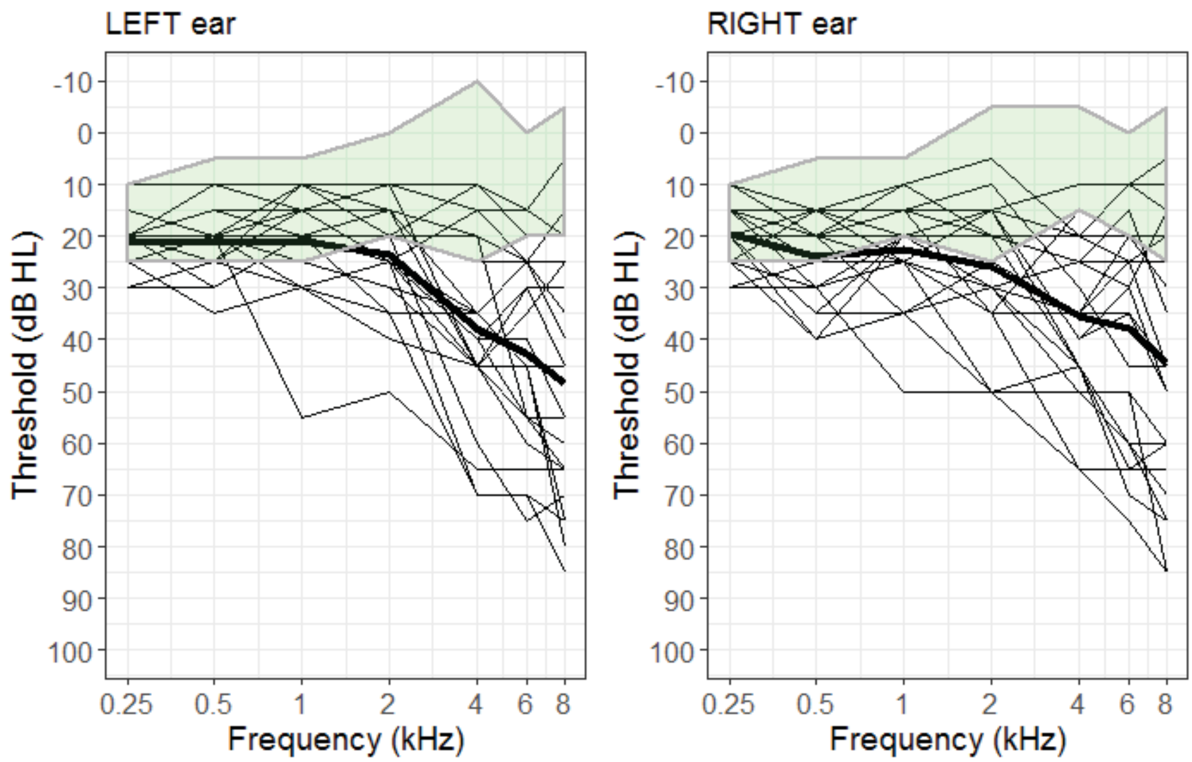

Figure 4.3. Audiogram Results for the Left and Right Ears. The bold black line represents the mean threshold for older adults as a function of frequency. The fine black lines represent individual audiograms for older adults as a function of frequency. The green shaded area represents the audiometric threshold range for younger adults.

Better Ear Average scores were calculated by averaging hearing thresholds across all tested frequencies for each ear and selecting the lower average threshold. The within group variation for the Better Ear Average was greater for older adults (*Min.* = 12.14, *Max.* = 47.86, *M* = 28.24, *SD* = 9.41) than younger adults (*Min.* = 5.00, *Max.* = 18.57, *M* = 10.83, *SD* = 3.46). For younger adults, the Better Ear Average was not significantly related to performance on any of the display conditions (all p values ≥ .20; r values -1dB: Static = -.03, 1TF = -.19, Ambiguous Cue = .05, Valid Cue = .02; r values -4dB: Static = .27, 1TF = -.01, Ambiguous Cue = -.05, Valid Cue = -.15). For older adults, the Better Ear Average was significantly related to performance on the One Talking Face Condition at -1dB (r = -.63, p < .01) and the Static Condition at -4dB (r = -.41, p < .05). The Better Ear average was not significantly related to performance for any other condition of the speech recognition task (all p values <

.08; *r* values -1dB: Static = -.39, Ambiguous Cue = -.18, Valid Cue = -.36; *r* values -4dB: 1TF = -.37, Ambiguous Cue = -.36, Valid Cue = -.37).

4.3.4 Cognitive Tasks

4.3.4.1 The Listening Span

Younger adults (*Min.* = 9.00, *Max.* = 49.00, *M* = 26.67, *SD* = 10.89) scored significantly higher on the listening span (i.e., LSPAN) than older adults (*Min.* = 0.00, *Max.* = 31.00, *M* = 11.40, *SD* = 9.14); $t(46) = 5.14$, $p < .001$). Younger adults' LSPAN scores were not significantly related to performance on the speech recognition task (all *p* values ≥ .23; *r* values -1dB: Static = .18, 1TF = .25, Ambiguous Cue = -.17, Valid Cue = .04, *r* values -4dB: Static = .01, 1TF = -.01, Ambiguous Cue = -.01, Valid Cue = -.12).

As can be seen in Figure 4.4, older adults' LSPAN scores were moderately positively correlated to performance on the One Talking Face Condition at -4dB (*p* = 0.02, *r* = 0.48) and the Ambiguous Cue Condition at -1dB (*p* = 0.01, *r* = 0.50). Older adults LSPAN scores were weakly positively correlated to performance on the Valid Cue Condition at -4dB (*p* = 0.05, *r* = 0.39) but not significantly related to performance on any other condition for either SNR (all *p* values ≥ .09; *r* values -1dB: Static = .12, 1TF = .27, Valid Cue = .25, *r* values -4dB: Static = .27, Ambiguous Cue = .34).

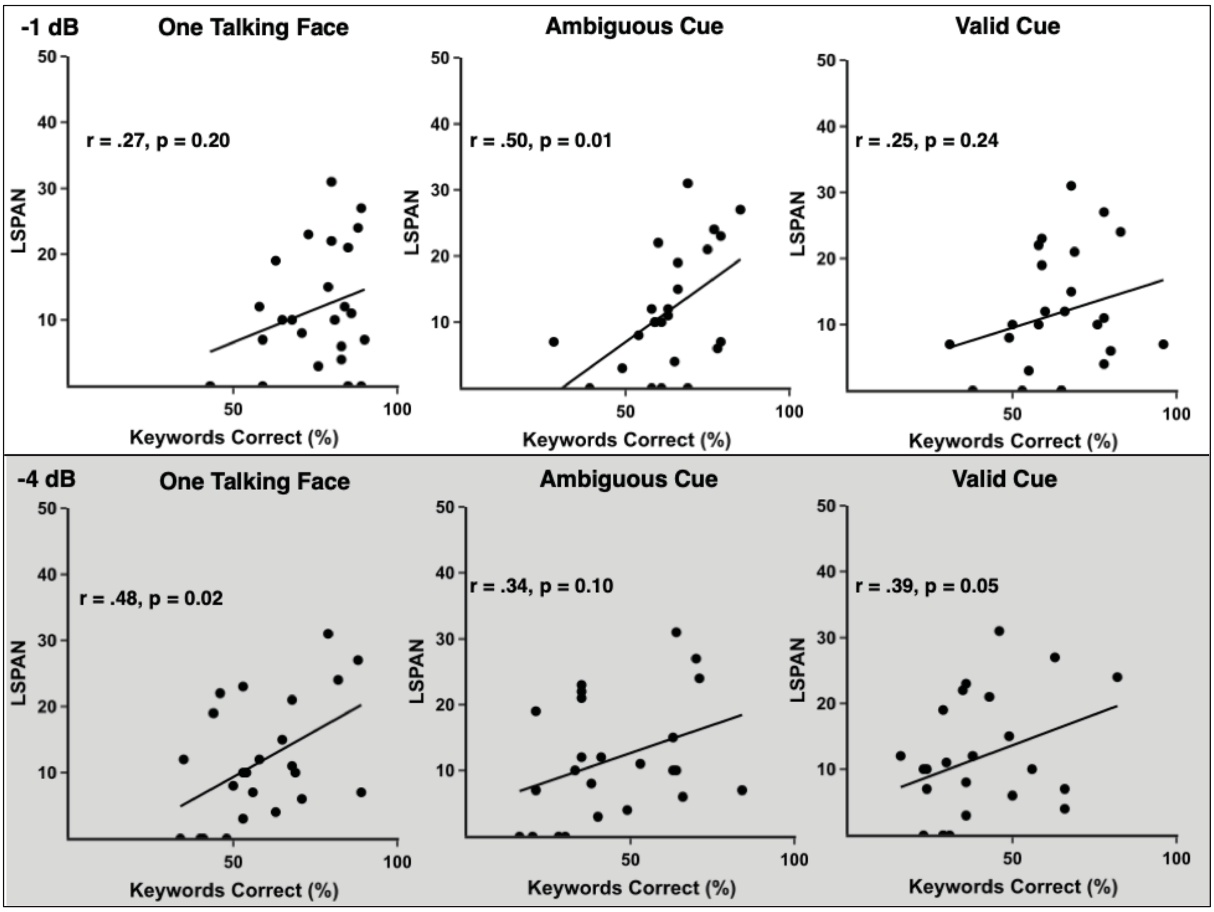

Figure 4.4. Pearson Correlations Between the Listening Span and Performance on the One Talking Face, Ambiguous Cue, and Valid Cue Conditions for Older Adults. The white background indicates -1dB SNR and the grey background -4dB SNR.

4.3.4.2 The Symmetry Span

Younger adults (*Min.* = 4.00, *Max.* = 39.00, *M* = 19.96, *SE* = 1.82) scored significantly higher on the symmetry span (i.e., SSPAN) than older adults (*Min.* = 0.00, *Max.* = 24.00, *M* = 8.96, *SE* = 1.17); $t(46) = 4.83$ $p < .001$.Younger adults' LSPAN scores were not significantly related to performance on the speech recognition task (all p values $\geq .21$; r values -1dB: Static = -.17, 1TF = -.26, Ambiguous Cue = -.17, Valid Cue = -.02, r values -4dB: Static = .15, 1TF = .09, Ambiguous Cue = .10, Valid Cue = .16). Older adults' SSPAN scores were not significantly related to performance on the speech recognition task (all p values $\geq .10$; r values -1dB: Static = -.01, 1TF = .04, Ambiguous Cue = .35, Valid Cue = .05, r values -4dB: Static = -.14, 1TF = -.10, Ambiguous Cue = -.08, Valid Cue = -.18.

4.4 Discussion

The main aim of the current study was to test whether visually cueing a talker's face that matches the auditory signal (i.e., a target face) helps younger and older adults get a standard visual speech benefit when two talking faces (one target and one distractor) are presented. In summary, the results indicated that younger adults were able to benefit from the cue, whereas older adults were not. That is, when two talking faces were presented and a visual cue surrounded only the target talker's face, younger adults' speech recognition in noise performance was not significantly different from when only one talking face was presented. In contrast, older adults recognised significantly less speech in noise when two talking faces were presented in comparison to when one talking face was presented, regardless of whether the visual cue was valid (i.e., surrounded just the target talker) or ambiguous (i.e., surrounded both target and distractor talkers).

Although it possible that older adults did not gain a standard visual speech benefit for the Valid Cue Condition because they were not able to perceive the cue as efficiently as younger adults, this explanation seems unlikely as the colours of the cue and the background had a high contrast (i.e., a white cue on a black background) and as older adults' visual acuity scores (as measured by the FrACT) were not related to performance on any of the experimental conditions. Age-related declines in the speed at which older adults are able to shift their gaze to a cued location also seems to be an unlikely explanation for older adults' performance on the Valid Cue Condition. That is, even though older adults are generally slower at overtly switching their gaze to a location in the periphery from a central fixation point than younger adults, both younger and older adults can complete a prosaccade (i.e., a saccade in the direction that a cue has previously indicated) to a target at up to 8º eccentricity within 200 ms

(Bojko, Kramer, & Peterson, 2004; Brett & Machado, 2017; Wang, Tian, Wang, & Benson, 2013). As the valid cue for the current experiment was only 3º from fixation and presented for 1000 ms prior to stimulus onset, older adults should have been able to adjust their gaze to the validly cued location before the auditory-visual stimulus was presented.

It is also possible that age-related declines in covertly (i.e., independent of eye gaze) switching visual-spatial attention could have prevented older adults from gaining a standard visual speech benefit for validly cued trials (Erel & Levey, 2016). However, this interpretation is inconsistent with studies that have tested the effects of ageing on orienting visual-spatial attention by adapting Posner's (1980) attentional cueing paradigm (e.g., Folk & Hoyer, 1992; Langley, Friesen, Saville & Ciernia, 2011). Langley et al. (2011), for example, showed that older (i.e., between the ages of 60-74) adults' ability to covertly orient visual-spatial attention was not significantly different from younger adults when, as in the current study, a valid peripheral cue was presented for 1000 ms, and remained visible for the duration of a validly cued trial. When the cues in a classic cueing paradigm are presented for shorter durations (i.e., 50-200ms) and do not remain visible for the duration of a trial, older adults tend to have slower response times than younger adults; however, older adults are still able to correctly respond to validly cued trials within approximately 400-600ms (Folk & Hoyer, 1992; Langley et al., 201; Lincourt, Folk, Hoyer & 1977; Madden, Connelly, & Pierce, 1994; Tales, Muir, Bayer & Snowdren, 2002). Thus, older adults' performance on traditional cueing tasks suggests that the older adults in the current study should have had sufficient time to switch visual-spatial attention to the location of the target talker for the current study where visual cues were always presented for 1000 ms.

As older adults should have been able to switch their eye-gaze and/or scope of visual-spatial attention to the cued location prior to the presentation of the auditory-visual stimulus, we suggest that age-related declines in inhibitory control reduced older adults' ability to get a standard visual speech benefit for the Valid Cue Condition. In summary, the inhibitory deficit theory of cognitive ageing suggests that older adults are more distractible than younger adults in that they are less able to inhibit (i.e., ignore) information that is irrelevant to a goal (e.g., understanding speech in noise; Hasher & Zacks 1988; Lustig, Hasher, & Zacks, 2007). That is, older adults show impairments in preventing irrelevant information from gaining access to the focus of attention (i.e., irrelevant information is more likely to capture older adults attention than younger adults) and in the ability to filter out irrelevant information once it has reached higher levels of the processing stream (Hasher & Zacks 1988; Lustig, Hasher, & Zacks, 2007). In the context of the current study, the inhibitory deficit hypothesis would suggest that the non-matching talking face presented during trials from the Valid Cue Condition could have captured older adults visual-spatial attention, even if they had previously oriented their gaze and/or attention to the location of the valid cue. This attentional capture could have prevented older adults from combining visual information from the target talking face with the auditory signal, thus reducing the magnitude of the visual speech benefit.

As research on ageing and cross-modal distraction (e.g., Guerreiro, Murphy, and Van Gerven, 2013) suggests that the modality of a distractor in relation to the modality of a primary task affects older adults' ability to inhibit irrelevant information, it would be interesting for future research on ageing and auditory-visual speech perception to vary the characteristics of the auditory, visual, or auditory-visual distractors presented. Cohen and Gordon-Salant (2017), for example, presented younger and older listeners with different types of irrelevant visual information (i.e., text, an additional talking face, or a video of a person performing a simple action) next to a talking face that matched the auditory signal. When speech recognition performance on the conditions with visual distraction was compared to performance on standard auditory-only and auditory-visual conditions, the results indicated that for both age groups, the only visual distractor that caused speech reception thresholds (at 50% correct) to be significantly worse than the standard auditory-visual condition was the action video.

It is possible that when two talking faces were presented for Cohen and Gordon-Salant’s (2017) study, older adults were able to inhibit the distractor talking face as the target always appeared at the same location on the screen (i.e., there was no need to switch eye-gaze and/or visual-spatial attention throughout the experiment). However, there may have been another reason why the presentation of an additional talking face did not have a detrimental effect on older adults. This is because, unlike the experiments in this thesis, Cohen and Gordon-Salant (2017) presented auditory speech as an auditory masker (i.e., informational masking) and this auditory speech matched the visual speech that was presented as a visual distractor for the Two Talking Faces Condition. This synchronous auditory-visual distractor could have potentially facilitated older adults’ ability to inhibit the auditory and visual distraction (and thus gain a visual speech benefit) in addition to (or instead of) knowing where to look (Driver, 1996). Thus, future studies should compare the effects of bi-modal distractors (e.g., synchronous auditory-visual speech) and unimodal distractions (e.g., visual speech) on younger and older adults’ speech recognition in noise performance.

One difference between the results from the current study and Experiment 2 (Chapter 3) is younger adults’ performance when two talking faces were presented and the SNR was -1dB. That is, for the current study, younger adults' performance on the ambiguous cue condition was significantly poorer than performance on the One Talking Face Condition; whereas the previous experiment showed no significant difference between the One Talking Face Condition and the Two Talking Faces Condition for younger adults. One explanation for this difference between studies is that the SNRs were blocked for previous experiments but mixed for the current study (i.e., -1dB trials and -4dB trials were presented in a random order). Switching back and forth between SNRs may have been a more attentionally demanding task than listening to only one SNR, and any attentional resources that younger adults were able to devote to combining auditory information with two talking faces at once for previous experiments may have been allocated to auditory (and not visual) processing for the current study. A task that requires a participant to continuously adjust to different SNRs is arguably more realistic than a task that has a consistent SNR; thus, the results from the current study may be a more ecologically valid representation of the constraints of attention on younger adults' ability to gain a visual speech benefit in noise.

Although future studies should track participants' eye movements to verify the visual and/or attentional processes that affected older adults' performance on the Valid Cue Condition, the results from the listening span task suggest that variation in the older adults' speech recognition performance could be due to cognitive rather than sensory mechanisms. For example, when the SNR was -1dB, older adults with higher LSPAN scores recognised more speech in noise for the Ambiguous Cue Condition than older adults with lower LSPAN scores ($p = 0.01$, $r = 0.50$). This suggests that when the SNR was -1dB, older adults had sufficient attentional (or working memory) resources to combine auditory information with two talking faces at once for at least some of the trials. In addition, there was a significant (yet weak) relationship between older adults LSPAN scores and performance on the Valid Cue Condition when the SNR was - 4dB ($p = 0.05$, $r = 0.39$) but not when the SNR was -1dB ($p > 0.05$, $r = 0.25$). This suggests that when auditory processing is attentionally demanding (i.e., -4dB), older adults with a larger resource capacity were able to benefit more from the valid cue than those with a smaller resource capacity.

Together, the results from the current study and Chapter 3 suggest that the ability to gain a visual speech benefit when visual distractors are presented changes as a function of age and auditory processing demands, and that these changes may be due to the way in which visual-spatial selective attention is directed and controlled rather than deficits in gaze switching or peripheral sensory processing (i.e., hearing sensitivity or visual acuity). As older adults were not able to gain a visual speech benefit for the Valid Cue Condition (i.e., when the visual cue was clearly presented and remained accessible for the duration of each trial), it seems unlikely that that they would be able to benefit from social cues that indicate the location of a relevant talker in a realistic cocktail party environment. That is, visual cues that could direct a listeners attention to a relevant talker during a real-life conversation, such as people directing their eye-gaze to a relevant talker or a gesture (e.g., a hand raise) from the relevant talker themselves, are arguably more subtle and fleeting than the type of visual cue presented for the current study.

On the other hand, it is possible that older listeners could be more sensitive to ecologically valid cues (e.g., gesture and eye-gaze) than the types of visual cues typically used for perception research (e.g., squares and arrows; Gayzur, Langley, Kelland, Wyman, Saville, Ciernia, 2014). Future research should investigate how different types of cues (i.e., endogenous and exogenous) that range in ecological validity affect younger and older adults' ability to gain a visual speech benefit in noise when there are visual distractors within the visual field. Furthermore, as familiarity with a talker's voice can help older adults recognise speech in noise (Johnsrude et al., 2013), how familiarity with a talker's auditory-visual speech (i.e., face and voice) affects older adults' ability to get a visual speech benefit when visual distraction is presented should also be investigated. That is, if a listener is more familiar with a talker's face and voice, it might be easier to avoid visual distraction. For these future studies, measuring eye-movement, in addition to speech recognition, could help to distinguish between the visual and/or attentional processes contributing to any potential differential effects of age.

Chapter 5

Does a visual distractor impair older adults' performance on an auditory-visual speech understanding in noise task?

5.1 Introduction

Older adults often report that understanding speech in situations with background noise (e.g., a busy restaurant) is challenging (CHABA, 1988; Pichora-Fuller, 2003). To assess an individual's ability to understand speech in noise, a clinician or researcher typically uses a speech recognition in noise task, where listeners are presented with words or short sentences and asked to recall what they hear by typing or speaking (e.g., The Hearing in Noise Test; Nilsson, Sigfrid, & Sullivan, 1996). Although speech recognition tasks provide an accurate and repeatable measure of speech intelligibility, they do not require participants to use listening skills that are necessary for participating in conversations, such as extracting meaning from speech and switching between multiple talkers (Best, Streeter, Roverud, Mason, & Kidd Jr., 2016).

In order to understand the day-to-day communication difficulties that people experience and to predict the real-world outcomes of potential interventions (e.g., hearing aids), researchers have started to develop speech perception tests that incorporate important aspects of listening in real life (Best, Keidser, Buchholz & Freeston, 2016). Although the specific designs vary across studies, these "real-life listening tests" generally aim to measure a person's ability to continuously comprehend what is said during a conversation between multiple people (i.e., a conversation-comprehension task; Best, Keidser, Buchholz, & Freeston, 2016; Best, Keidser, Freeston, & Buchholz, 2018; Best, Streeter, Roverud, Mason, & Kidd Jr., 2016). Significant advances have been made in incorporating these conversation-comprehension tasks into dynamic auditory scenes (Weisser, Buchholz, Oreinos et al., 2019); however, typically, the conversation-comprehension tasks that have been developed do not include any components of a realistic visual scene (e.g., visual speech that matches the auditory signal or visual distractors). Although seeing a talker's face generally facilitates speech perception in noise for both younger and older adults (Tye-Murray, Spehar, Myerson, Hale & Sommers, 2016), it is possible that susceptibility to distraction could reduce older adults ability to benefit from visual speech in a complex visual scene (Ch. 4 this thesis; Wascher, Schneider, Hoffman, Beste, & Sänger, 2012).Thus, the current study investigated how younger and older adults perform on a conversational speech understanding task (i.e., The Question-and-Answer Task) when visual speech that matches the auditory signal and visual distraction are presented.

Although components of realistic visual scenes have not been included in conversation-comprehension tasks, they have started to be included in traditional speech recognition tasks. Devesse, van Wieringen, and Wouters (2019), for example, developed the Audiovisual True-to-Life Assessment of Auditory Rehabilitation (AVATAR), an auditory-visual sentence recognition in noise task that includes a virtual restaurant scene with five virtual humans seated at a dining table. For an initial evaluation of the AVATAR, younger adults' performance on the speech recognition in noise task was measured when one of the virtual humans produced visual speech that matched the auditory signal. As (to the best of our knowledge) the other four virtual humans remained relatively still throughout the experiment, visual distraction (i.e., movement from other people or objects within the scene) was not included in the speech recognition task itself.

Devesse et al. (2019) did test whether speech recognition in noise performance was affected by having participants complete this task (i.e., the primary task) at the same time as a visual working memory task (i.e., a secondary task involving keeping track of numbers on a menu) that could have drawn attention away from the virtual human producing visual speech. They found that there was no significant difference between performance on the speech recognition in noise task for the dual-task (speech recognition and visual working memory) and single-task (i.e., just speech recognition) conditions. However, participants' performance on the visual working memory task was poorer for the dual-task condition in comparison to when the visual working memory task was completed by itself (i.e., a dual task cost). This pattern of results suggests that participants withdrew from the visual working memory task in order to sustain performance on the speech recognition in noise task (i.e., both tasks could not be completed simultaneously).

Although the restaurant scene presented by Devesse et al. (2019) has an attractive aesthetic and presents a common listening situation (i.e., a restaurant), the speech recognition task used in the initial evaluation study did not include visual distraction. Cohen & Gordon-Salant (2017), on the other hand, specifically tested how different types of visual distraction affect younger and older adults' performance on an auditory-visual speech recognition in noise task. They compared younger and older adults' performance on two conditions without visual distraction (i.e., an auditory only condition and a standard auditory-visual condition with one talking face that matched the auditory signal) to their performance in three conditions with different types of visual distraction (i.e., text, a talking face, or a video of a person performing an action).

Cohen and Gordon-Salant's (2017) results indicated that although speech recognition in noise performance was poorer for older adults than younger adults overall, the only condition with visual distraction that was significantly different from the standard auditory-visual condition for either age group was the video of a person performing a simple action (e.g., watering a plant). Presenting text or an additional talking face next to a relevant talking face, however, did not affect the visual speech benefit for younger or older adults. It is possible that the movements made by the talking face distractor (i.e., speech utterances) were not dynamic enough to be distracting in comparison to the more dynamic simple action. Another quality of the talking face distractor that could have made it easier to ignore in comparison to the action video is that the movement (i.e., speech utterances) always matched the content of the auditory noise. That is, this auditory-visual synchrony could have helped segregate the auditory signal from the auditory noise, and thus offset any effects of visual distraction (Driver, 1996).

Indeed, in the previous chapter (i.e., the cueing experiment), when a talking face that did not move in synchrony with auditory noise was presented as a visual distractor, older adults were unable to gain a full visual speech benefit, even when a salient visual cue indicating the location of the target talker was presented. It is also possible that the relevance of the visual distractor to the task affected older adults' ability to inhibit the visual distractor for both the cueing experiment and Cohen and Gordon-Salant's (2017) study. That is, the talking face distractor presented for the cueing experiment may have been more relevant to the task (and thus harder to ignore) than the talking face distractor presented by Cohen and Gordon-Salant (2017), as for the cueing experiment, the facial characteristics of target and distractor talkers were identical rather than distinct, and there was an equal chance of the valid cue and target appearing in one of two locations (left or right) rather than the target and distractor remaining in the same location for the entire experiment.

Although the cueing experiment and Cohen and Gordon-Salant (2017) incorporated basic components of a real-life visual scene into their visual display (i.e., visual speech that matches the auditory signal and visual distraction), both studies used a traditional speech recognition task. As previously stated, speech recognition tasks capture the word recognition component of speech perception but not the comprehension component of speech perception. Thus, the current study adapted an auditory-only, conversation-comprehension style task (i.e., The Question-and-Answer Task; Best, Streeter, Roverud, Mason, & Kidd Jr., 2016) and tested how visual speech and visual distraction affect younger and older adults' performance on this task.

The Question-and-Answer Task is inspired by the "Helen Test" which was originally developed to test the speech reading abilities of individuals with profound hearing loss (Best, Streeter, Roverud, Mason, & Kidd Jr., 2016; Ludvigsen, 1974). Each item of the Question-and-Answer Task consists of an unambiguous question (e.g., What colour is a lime?) followed by a one-word answer (e.g., Green). For this task, participants are charged with identifying whether the answer presented is true or false via a response button. To investigate the effects of visual speech and visual distraction on speech understanding in noise for younger and older adults, the current study pseudo-randomly presented listeners with Question-and-Answer items (mixed with speech shaped noise) in three visual display conditions: Static (i.e., a static image of faces was shown), Auditory-Visual (i.e., relevant talking faces were shown), and Auditory-Visual with Visual Distraction (i.e., relevant and irrelevant talking faces were shown). Both response time and accuracy were measured for each trial.

Although Best, Streeter, Roverud, Mason, and Kidd Jr. (2016) did not examine response time, it was included as a measure in the current study as it has been shown that response time can reveal effects independent of accuracy for speech in noise tasks (van den Tillaart-Haverkate, de Ronde-Brons, Dreschler, & Houben, 2017). For the response time measure, it was predicted that older adults would respond slower than younger adults overall and that both age groups would gain a visual speech benefit (i.e. response times would be faster for the Auditory-Visual Condition than the Static Condition). Furthermore, it was expected that the response time measure would be sensitive to effects of distraction. That is, if older adults are more distractible than younger adults, then the visual speech benefit should reduce (i.e., response times should increase) for older adults, but not for younger adults, when visual distraction is presented.

In order to minimise the effect of age-related hearing loss on task difficulty, older adults received a less adverse SNR (-8dB) than the younger adults (-10dB). As the results from a pilot study indicated that both age groups performed at approximately 80% correct for these respective SNRs when no visual cues were provided, it was predicted that there would not be a significant difference in accuracy scores between age groups. For the accuracy measure, it was predicted that younger and older adults would be less accurate for the Static Condition in comparison to the Auditory-Visual Condition and the Auditory-Visual with Visual Distraction Condition, and that the two auditory-visual conditions would not be significantly different from each other. That is, as the lexical complexity of the question-answer stimuli is low (which is likely due to the task's history with severely hearing-impaired subjects), visual distraction was not expected to affect younger or older adults' ability to accurately respond.

The current study also measured participants' visual acuity, hearing sensitivity, working memory capacity, and executive functioning. These additional tasks were administered to test for differences between age groups and to test for correlational relationships between these skills and performance on a speech understanding in noise task. As working memory capacity and hearing sensitivity have been related to performance on speech recognition in noise tasks (Dryden, Allen, Henshaw & Heinrich, 2017; Humes, 2013), it was predicted that these skills would also be related to performance on the Question-and-Answer Task.

5.2 Method

5.2.1 Participants

Twenty-five younger adults (17 Females, M_{Age} = 22) and 25 older adults (12 Females, M_{Age} = 72) participated in this study. Younger adults were students at Western Sydney University and participated for course credit (6 credits/hour) or monetary reimbursement ($20/hour). Older adults were recruited from the community and participated for monetary reimbursement ($20/hour). Session One took approximately one and a half hours and Session Two took approximately one hour.

5.2.2 Stimuli

An Auditory-Visual (AV), Australian- English, version of Best at al.'s (2016) Question-and-Answer Task was created. Each trial of the Question-and-Answer Task consists of a simple, unambiguous question, and a one-word answer. The questions cover six broad categories (i.e., Days, Months, Colours, Opposites, Sizes, and Numbers). See Table 5.1 for examples of questions and answers from each category.

Table 5.1
Description of the Six Question Categories from the Question-Answer Task (adapted from Best, Streeter, Roverud, Mason, & Kidd Jr., 2016)

Category	Number of Questions	Example Question	True Answer	False Answer
Days	14	What day comes after Tuesday?	Wednesday	Monday
Months	24	What month comes before June?	May	July
Colours	19	What colour is a lime?	Green	Silver
Opposites	18	What is the opposite of on?	Off	Closed
Sizes	21	Which is bigger, a moose or a bee?	Moose	Bee
Numbers	129	What is half of 10?	Five	Eight

A native Australian-English female talker was recorded uttering 225 questions and 113 answers in a sound attenuated booth. The talker was seated in front of a monitor that displayed each question and each answer one at a time. The video camera (Sony NCCAM HXR-NX30p) was situated directly above the monitor and captured video at 1920 x 1080 full HD resolution at 50 frames per second. The microphone (AT 4033a Transformerless

Capacitor Studio Microphone) was placed approximately 20 cm away from the talkers' mouth out of the cameras view and captured auditory speech at 48 kHz. All audio recordings were sent through a Motu Ultralite mk3 audio interface with FireWire connection to a PC running CueMix FX digital mixer and then to Audacity (Version 2.1.1).

Appendix A lists all of the question-answer pairs used for the current study. Incorrect answers were selected by the first author from other valid answer options (i.e., answers from the same category) in the corpus. One of the questions used in Best, Streeter, Roverud, Mason, & Kidd Jr.'s (2016) study was adapted for use in an Australian context (i.e., "What colour is a dime?" was changed to "What colour is a ten-cent coin?"). Two additional Australian questions and their respective answers were recorded and used (e.g., "Which is bigger, a kangaroo or a koala?").

5.2.2.1 Video Editing

The video recordings for each condition were edited using FFmpeg. All video recordings were scaled and cropped to measure 450px (height) x 340px (width). Video recordings were then further edited to produce six experimental conditions that counterbalanced two variables: Video type (Static vs. Auditory-Visual vs. Auditory-Visual with Visual Distraction) and Answer Type (Correct vs. Incorrect). See Figure 5.1 for a summary of the visual stimuli presented for each condition.

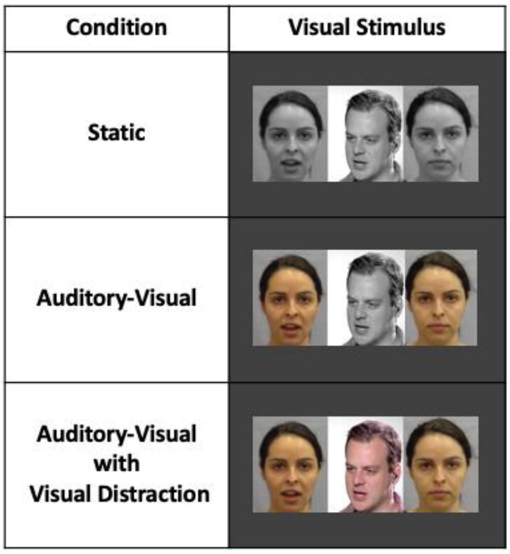

Figure 5.1. Visual Stimuli for Each Condition. Grey scale represents a static image whereas colour represents a visual speech video. Videos were presented as 12cm (height) by 21cm (width) with a visual angle of 121°. Trials from each condition were presented pseudo-randomly (i.e., conditions were not blocked). For trials from the Auditory-Visual and Auditory-Visual with Visual Distraction conditions, the face uttering the question appeared on the right side (and the face uttering the answer appeared on the left side) 50% of the time. The location of the faces uttering questions and answers (right vs. left) was also pseudo-randomised.

Each section of the videos shown in Figure 5.1 (i.e., Left, Middle, Right) were individually edited and then concatenated to create a single video file. For the Auditory-Visual with Visual Distraction Condition, the middle section of the video displayed a silent video of a male talker participating in a conversation. One out of eighteen possible distractor videos were randomly assigned to each question.

5.2.2.2 Auditory Editing

Two versions of the auditory recordings were created: one with a SNR of -8dB and one with a SNR of -10dB. Speech-shaped noise was created based on the long-term average spectrum of the original clear speech stimuli and then mixed with a copy of the clear stimuli at -8dB and -10dB respectively. Both versions were normalized to 70 dB SPL.

Questions from each version of the auditory recordings were then concatenated twice, once with the correct answer and once with the preselected incorrect (but valid) answer. A 0.5 second blank audio file was always included between the offset of each question and onset of each answer. The concatenated audio recordings were then mapped with the Auditory-Visual and Auditory-Visual with Visual Distraction videos to create auditory-visual stimuli. For each SNR (i.e., -8dB, -10dB, and no noise), six versions of the experiment were created so that each item could appear in all conditions without being repeated to a participant. Table 5.2 shows the time course of an auditory-visual trial.

Table 5.2.
Time Course of a Video from the Auditory Visual Condition

Segment	**Time Course**	**Video Portion**		
		Left	**Middle**	**Right**
Question Utterance	0s-2s	"What is two times seven?"	Static image of a male	Static image of the female talker
Pause	2s-2.5s	Static image of the female talker	Static image of a male	Static image of the female talker
Answer Utterance	2.5s-3.5s	Static image of the female talker	Static image of a male	"Fourteen"

Note. Other trials follow the same format (i.e., Question Utterance, Pause, Answer Utterance), however the time course varied depending on the content of the question and answer. The location (right vs. left) and accuracy (true vs. false) of the answers were evenly distributed across trials (and pseudo-randomly presented) for each version of the experiment.

5.2.3 Apparatus

Stimuli were presented using DMDX software (Forster & Forster, 2003) on a Dell T7810 computer with Windows 7 software. Stimuli were presented on a monitor measuring 30cm x by 53cm and through Sennheiser HD280pro headphones. The response button-box interfaced with the DMDX program via a parallel input/output card (Measurement Computing PCI-DIO24) to provide millisecond accurate response timing.

5.2.4 Procedure

After providing informed consent, participants completed a questionnaire that asked about their age, sex, and native language. Next, participants completed the Question-and-Answer Task with noise. In an attempt to equalise task difficulty between age groups, the experiment was presented to younger and older participants at -10dB and -8dB respectively.

Participants were seated in a sound attenuating booth approximately 70cm from the computer monitor. Participants were told that they would hear a question followed by a one-word answer, and that their task was to respond (as quickly and as accurately as possible) by indicating whether each answer was true or false on the button box provided. The researcher familiarized the participant with the button box; the left button was always labelled “FALSE” and the right button “TRUE”. Participants were also told that they would see a fixation cross and then a video for each trial. Participants were instructed to attend to each fixation cross and video, and to avoid closing their eyes during the experiment. Each participant completed a practice session that consisted of two items from the Static Condition presented with noise at -1dB, two items from the Auditory-Visual condition presented with noise at -8dB (older) and -10dB (younger), and two practice catch trials. Catch trials were identical to items from the Auditory-Visual Condition, however, a red border surrounded the exterior of the video. Participants were instructed to not press either button (i.e., True or False) when catch trials were presented.

After the practice session, participants completed 234 trials (216 test trials and 18 catch trials) presented in a pseudo-randomised order with an enforced break after 117 trials (i.e. halfway). For each trial, participants had ten seconds from the onset of each question to respond. The following trial always started after the ten seconds had passed, regardless of when the participant responded. Accuracy and response time (from the onset of the answer) were measured. For trials from the Auditory-Visual and Auditory-Visual with Visual Distraction conditions, the face uttering the question appeared on the right side (and the face uttering the answer appeared on the left side) 50% of the time. The location (right vs. left) and accuracy (true vs. false) of the answers were evenly distributed across trials (and pseudo-randomly presented) for both auditory-visual conditions, for each version of the experiment.

After completing the Question-and-Answer Task, participants completed a visual acuity test (FrACT) and pure-tone thresholds (Diagnostic Audiometer, AD229e) were measured at seven different frequencies (0.25,0.5, 1, 2, 4, 6, and 8 kHz). Finally, participants completed two cognitive tasks: The Listening Span (Conway et al., 2005) and the Trail Making Task (Reitan, 1992).

5.3. Results

5.3.1 Speech Understanding Task

The response time and accuracy data from the Question-and-Answer Task were evaluated to answer two questions: first, does presenting auditory-visual targets help younger and older adults' performance (i.e., response time and/or accuracy) on a conversational speech understanding in noise task, and second, does a visual distractor reduce any performance benefits (in response time and/or accuracy) that are gained when only target auditory-visual stimuli are presented?

5.3.1.1 Response Time

Participants' response times were measured from the answer onset of each item. Figure 5.2 shows the mean response times for trials that received a correct response as a function of Age, Answer Type, and Display Condition. As can be seen in the figure, younger and older adults responded approximately 200ms faster to trials presented in the Auditory-Visual Condition in comparison to the Static Condition for both answer types. This response time benefit seemed to persist when visual distraction was presented for both age groups.

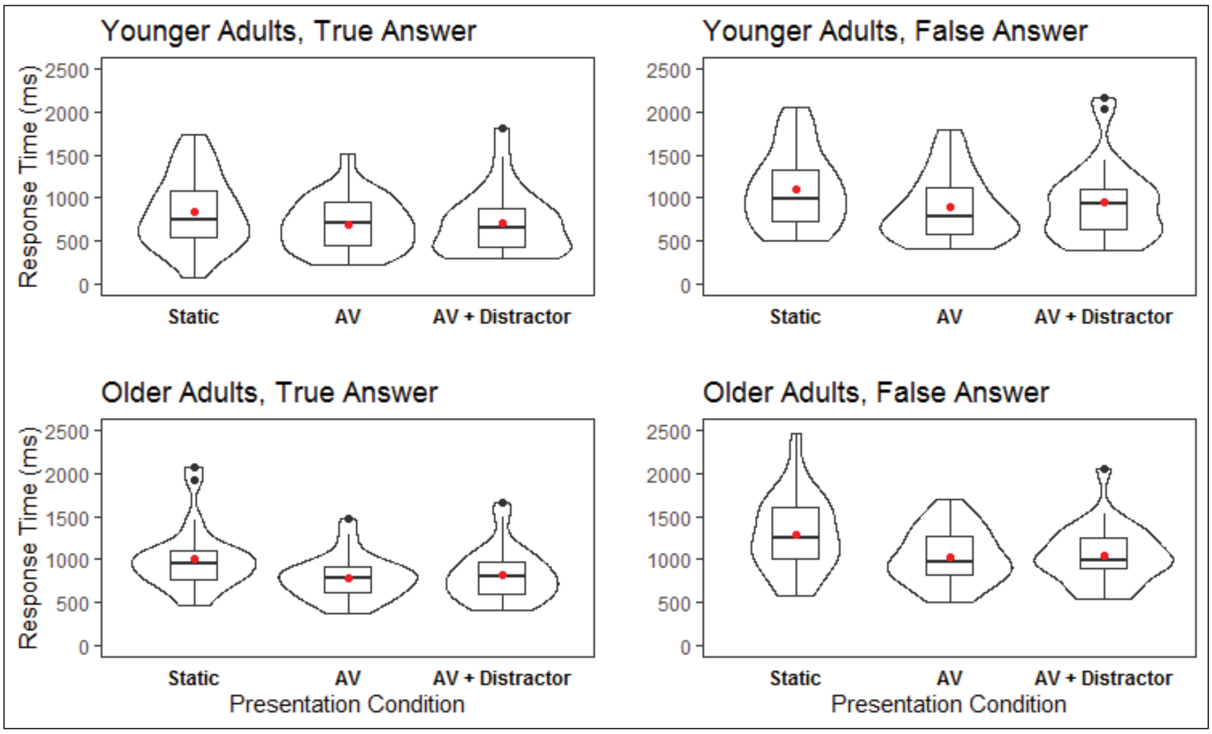

Figure 5.2. Mean Response Times. Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution. Circles filled in red represent the mean response time for each condition, Circles filled in grey represent values ≥ 1.5 IQRs.

To test whether younger and older adults' response times for a speech understanding in noise task were affected by the presence of visual speech and visual distraction, a repeated measures ANOVA with Age (Younger vs. Older) as the between participants factor and Display Condition (Static vs. Auditory-Visual vs. Auditory-Visual with Visual Distraction) and Answer Type (True vs. False) as the within participants factors was conducted. Although the difference in response times was in the expected direction, the main effect of Age was not significant. That is, older adults' response times (M = 864.18, SE = 71.14) were not significantly different from younger adults' response times (M = 999.32, SE = 71.14), $F(1, 48) = 1.81$, $p = 0.19$, $\eta_2 = 0.04$.

A significant main effect of Display Condition was found, $F(2, 96) = 55.53$, $p < .001$, $\eta_2 = 0.54$. Bonferroni corrected pairwise comparisons indicated that participants' response times for the Static Condition (M = 1055.71, SE = 57.15) were significantly slower than participants response times for the Auditory-Visual Condition (M = 856.40, SE = 45.80, p <

.001) and the Auditory-Visual with Distraction Condition (*M* = 883.14, *SE* = 51.46, $p < .001$), which were not significantly different from each other ($p = .370$). A significant main effect of Answer Type was also found. Participants' responded faster when the Answer Type was True (*M* = 810.28, *SE* = 46.91) in comparison to when the Answer Type was False (*M* = 1053.22, *SE* = 55.07), $F(1, 48) = 170.52$, $p < .001$, $\eta^2 = 0.78$. No significant interaction effects were found. This pattern of results did not change when any values that were greater than or equal to three interquartile ranges were replaced with mean values. In summary, the response time results suggest that both age groups responded faster to items that were presented with visual speech (in comparison to static faces) and to items that had a true answer type (in comparison to a false answer type), but that there was no effect of visual distraction on response time for either age group.

5.3.1.2 Accuracy

Figure 5.3 shows the mean percentage of correct responses for the Question-and-Answer Task as a function of Age, Display Condition, and Answer Type. As can be seen in Figure 5.3, the vast majority of both younger and older adults performed at above chance levels for all Display Conditions and Answer Types, however, both age groups were less accurate for the Static Condition in comparison to the Auditory-Visual Condition and Auditory-Visual with Visual Distraction Condition.

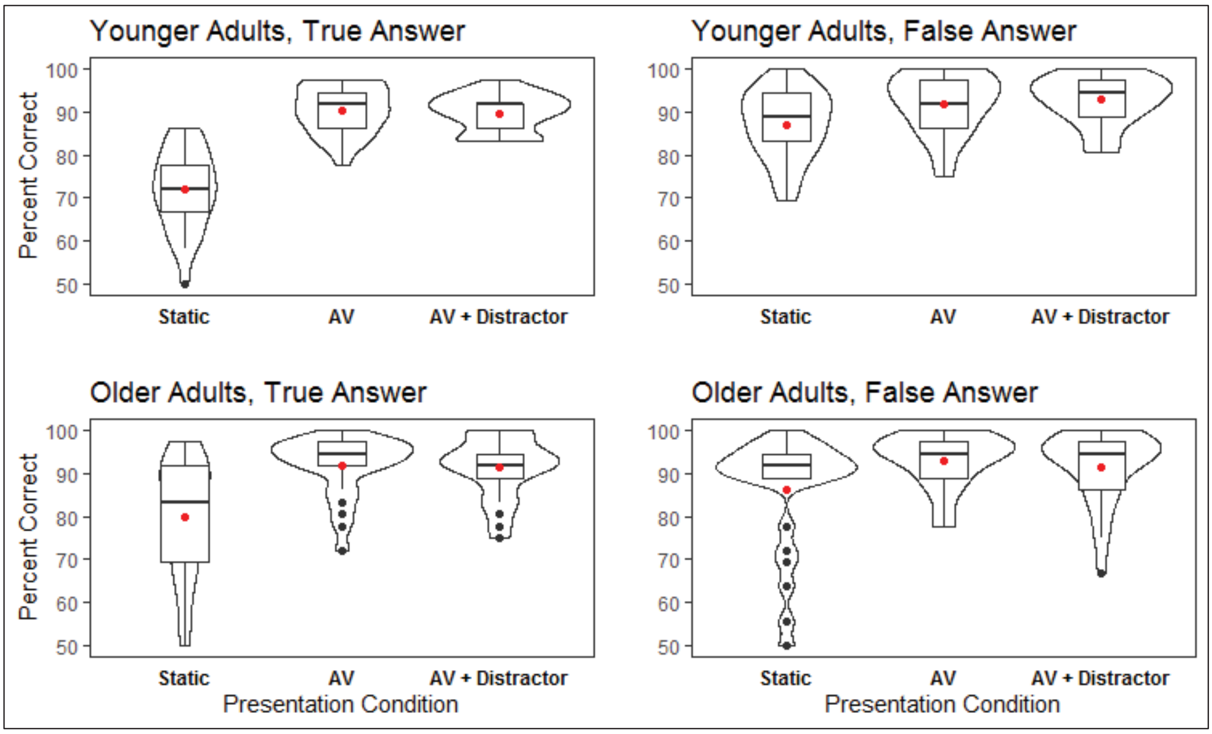

Figure 5.3. Mean Accuracy Scores. Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution. Circles filled in red represent the mean accuracy score for each condition, circles filled in grey represent values ≥ 1.5 IQRs. Note that the y axis scale starts at 50% correct.

5.3.1.2.1 Main Effects

To test whether younger and older adults' accuracy for a speech understanding in noise task was affected by the presence of visual speech and visual distraction, a repeated measures ANOVA with Age (Younger vs. Older) as the between participants factor and Display Condition (Static vs. Auditory-Visual vs. Auditory-Visual with Visual Distraction) and Answer Type (True vs. False) as the within participants factors was conducted. The main effect of Age was not significant. That is, the percentage of items that older adults responded to correctly (M = 88.89, SE = 1.18) was not significantly different from the percentage of items that younger adults responded to correctly (M = 87.26, SE = 1.18), $F(1, 48) = 0.95$, p = 0.33, $\eta_2 = 0.02$.

A significant main effect of Display Condition was found, $F(1.31, 63.09) = 75.85$, $p <$ .001, $\eta_2 = 0.61$(Greenhouse-Geisser correction; $\chi^2(2) = 34.66$, $p < .05$). Bonferroni corrected pairwise comparisons indicated that participants were significantly less accurate for the Static

Condition (*M* = 81.19, *SE* = 1.37) in comparison to the Auditory-Visual Condition (*M* = 91.72, *SE* = 0.74, $p < .001$) and the Auditory-Visual with Visual Distraction Condition (*M* = 91.31, *SE* = 0.78, $p < .001$), which were not significantly different from each other ($p = 1.00$). A significant main effect of Answer Type was also found. Participants' were less accurate when the Answer Type was True (*M* = 85.78, *SE* = 0.90) in comparison to when the Answer Type was False (*M* = 90.37, *SE* = 1.03), $F(1, 48) = 22.97$, $p < .001$, $\eta^2 = 0.32$.

5.3.1.2.2 Interaction Effects

5.3.1.2.2.1 Condition x Answer Type

There was a statistically significant Condition x Answer Type interaction, $F(1, 48) = 25.30$, $p < .001$, $\eta^2 = 0.35$. When the Answer Type was True, there was a main effect of Condition, $F(2, 98) = 77.71$, $p < .001$, $\eta^2 = 0.6$. That is, participants were significantly less accurate for the Static Condition (*M* = 75.83, *SE* = 1.74) in comparison to the Auditory-Visual Condition (*M* = 91.06, *SE* = 0.89, $p < .001$) and the Auditory-Visual with Visual Distraction Condition (*M* = 90.44, *SE* = 0.82, $p < .001$), with the difference between these latter two conditions not significantly different ($p = 1.00$). When the Answer Type was False, there was also a main effect of Condition, but the effect size was approximately 50% less than the effect size when the Answer was True, $F(2, 98) = 18.53$, $p < .001$, $\eta^2 = 0.27$. That is, when the Answer Type was False, participants were significantly less accurate for the Static Condition (*M* = 86.56, *SE* = 1.55) in comparison to the Auditory-Visual Condition (*M* = 92.39, *SE* = 0.93, $p < .001$) and the Auditory-Visual with Visual Distraction Condition (*M* = 92.17, *SE* = 1.00, $p < .001$), which were not significantly different from each other ($p = 1.00$).

5.3.1.2.2.2 Condition x Answer Type x Age Group

There was a statistically significant three-way interaction between Condition, Answer Type and Age Group, $F(2, 96) = 4.51$, $p = .01$, $\eta^2 = 0.09$. In summary, simple effects testing suggested that younger and older adults were significantly less accurate for the Static

Condition in comparison to the Auditory-Visual and Auditory-Visual with Visual Distraction conditions for both answer types. When the Answer Type was True, the size of this effect was approximately 50% greater for younger adults ($\eta^2 = 0.80$) than for older adults ($\eta^2 = 0.45$). When the Answer Type was False, the effect size for younger ($\eta^2 =0 .26$) and older ($\eta^2 = 0.30$) adults were small and not meaningfully different. Accuracy for the Auditory-Visual and the and Auditory-Visual with Visual Distraction conditions were not significantly different from each other for either age group or answer type. The descriptive statistics and statistical summaries from the simple effects analyses are summarised below.

5.3.1.2.2.2.1 Simple Effects

Bonferroni corrected simple effects analyses indicated that, for both Answer Types, younger adults were significantly less accurate for the Static Condition (True: $M = 71.88$, $SE = 1.85$; False: $M = 87.00$, $SE = 1.60$) in comparison to the Auditory-Visual Condition (True: $M = 90.44$, $SE = 1.11$; False: $M = 91.78$, $SE = 1.39$; $p < .05$) and the Auditory-Visual with Visual Distraction Condition (True: $M = 89.67$, $SE = 0.87$; False: $M = 92.78$, $SE = 1.13$; $p < .05$), which were not significantly different from each other ($p = 1.00$ for both answer types; True: $F(2, 48) = 84.75$, $p < .001$, $\eta^2 = 0.78$; False: $F(2, 48) = 8.61$, $p = .001$, $\eta^2 = 0.26$. Older adults were also significantly less accurate for the Static Condition (True: $M = 79.78$, $SE = 2.76$; False: $M = 86.11$, $SE = 2.69$) in comparison to the Auditory-Visual Condition (True: $M = 91.67$, $SE = 1.39$; False: $M = 93.00$, $SE = 1.25$; $p < .05$) and the Auditory-Visual with Distraction Condition (True: $M = 91.22$, $SE = 1.39$; False: $M = 91.55$, $SE = 1.67$ $p < .05$), which were not significantly different from each other ($p = 1.00$), for both answer types (True: $F(1.38, 33.17) = 19.67$, $p < .001$, $\eta^2 = 0.45$; False: $F(1.40, 33.48) = 10.44$, $p = .001$, $\eta^2 = 0.30$; Greenhouse-Geisser correction).

5.3.1.3 Effect of Outliers

As can been seen in Figure 5.4, there were several accuracy scores that were greater than the interquartile range of each condition. To test whether these outliers affected the pattern of results reported above, an additional repeated measures ANOVA was run with any accuracy scores that were greater than or equal to three interquartile ranges (i.e., outliers) replaced with the mean. The results from this additional repeated measures ANOVA suggested that when outliers were replaced with mean scores, the three-way interaction between Condition, Answer Type, and Age, was not significant ($F(2, 96) = .281$, $p = .756$, $\eta^2 = 0.01$). To identify what contributed to the non-significant three-way interaction for the second ANOVA, simple effects analyses were conducted using the data that replaced outliers with mean values. The descriptive statistics and statistical summaries from the simple effects analyses are summarised below and suggest that, when outliers were replaced with a mean value and the Answer Type was False, there was no longer a significant difference in accuracy between the Static Condition and the Auditory-Visual Condition or between the Static Condition the Auditory-Visual with Distraction Condition for older adults.

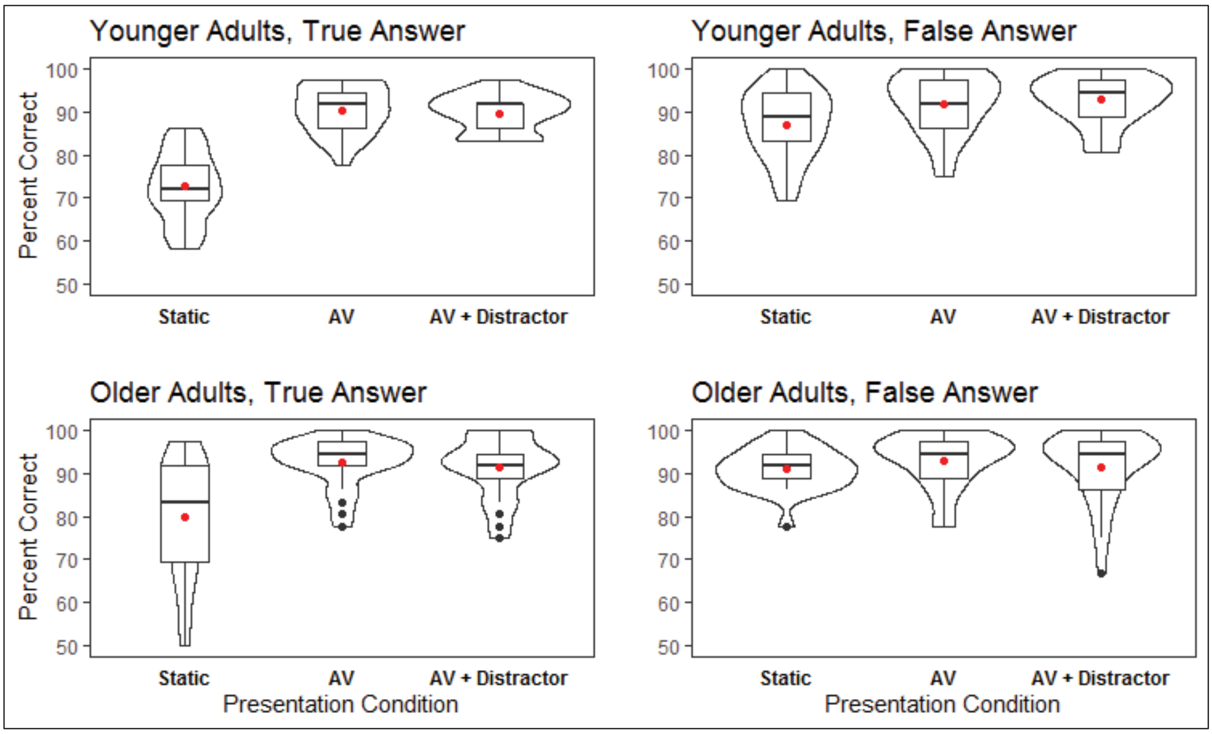

Figure 5.4. Mean Accuracy Scores when Outliers (i.e., values ≥ 3 IQRs) are Replaced with Mean Values). Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution. Circles filled in red represent the mean accuracy score for each condition, circles filled in grey represent values ≥ 1.5 IQRs.

5.3.1.3.1 Simple Effects

For both Answer Types, younger adults were significantly less accurate for the Static Condition (True: $M = 72.76$, $SE = 1.61$; False: $M = 87.00$, $SE = 1.61$) in comparison to the Auditory-Visual Condition (True: $M = 90.44$, $SE = 1.11$; False: $M = 91.78$, $SE = 1.39$, $p < .05$) and the Auditory-Visual with Distraction Condition (True: $M = 89.67$, $SE = .87$; False: $M = 92.77$, $SE = 1.13$; $p < .05$), which were not significantly different from each other ($p = 1.00$; True: $F(2, 48) = 86.59$, $p < .001$, $\eta_2 = 0.78$; False: $F(2, 48) = 8.61$, $p = .001$, $\eta_2 = 0.26$.

When the Answer Type was True, older adults were significantly less accurate for the Static Condition ($M = 79.78$, $SE = 2.76$) in comparison to the Auditory-Visual Condition ($M = 92.44$, $SE = 1.13$, $p < .001$) and the Auditory-Visual with Distraction Condition ($M = 91.22$, $SE = 1.39$, $p = .001$), which were not significantly different from each other ($p = 1.00$), $F(1.44, 34.50) = 19.09$, $p < .001$, $\eta_2 = 0.44$, (Greenhouse-Geisser correction: $\chi^2(2) = 11.42$, p

< .005). When the Answer Type was False, there was no significant difference in older adults' accuracy for the Static Condition (*M* = 90.89, *SE* = .97) in comparison to the Auditory-Visual Condition (*M* = 93.00, *SE* = 1.25, *p* = .317) or the Auditory-Visual with Distraction Condition (*M* = 91.55, *SE* = 1.67, *p* = 1.00), and the Auditory-Visual Condition and the Auditory-Visual with Distraction Condition were not significantly different from each other (*p* = .425), $F(1.55, 37.22) = 1.42$, $p = .251$, $\eta^2 = 0.06$, (Greenhouse-Geisser correction: $\chi^2(2) = 7.87$, $p < .005$).

5.3.1.4 Sensitivity (d') and Bias

As participants' responses for the Question-and-Answer Task were true/false judgements, it is possible to consider these data in terms of signal detection theory (Green & Swets, 1966). That is, a true response when the answer type is true is considered a "hit", a false response when the answer type is true is considered a "miss", a true response when the answer type is false is considered a "false alarm", and a false response when the answer type is false is considered a "correct rejection". With signal detection theory in mind, participants' sensitivity to the signal (i.e., d') and bias were calculated. Mean d' and bias scores for younger and older adults are shown in Figure 5.5. In summary, younger and older adults were less sensitive and more biased to respond false for the Static Condition in comparison to both auditory-visual conditions. Visual distraction did not affect younger or older adults' sensitivity or bias. The results from the repeated measures ANOVAs for d' and bias are presented below.

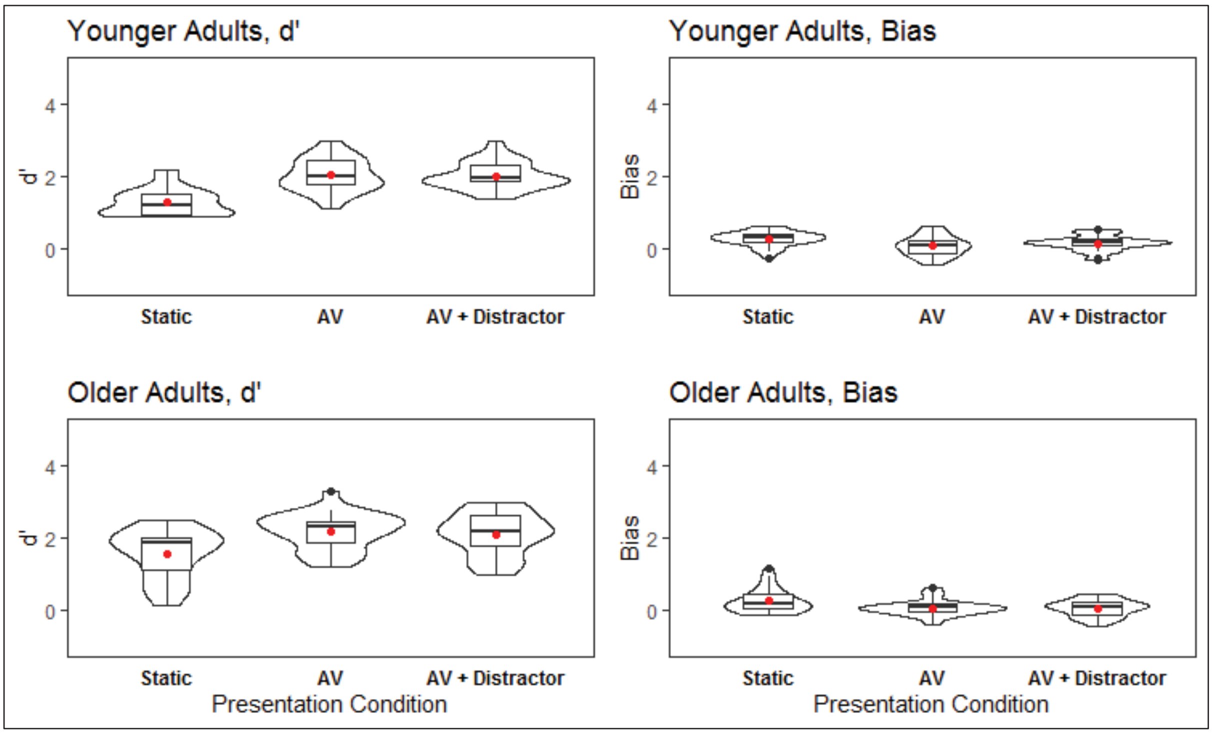

Figure 5.5. Mean d' and Bias Scores for Younger and Older Adults as a Function of Presentation Condition.

5.3.1.4.1 d'

As can be seen in Figure 5.5, younger and older adults had lower d' scores for the Static Condition in comparison to both auditory-visual conditions. Lower d' scores indicate poorer sensitivity. A repeated measures ANOVA with Display Condition as the within participants factor (Static vs. Auditory-Visual vs. Auditory-Visual with Distraction) and Age (Younger vs. Older) as the between participants factor indicated that d' was significantly lower for the Static Condition (M = 1.42, SE = 0.08, p = .000) in comparison to the Auditory-Visual Condition (M = 2.11, SE = 0.07, p = .000) and the Auditory-Visual with Distraction Condition (M = 2.07, SE = 0.07, p = .000), which were not significantly different from each other, p = 1.00, $F(1.31, 63.09) = 75.85$, $p < .001$, $\eta_2 = 0.53$. There was no significant difference in d' between younger adults (M = 1.79, SE = 0.08) and older adults (M = 1.95, SE = 0.08), $F(1, 48) = 2.19$, $p = 0.15$, $\eta_2 = 0.04$.

5.3.1.4.1 Bias

As can be seen in Figure 5.5, mean bias scores were slightly positive (i.e., above zero but less than one) for both age groups and for all display conditions. This suggests that participants were more likely to respond false than true for the Question-and-Answer Task. A repeated measures ANOVA indicated that participants' bias scores were significantly (yet only slightly, $\eta_2 = 0.18$) higher for the Static Condition ($M = .26$, $SE = 0.04$) in comparison to the Auditory-Visual Condition ($M = .06$ $SE = 0.04$, $p = .000$) and the Auditory-Visual with Distraction Condition ($M = 0.10$, $SE = 0.03$, $p = .000$), which were not significantly different from each other, $p = 1.00$, $F(2, 48) = 10.43$, $p < .001$, $\eta_2 = 0.18$. There was no significant difference in bias scores between younger adults ($M = 0.16$, $SE = 0.04$) and older adults ($M = 0.12$, $SE = 0.04$), $F(1, 48) = .66$, $p = 4.22$, $\eta_2 = 0$.

5.3.1.5 Learning Effects

Although the results suggest that the visual distractor did not seem to make a meaningful difference to Accuracy or Response Time performance for the Question-and-Answer Task, the amount of exposure to the visual distractor that participants had received over the duration of the experiment was not considered. It is possible that with repeated exposure to trials from the Auditory-Visual with Visual Distraction condition, younger and older adults learned to supress the visual distractor, limiting the distractor's effect on accuracy and/ or response time (i.e., a learning effect). To test whether performance on the Auditory-Visual with Visual Distraction Condition was susceptible to learning effects, younger and older adults' performance on each condition for the first half of the experiment was compared to their performance on the second half of the experiment.

5.3.1.5.1 Response Time

Figure 5.6 shows mean response times for younger and older adults as a function of Display Condition and Presentation Period. As can be seen in Figure 5.6, both age groups responded approximately 200ms slower to trials presented during the first half of the experiment in comparison to trials presented during the second half of the experiment for all display conditions.

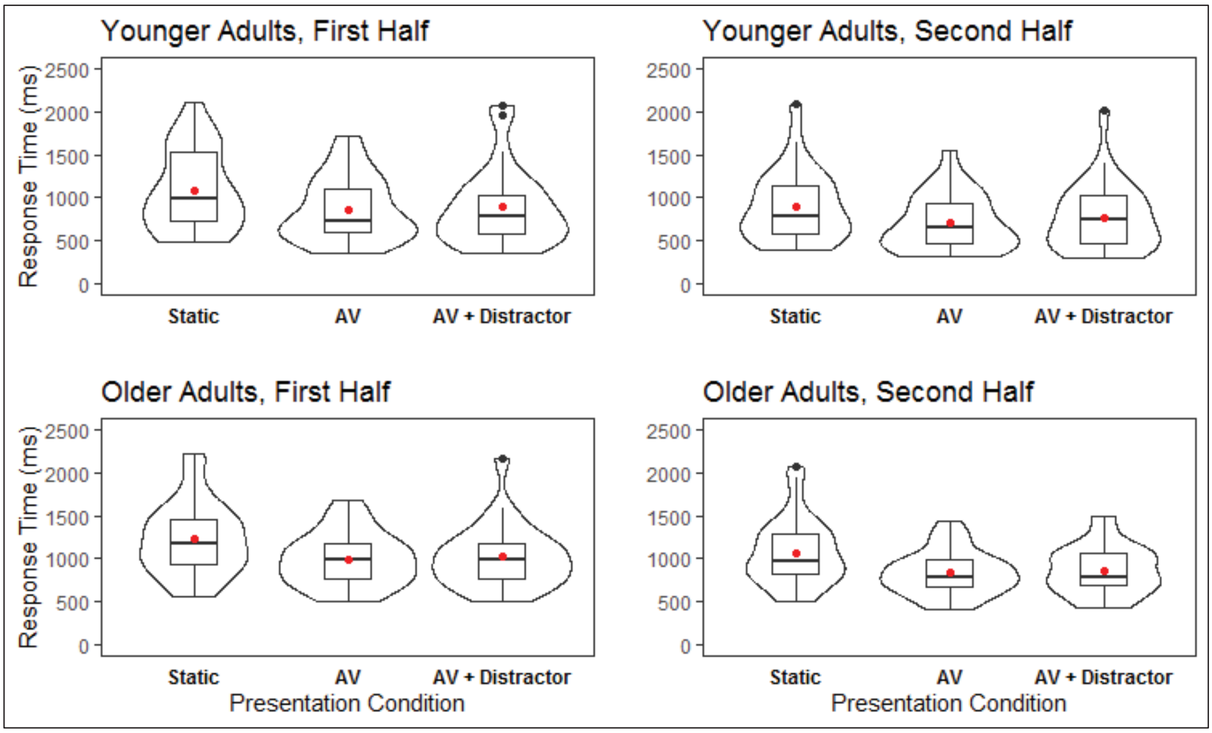

Figure 5.6. Mean Response Time for the First and Second Halves of the Experiment. Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution. Circles filled in red represent the mean accuracy score for each condition, circles filled in grey represent values ≥ 1.5 IQRs.

To test whether younger and older adults' response times for the Auditory-Visual with Visual Distraction Condition were affected by the amount of exposure to the task, a repeated measures ANOVA with Age (Younger vs. Older) as the between participants factor and Display Condition (Static vs. Auditory-Visual vs. Auditory-Visual with Visual Distraction) and Presentation Period (First Half vs. Second Half) as the within participants factors was conducted. There was a significant main effect of Presentation Period. Participants responded significantly slower to trials from the First Half of the experiment (M = 1013.52, SE = 54.58) in comparison to trials from the Second Half of the experiment (M = 859.00, SE = 47.39), $F(1, 48) = 45.49$, $p < .001$, $\eta_2 = 0.49$. There was also a significant main effect of Display Condition, $F(1.77, 96) = 64.53$, $p < .001$, $\eta_2 = 0.57$ (sphericity corrected with Greenhouse-Geisser, ($\chi^2(2) = 6.39$, $p < .005$)). Participants' response times for the Static Condition (*M* = 1071.11, *SE* = 57.47) were significantly slower than participants response times for the Auditory-Visual Condition (*M* = 850.40, *SE* = 43.84, $p < .001$) and the Auditory-Visual with Visual Distraction Condition (*M* = 887.28, *SE* = 51.49, $p < .001$), which were not significantly different from each other ($p = .096$). No significant interaction effects were found. Taken together, these results suggest that younger and older adults' response times for a conversational speech understanding task were not affected by the visual distractor used in this study (i.e., a talking face), regardless of the amount of exposure to the visual distractor.

5.3.1.5.2 Accuracy

Figure 5.7 shows the mean percentage of items correctly answered for younger and older adults as a function of Display Condition and Presentation Period. As can be seen in Figure 5.7, there was no meaningful difference between accuracy levels for the Auditory-Visual with Visual Distraction condition for either Presentation Period or Age Group. A repeated measures ANOVA with Age (Younger vs. Older) as the between participants factor and Display Condition (Static vs. Auditory-Visual vs. Auditory-Visual with Visual Distraction) and Presentation Period (First Half vs. Second Half) as the within participants factors suggested that there was a significant, yet very small ($\eta_2 = 0.08$), main effect of Presentation Period. Participants (i.e., younger and older adults) were approximately one percent more accurate for the First Half of the experiment (*M* = 88.67, *SE* = .87) in comparison to the Second Half of the experiment (*M* = 87.39, *SE* = .91), $F(1, 48) = 4.11$, $p < .05$, $\eta_2 = 0.08$. A significant main effect of Display Condition was also found, $F(1.32, 63.25) = 78.18$, $p < .001$, $\eta_2 = 0.62$ (sphericity corrected with Greenhouse-Geisser, ($\chi^2(2) = 34.27$, $p < .005$)). Participants were significantly less accurate for the Static Condition (*M* = 81.08, *SE* = 1.37) in comparison to the Auditory-Visual Condition (*M* = 91.76, *SE* = .72, $p < .001$) and the Auditory-Visual with Visual Distraction Condition (*M* = 91.23, *SE* = .78, $p < .001$), which were not significantly different from each other ($p = .917$).

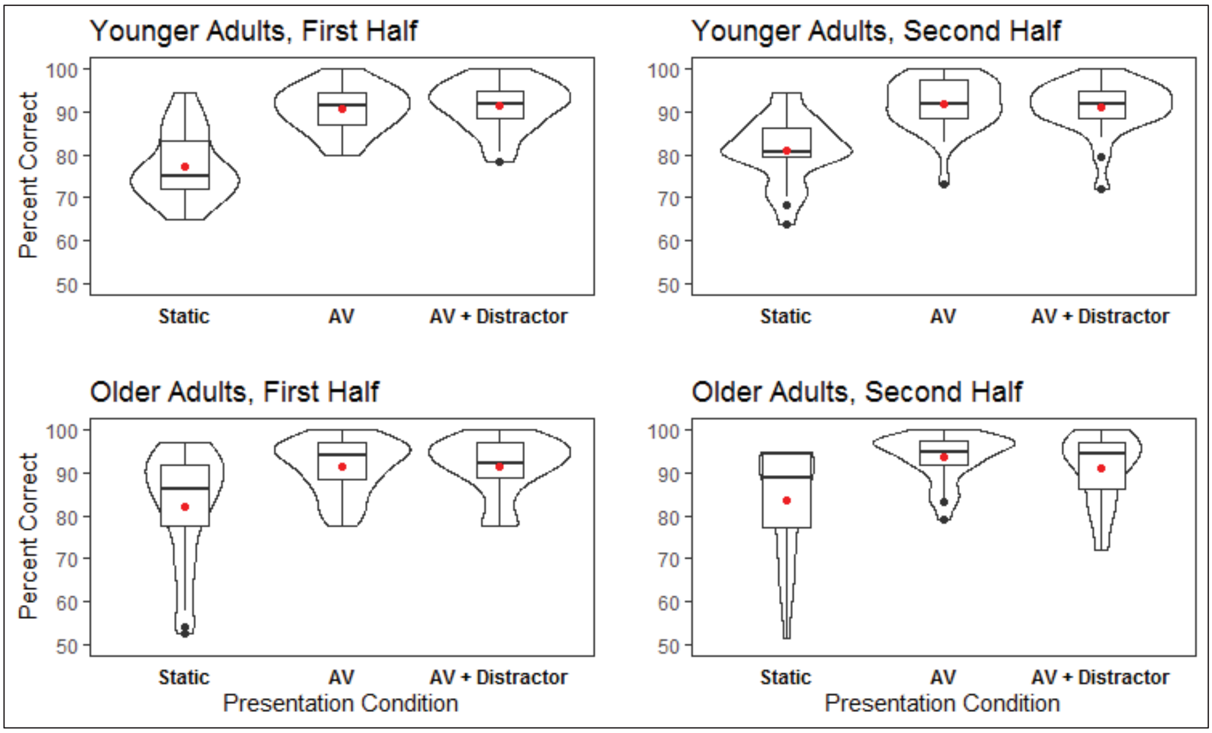

Figure 5.7. Mean Accuracy Scores for the First and Second Halves of the Experiment. Tukey's box plots represent the median and interquartile range (Q3-Q1). Violin plots represent the probability density of the data across the distribution. Circles filled in red represent the mean accuracy score for each condition, circles filled in grey represent values ≥ 1.5 IQRs.

5.3.2 Vision, Hearing, and Cognitive Tasks

Visual acuity, hearing sensitivity, executive functioning, and working memory capacity, were measured to evaluate any differences in these skills between age groups, and to evaluate whether these skills are related to performance on a speech understanding in noise task (i.e., the Question-and-Answer Task).

5.3.2.1 Visual Acuity

5.3.2.1.1 Younger Adults

All younger participants had normal or corrected to normal vision (i.e., ≥1.0 on the FrACT visual acuity measure; Bach, 2007). Younger adults' visual acuity scores ranged from 1.11 to the maximum score of 2.0 (M = 1.63, SD = .25).

5.3.2.1.2 Older Adults

Six older adults had worse than normal vision (i.e., < 1.0 on the FrACT visual acuity measure) with visual acuity scores ranging from 0.76 to the maximum score of 2.0 (M = 1.18, SD = .32). Pearson product-moment correlation coefficients were computed to test whether visual acuity was related to performance on the Question-and-Answer Task. The results indicated that older adults' visual acuity scores were not significantly associated with Response Time or Accuracy for the Static Condition (RT: r = -.04, p = .81; Accuracy: r = .12 , p = .39), Auditory-Visual Condition (RT: r = -.04 , p = .81; Accuracy: r = .11 , p = .45), or the Auditory-Visual with Visual Distraction Condition (RT: r = -.07 , p = .62; Accuracy: r = .10 , p = .47).

5.3.2.2 Hearing Sensitivity

Audiometric thresholds for each age group, as a function of ear and frequency, are summarised in Figure 5.8. All younger participants had normal hearing (i.e., ≤ 25dB HL at .25, .5, 1, 2, 4 kHz). Older adults' hearing levels were more diverse, ranging from normal to moderately-severe hearing loss (i.e., > 40dB and ≤ 70dB HL at .25, .5, 1, 2, or 4 kHz in the better ear), with the majority of older participants having mild hearing loss (12 participants; > 25 ≤ 40dB HL at .25, .5, 1, 2, or 4 kHz in the better ear) or normal hearing (6 participants). Younger adults had significantly lower thresholds than older adults for both ears at all tested frequencies (all p values ≤ .01).

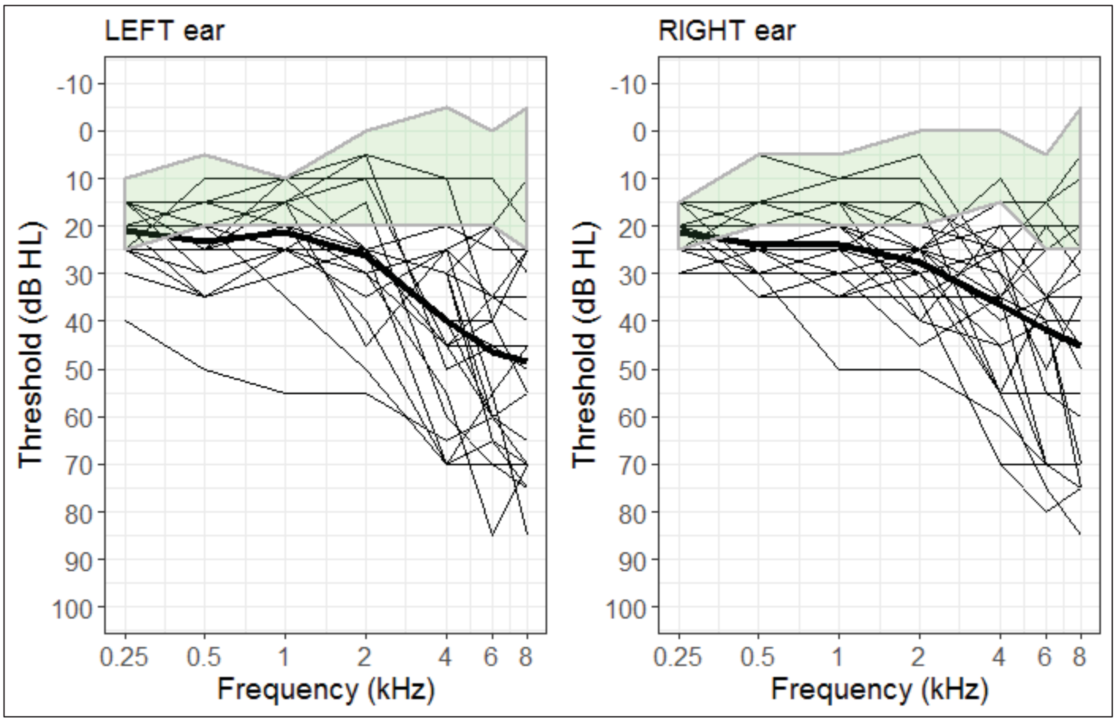

Figure 5.8. Audiogram Results for the Left and Right Ears. The bold black line represents the mean threshold for older adults as a function of frequency. The fine black lines represent individual audiograms for older adults as a function of frequency. The green shaded area represents the audiometric threshold range for younger adults.

Better Ear Average scores were calculated by averaging hearing thresholds across all tested frequencies for each ear and selecting the lower average threshold. The within group variation for the Better Ear Average was greater for older adults (*Min.* = 13.00, *Max.* = 39.00, *M* = 24.28, *SD* = 6.45) than younger adults (*Min.* = 7.00, *Max.* = 17.00, *M* = 11.24, *SD* = 2.34). When Question-and-Answer performance scores were averaged across Answer Types, younger adults' Better Ear Average scores were not significantly correlated with Response Time or Accuracy for the Static Condition (RT: *r* = .29 , *p* = .16; Accuracy: *r* = -.24, *p* = .25), Auditory-Visual Condition (RT: *r* = .18, *p* = .39; Accuracy: *r* = .05 , *p* = .83), or the Auditory-Visual with Visual Distraction Condition (RT: *r* = .20, *p* = .35; Accuracy: *r* = -.17 , *p* = .41).

Pearson correlations between older adults' performance on the Question-and-Answer Task (with performance averaged across Answer Types) and Better Ear Average scores are illustrated in Figure 5.9. There was a significant association between older adults' Better Ear

Average scores and both Response Time and Accuracy performance for certain conditions. That is, in comparison to older adults with lower Better Ear Average scores, older adults with higher Better Ear Average scores (i.e., poorer hearing sensitivity) were slower to respond for the Auditory-Visual with Visual Distraction Condition ($r = .40$, $p = .05$), and were less accurate for the Static Condition ($r = -.48$, $p = .01$) and the Auditory-Visual Condition ($r = -.57$, $p < .01$). There was not a significant relationship between older adults' Better Ear Average scores and Response Time for the Static Condition ($r = .37$, $p = .07$) or the Auditory-Visual Condition ($r = .31$, $p = .13$). Older adults' Better Ear Average scores were also not significantly associated with Accuracy for the Auditory-Visual with Visual Distraction Condition ($r = -.28$, $p = .17$).

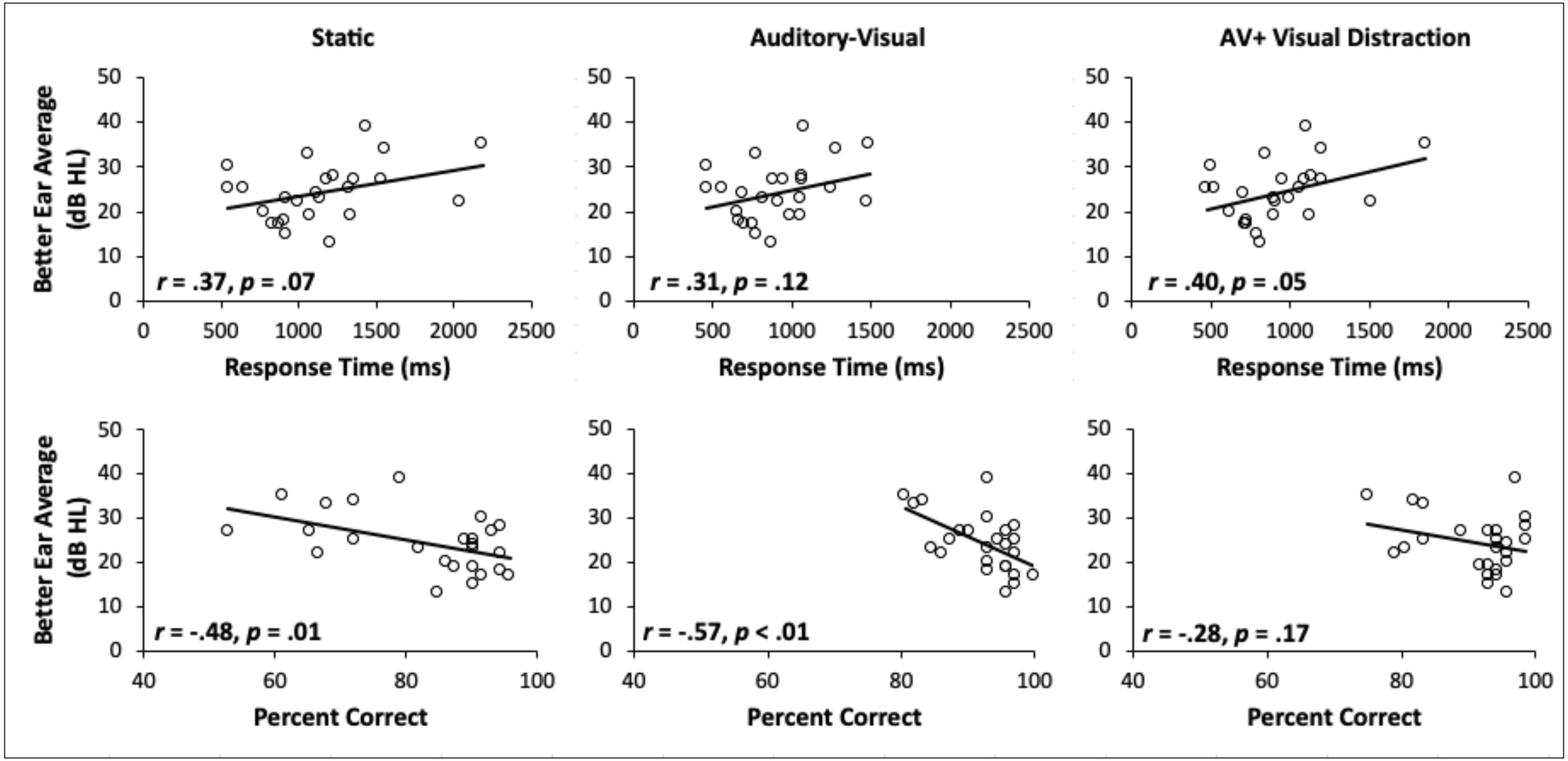

Figure 5.9. Pearson Correlations between Older Adults' Better Ear Average Scores and both Response Time (Top) and Accuracy (Bottom). Circles represent the average score from True and False Answer Types. Solid black lines represent the lines of best fit.

5.3.3 Cognition

Scores from parts A and B of the Trail Making Test were computed to assess age differences in executive control [(Part B-PartA)/PartA]. There was no significant difference between younger (M = 1.00, SD = .50) and older adults' (M = .90, SD = .55) computed scores ($t(98) = 0.98$, $p = 0.33$). When Question-and-Answer Task performance scores were averaged across Answer Type, Trail Making Test computed scores were not significantly related to younger or older adults' response times for the Static Condition (Younger: r = .06 , p = .80; Older: r = -.06, p = .79), Auditory-Visual Condition (Younger: r = -.01 , p = .97; Older: r = -.13, p = .56), or Auditory-Visual with Visual Distraction Condition (Younger: r = -.10, p = .66; Older: r = -.13, p = .55), or accuracy for the Static Condition (Younger: r = .14 , p = .50; Older: r = .13 , p = .55), Auditory-Visual Condition (Younger: r = -.07 , p = .74; Older: r = .10, p = .65), or Auditory-Visual with Visual Distraction Condition (Younger: r = -.05, p = .81; Older: r = .19, p = .38).

The listening span (i.e., LSPAN) was used to measure working memory capacity. Younger adults (M = 15.55, SD = 2.20) scored significantly higher on the LSPAN than older adults (M = 9.41, SD = 1.33; $t(98) = 6.91$, $p < 0.01$). Pearson correlation coefficients were calculated to test the relationship between LSPAN performance and performance on each condition of the Question-and-Answer Task when performance scores were averaged across Answer Types. However, no significant associations were found for either age group. That is, younger adults' LSPAN scores were not significantly correlated with Response Time or Accuracy for the Static Condition (RT-young, r = -.08 , p = .70 RT-old, r = -.38 , p = .06; Accuracy-young: r = .27, p = .20 Accuracy-old: r = .20, p = .33), the Auditory-Visual Condition (RT-young, r = .02 , p = .93 RT-old, r = -.31 , p = .14; Accuracy-young: r

= .30, *p* = .14 Accuracy-old: *r* = .18, *p* = .40), or the Auditory-Visual with Visual Distraction Condition (RT-young, *r* = -.06 , *p* = .78 RT-old, *r* = -.34 , *p* = .09; Accuracy-young: *r* = .34, *p* = .09 Accuracy-old: *r* = .11, *p* = .60).

5.4 Discussion

The current study had two primary aims. The first was to test, on a speech understanding in noise task, whether seeing visual speech that matches the auditory signal improves younger and older adults' performance (i.e., accuracy and response time) in comparison to an auditory-only condition. The second aim was to test whether this benefit would be reduced, for either performance measure, when a visual distractor was additionally presented. Consistent with investigations of the visual speech benefit that have used standard sentence recognition tasks, younger and older adults' speech understanding was more accurate when visual speech that matched the auditory signal was presented in comparison to when no visual speech was presented (i.e., both age groups gained a visual speech benefit; Cienkowski & Carney, 2002; Jesse & Janse, 2012; Middelweerd & Plomp, 1987, Sommers, Tye-Murray, Spehar, 2005; Tye-Murray, Spehar, Myerson, Hale & Sommers, 2016; Winneke & Phillips, 2011).

A novel finding from the current study was that this visual speech benefit was also observed in the response time measure. That is, both younger and older adults responded faster to question-and-answer task trials when matching visual speech was presented in comparison to the auditory-only condition. Response time has typically been used to objectively measure how different SNRs and hearing aid settings affect listening effort (i.e., the level of fatigue experienced by a listener due to the allocation of cognitive resources to a listening task; Gatehouse & Gordon, 1990; Houben, van Doorn-Bierman, & Dreschler, 2013; Meister, Rahlmann, Lemke, Besser, 2018; van den Tillaart-Haverkate, de Ronde-Brons, Dreschler, & Houben, 2017); however, to the best of our knowledge, the effect of visual speech on response times in a speech understanding in noise task has not been previously measured. Thus, this study is one of the first to show a new type of visual speech benefit in that seeing a talker's face can significantly reduce response time for speech understanding in noise for both younger and older adults.

Due to age-related declines in attentional control (Lustig, Hasher, & Zacks, 2007; Madden, Connelly & Pierce, 1994), it was expected that there would be a smaller visual speech benefit for older adults (for the response time measure) when the visual distractor was presented. However, the results indicated that both age groups were able to successfully ignore the visual distractor. That is, there was not a significant difference in performance (response time or accuracy) between the auditory-visual condition and the auditory-visual with distraction condition for either age group.

5.4.1 Why were older adults able to ignore the visual distractor?

Although it is possible that the abilities of the current older adults to inhibit the distractor were similar to those of the younger adults, this explanation seems unlikely, as older adults performed significantly worse than younger adults on the working memory capacity measure, and working memory capacity is thought to be indicative of an individuals' ability to inhibit distraction (Engle, 2010; McCabe et al., 2010). Alternatively, there are several properties of the visual distractor that could have helped younger and older adults to ignore the visual distractor and in turn facilitated participants' performance on the Auditory-Visual with Visual Distraction Condition.

First, the visual speech from the distractor talker was never a target (i.e., it never matched the auditory question or answer presented). That is, participants would have quickly realised that visual information from the distractor was never relevant to the task, which could have made the distractor easier to ignore. If as in the cueing experiment (Chapter 4), the speech of the distractor could have potentially been a target for some trials, then the visual distractor from the current study may have been more difficult to ignore. Second, the visual distractor used in the current study was very distinct from the target faces. That is, the distractor talker was male whereas the target talkers were female, the distractor video had a white background whereas the target videos had a grey background, and the distractor video always appeared in the same location on the screen (i.e., in between two target videos). If one or more visual properties of the visual distractor were consistent with the visual targets, then older adults' ability to ignore the distractor may have be impaired, particularly if (as discussed above) the speech of the distractor could have potentially been a target (Lien, Ruthruff, & Cornett, 2010). In summary, younger and older adults may have been able to ignore the visual distractor used for the current study as the distractor was clearly irrelevant to the speech understanding task.

5.4.2 Speech understanding and key concepts of cognitive hearing science

As significant associations have been found between standardised cognitive and auditory assessments and performance on traditional speech recognition in noise tasks (Dryden, Allen, Henshaw & Heinrich, 2017; Humes, 2013), a secondary aim of this study was to investigate whether these standard assessments would be significantly associated with performance on a conversation-comprehension style task (i.e., the Question-and-Answer Task). This analysis revealed two patterns of results that are relevant to cognitive hearing science.

5.4.2.1 Older Adults' Hearing Sensitivity and Accuracy

Older adults Better Ear Average (BEA) scores were moderately negatively related with older adults' accuracy scores for both the Static (r = -.48, p = .01) and Auditory-Visual (r = -.57, p < .01) conditions. These associations are consistent with the results from the cueing experiment (Chapter 4 this thesis) which showed moderate and strong negative correlations between speech recognition in noise performance and BEA for a static condition with no visual speech (r = -.41, p < .05) and a standard one talking face condition (r = -.63, p < .01). Although the significant association between the static conditions and hearing sensitivity is consistent with previous research (Humes, 2013; Schoof & Rosen, 2014), the negative relationship between BEA and the standard auditory-visual condition is inconsistent with other studies that have tested the relationship between hearing sensitivity and the ability to combine auditory and visual speech. Helfer (1998), for example, tested older adults with a range of hearing abilities and found that hearing sensitivity was not significantly related to the magnitude of the visual speech benefit. However, Tye-Murray, Sommers, and Spehar's (2007) results were in the opposite direction of the current study in that older adults with hearing loss gained a larger visual speech benefit than older adults with normal hearing for a sentence recognition in noise task. Tye-Murray, Sommers, and Spehar's (2007) results are consistent with Rosemann and Thiel's (2018) study which showed a stronger McGurk Effect for listeners with mild to moderate hearing loss than normal hearing listeners. Comparing auditory-visual speech understanding performance between groups of listeners with clearly defined hearing abilities could help to clarify the relationship between hearing sensitivity and speech understanding in noise.

Older adults BEA scores were not significantly related to accuracy performance on conditions with visual distraction for the current study or the cueing experiment, suggesting that peripheral hearing did not affect speech understanding in noise performance when visual distraction was within the visual field. Although the relationship between hearing sensitivity and auditory-visual speech perception has not been sufficiently tested, the results from the current study are consistent with those of Cohen and Gordon-Salant (2017), which showed that hearing sensitivity was not predictive of younger or older adults speech recognition in noise performance when visual speech and visual distractors were presented.

5.4.2.2 Working Memory Capacity and Question-and-Answer Task Performance

The results from the current study did not indicate any significant associations between LSPAN performance and Question-and-Answer Task performance (Response Time or Accuracy) for either age group. This pattern of results is different than studies showing that measures of working memory capacity are significantly related to performance on traditional sentence recognition tasks (Akeroyd, 2008; Dryden, Allen, Henshaw, & Heinrich, 2017) and performance on conversation-comprehension style tasks (Best, Keidser, Freeston, & Buchholz, 2018). Although the Question-and-Answer Task involves a comprehension component that was expected to place a demand on cognitive resources, it is possible that presenting linguistically simple questions and answers, with consistent onset times, was not a particularly cognitively demanding task. Indeed, performance on the Question-and-Answer Task may be more dependent on hearing the speech signal, as indicated by the significant associations with hearing sensitivity discussed above. The results from the current study may also differ from other studies investigating the relationship between speech perception in noise and working memory capacity as researchers have not agreed on the best method to measure working memory capacity. That is, inconsistent methods used between studies may also contribute to differing results (Dryden, Allen, Henshaw, & Heinrich, 2017; Wayne, Hamilton, Huyck, & Johnsrude, 2016).

5.4.3 Positive Response Bias and Signal Detection Theory

Consistent with Best, Streeter, Roverud, Mason, & Kidd Jr.'s (2016) initial evaluation study of the Question-and-Answer Task, the results from the current study indicate that participants were more biased to respond false. As discussed in Best et al. (2016), it is possible that this bias occurred due to participants adopting a response strategy of responding false if she/he did not hear the question or answer clearly. Our results partially support this interpretation, as when visual speech was presented (i.e., when it was easier for people to perceive the auditory signal) participants' responses were less biased than when visual speech was not presented. In contrast, even though the older adults tested had some age-related hearing loss (and the younger adults did not), older adults' bias was not significantly greater than younger adults' bias, and older adults' bias was not significantly related to the BEA. It is possible that the different SNRs presented to each age group (Younger: -10dB, Older -8dB) equalised the bias levels between age groups. It is also possible that a factor other than not perceiving the auditory signal clearly may have contributed to the positive response bias observed. For example, the knowledge that there were multiple possible false answers, but only one true answer, may have influenced participants' response bias (Best, Streeter, Roverud, Mason, & Kidd Jr., 2016).

5.4.4 Conclusion

The current study represents an initial attempt to incorporate basic visual features from real-life listening situations (i.e., visual speech from a talker and visual distraction) into a conversation-comprehension task. Using an auditory-visual version of the Question-and-Answer Task, this study showed that younger and older adults were able to gain a visual speech benefit in the form of improved accuracy and reduced response time, and that this benefit persisted when a visual distractor was presented. Although the visual scene for the current study was arguably more ecologically valid than previous studies that have used conversation-comprehension style tasks, the auditory scene for the current study lacks ecological validity as speech-shaped auditory noise was used rather than spatialised competing speech. A collaboration between researchers who specialise in realistic auditory scenes and researchers who specialise in visual speech could lead to the development of an auditory-visual, real-life listening test. This auditory-visual test could be useful for expanding our understanding of older adults' day-to-day communication difficulties and for predicting the real-world outcomes of hearing aids in realistic auditory-visual listening conditions.

Chapter 6

General Discussion

6.1 Thesis Overview and Aims

The experiments included in this thesis investigated the effects of noise on speech recognition and understanding in older and younger adults. In Chapter 2, I evaluated older and younger adults' performance on an auditory-only speech recognition in noise task, and a series of well-established auditory, cognitive, and lifestyle assessments that have been previously associated with speech recognition in noise ability. In Chapter 3, I presented a new auditory-visual speech recognition in noise task that, in addition to the standard auditory-only and auditory-visual (i.e., one talking face) conditions, included conditions where both a visual signal and at least one visual distractor (i.e., visual speech that did not match the auditory signal) were presented. In Chapter 4, I adapted the paradigm presented in Chapter 3 and tested whether presenting a salient visual cue indicating the location of a talking face that matched the auditory signal helped older and younger adults get a visual speech benefit when two talking faces (i.e., one matching and one not matching) were presented. Finally, in Chapter 5, I evaluated older and younger adults' performance (response time and accuracy) on an auditory-visual version of a speech understanding in noise task (i.e., the Question-and-Answer Task) when a visual distractor was (and was not) presented.

Collectively, the experimental program described above had four key aims:

1) To provide a broad characterisation of the samples tested for each experiment
2) To examine a factor that may affect the visual speech benefit

3) To determine whether presenting a visual cue can help listeners get a visual speech benefit when two talking faces (one that matches the auditory signal and one that does not) are within the visual field
4) To examine the effect of visual speech and visual distraction on a speech understanding in noise task (i.e., The Question-and-Answer Task)

The following section will review the results from each chapter in relation to the main aims outlined above. This will be followed by a discussion of the implications, limitations, and future directions of this research.

6.1.1 Aim 1: To provide a broad characterisation of the samples tested for each experiment.

A standard approach used in cognitive hearing science is to evaluate younger and older adults' performance on a variety of standardised cognitive, auditory and lifestyle assessments (i.e., to characterise the participant samples), and then test the relationship between performance of these assessments and performance on a speech recognition in noise task. This approach was used in Chapter 2 to confirm that the participants recruited were relatively healthy, and to evaluate which tasks showed a significant effect of age.

The results presented in Chapter 2 demonstrated that, in comparison to younger adults, older adults performed poorer on the speech recognition in noise task, and the majority of cognitive and auditory measures included in the test battery, although there was considerable variability in older adults' performance for most tasks. The results also indicated that older adults who reported higher levels of physical activity performed significantly better on the speech perception in noise task in comparison to older adults who reported lower levels of physical activity; however, out of all the auditory and cognitive factors tested, the only factor that was significantly related to speech perception in noise performance for older adults was the auditory attention quotient of the IVA+ Continuous Performance Task.

To follow up the possible importance of attention indicated by the results in Chapter 2, the experiments in Chapters 3, 4, and 5 took a more direct approach to investigating the role of attention in speech perception in noise. That is, the subsequent experiments in this thesis tested whether the ability to benefit from seeing a talker's face in a noisy situation (i.e., to get a visual speech benefit) was reduced when the demands on visual-spatial selective attention were increased (i.e., when visual distractors were presented within the visual field), and whether this reduction was greater for older adults (with limited attentional resources and control) than younger adults. These auditory-visual experiments were largely motivated by the idea that, as the visual speech benefit is the largest benefit available to listeners in noisy situations, any reduction in this benefit, particularly for older adults, could exacerbate difficulties understanding speech in noise.

6.1.2 Aim 2: To examine a factor that may affect the visual speech benefit (Chapter 3).

As the visual speech benefit is one of the largest benefits that a listener can receive when perceiving speech in a noisy environment, it would be valuable to know if there are any factors that can reduce the visual speech benefit, and if for any reason, older adults may be more vulnerable to this factor than younger adults. Although some studies suggest that auditory and visual information can be combined automatically (McGurk & MacDonald, 1976), one factor that has reduced other non-speech auditory-visual processing effects is when additional visual information is presented within the visual field (i.e., visual distractors; Stacey et al., 2014, Alsuis & Soto-Faraco, 2011, Fujisaki et al. 2006).

The experiments in Chapter 3 identified that one type of visual distractor (i.e., visual speech that does not match the auditory signal produced by the same talker as the target talker) can reduce the visual speech benefit, and that older adults are more vulnerable to the effects of this visual distractor than younger adults in certain listening conditions. That is, when the SNR was -6 dB (Experiment 1) the visual speech benefit reduced by approximately 50% when just one visual distractor was presented for both age groups. However, when the SNR was -1dB, younger adults were able to get a full visual speech benefit when one visual distractor was presented, whereas older adults' visual speech benefit reduced as in the previous experiment, by approximately 50%.

As previously discussed, a likely explanation for these results is that combining auditory and visual speech information requires attentional resources, and when these resources need to be devoted to auditory processing, combining auditory and visual information is done in a serial fashion by directing visual-spatial selective attention to only one talker's face. That is, when the SNR was less adverse, younger adults likely had sufficient attentional resources to combine auditory and visual information from two talking faces at once (i.e., the scope of visual-spatial selective attention encompassed both faces regardless of where the perceiver was foveating), whereas older adults, with age-related hearing loss and reduced attentional resource capacity, likely only had sufficient resources to attend to one talking face at a time (i.e., the scope of visual-spatial selective attention only encompassed the area of one face). This interpretation is consistent with how studies testing the effect of visual distraction on auditory visual-processing effects of have interpreted their results (e.g., Stacey et al., 2014), and with theories of cognitive ageing (Craik & Byrd, 1982; Hasher & Zacks 1988).

Visual speech was selected as the visual distractor for the experiments in Chapter 3, since in a noisy situation there are likely other people talking within a listener's visual field whose actions and conversations may be potentially important to a listener. As the same talker was displayed in the target video and the distractor video(s), it was impossible for listeners to know the location of the target talker (i.e., where to look) for each trial. In other words, the visual distractors presented in Chapter 3 were all relevant to the task, as each of them had potential to be the target talker. If there was no reduction in the visual speech benefit when one, three, or five visual distractors with facial characteristics identical to the target talker were presented, then this would have suggested that the visual speech benefit is resilient to the presence of highly relevant visual distractors, and would therefore likely also be resilient to less relevant visual distractors. However, as this was not the pattern of results observed, it should not be assumed that listeners are guaranteed to get a visual speech benefit in all listening conditions just because they can get a visual speech benefit when one talking face that matches the auditory signal is presented. That is, the magnitude of the visual speech benefit may be affected by certain types of visual distractors and/or the need to take them into account while listening. Older listeners, with reduced attentional resources and abilities, may be particularly susceptible to the demands of relevant visual distractors during auditory-visual speech perception.

6.1.3 Aim 3: To determine whether presenting a visual cue can help listeners get a visual speech benefit when two talking faces (one target and one distractor) are within the visual field.

As a substantial reduction in the visual speech benefit was observed for older adults when just one visual distractor was presented, the experiment in Chapter 4 evaluated whether providing a visual cue to the location of the target talker (i.e., knowing where to look/foveate) could help older adults overcome this reduction. In summary, a salient visual cue indicating the location of a target talker did not help older adults get a visual speech benefit when two talking faces (i.e., one target and one distractor) were presented. Younger adults, on the other hand, were able to benefit from the visual cue. Since the cue was very clearly presented, it was suggested that age-related declines in inhibition likely contributed to older adults' inability to gain a standard visual speech benefit for the valid cue condition. Thus, the results from Chapter 4 suggest that when visual speech that does not match the auditory signal is within the visual field, older adults are not guaranteed to get a standard visual speech benefit, even if they know the location of a target talker.

6.1.4 Aim 4: To examine the effects of visual speech and visual distraction on a speech understanding in noise task (i.e., The Question and Answer Task).

The final experiment in this thesis evaluated whether the distraction effect observed in the cueing experiment would also occur for a speech understanding task (i.e., The Question-and-Answer Task). The Question-and-Answer Task was selected as it incorporates several aspects of real-life listening that are not present in the highly controlled sentence recognition tasks typically used to measure the visual speech benefit. That is, for the Question-and-Answer Task, participants are required to switch between multiple talkers, understand what was said, and provide a timely response.

The results indicated that seeing visual speech that matched the auditory signal improved younger and older adults' performance (accuracy and response time) on the Question-and-Answer Task, but that the presence of a visual distractor (i.e., a talker producing visual speech that did not match either auditory signal) did not reduce performance on this type of task for either age group. In the discussion of Chapter 5, I suggested that several qualities of the visual distractor used for this experiment (e.g., its consistent location, and lack of potential relevance to the task) may have supported older and younger adults in inhibiting this distractor. Characteristics of the speech stimuli presented (i.e., context, predictability, low lexical complexity), and the task itself (i.e., understanding the gist and guessing) could have also facilitated older adults' performance.

6.2 Implications

The results from Chapter 3 have implications for theories of the role of attention in auditory-visual speech processing (Navarra et al., 2011). Early research on auditory-visual speech interactions mainly focused on the McGurk illusion (i.e., when the auditory token /ba/ is perceived as /da/ when watching a face uttering /ga/) and the apparent automaticity of this effect (McGurk & MacDonald, 1976). Studies showing that the McGurk effect is resilient to manipulations such as repeated exposure to McGurk stimuli (McGurk & MacDonald, 1976), desynchronised McGurk stimuli (Munhall, Gribble, Sacco, & Ward, 1996), and spatial separation of McGurk stimuli (Jones & Munhall, 1990), suggested that auditory-visual speech processing is automatic and therefore not limited by selective attention or attentional resource capacity (Soto-Faraco, Navarra, & Alsius, 2004). More recent studies (e.g., Alsius et al., 2005, 2007) suggest that the McGurk Effect is dependent on attention, as the effect reduces under conditions of high attentional load (i.e., when a secondary auditory, visual, or tactile task must be completed concurrently with a McGurk style speech perception task) and when a visual distractor is presented (Andersen et al., 2009; Tiippana, Anderson, & Sams, 2004).

Even though research suggests that claims made on the basis of the McGurk Effect should not necessarily be generalised to natural auditory-visual speech (Alsius, Paré, & Munhall, 2018; Van Engen & Chandrasekaran, 2017), an automatic, pre-attentive understanding of auditory-visual speech processing seems to have influenced the conclusions of researchers investigating ageing and the visual speech benefit in noise. That is, once it was demonstrated that older adults were able to get the same sized visual speech benefit as younger adults when one talking face that matched the auditory signal was presented (e.g., Cienkowski & Carney, 2002), the issue of whether the attentional demands of a listening situation could affect older or younger adults' ability to get this benefit was not rigorously investigated. However, as the experiments in Chapter 3 show that age and SNR affected the magnitude of the visual speech benefit when two talking faces (one target and one distractor) were presented, the results from this thesis are consistent with the view that that the ability to combine auditory and visual speech is modulated by the availability of top-down attentional resources, and furthermore, that the availability of these resources can be influenced by internal (i.e., participant resource capacity) and external (i.e., environmental) factors.

The results from this thesis also have implications for understanding how listeners benefit from seeing a talker's face in noisy situations. Although existing models (i.e., the ELU and the FUEL; Rönnberg et al. 2013, Pichora-Fuller et al., 2016) do not explicitly state how visual speech facilitates speech perception in noise, one suggestion is that by making the auditory signal clearer, visual speech reduces the amount of cognitive resources that need to be devoted to auditory processing, leaving resources available for higher order functions (i.e., encoding the speech signal in working memory and matching it with a representation in long term memory; Brown

& Strand, 2019; Frtusova & Phillips, 2016; Rudner, Mishra, Stenfelt. Lunner, & Ronnberg, 2016; Schneider & Pichroa-Fuller, 2000; Wingfield, Amichetti, & Lash, 2015). However, as the results from Chapter 3 suggest that combining auditory and visual speech requires at least some cognitive resources, and as other studies have suggested that auditory-visual speech processing is equally cognitively demanding or potentially more cognitively demanding than auditory-only speech processing (Brown & Strand, 2019; Fraser et al., 2010; Gosselin & Gagné, 2011; Keisder, Best, Freeston, & Boyce, 2015), the provision of visual speech may not necessarily reduce the demand for or "release" cognitive resources that would have otherwise been allocated to auditory processing.

An alternative explanation for how visual speech facilitates speech perception in noise is that auditory-visual speech is more strongly represented in working memory than auditory-only speech, making it easier to match the speech signal with a representation in long term memory (Brown & Strand, 2019). This explanation is grounded in dual-coding theory, which suggests that dual-modality items (i.e., auditory-visual speech) are encoded in working memory twice and are therefore easier to recall than a unimodal item (i.e., auditory-only speech) which is only encoded once (Thompson & Paivio, 1994; Mastroberardino, Santangelo, Botta, Marucci, & Olivetti Belardinelli, 2008). In the context of Baddeley's (2012) multi-component model of working memory, auditory-visual speech would be represented in both the visual-spatial sketchpad and the phonological loop rather than the phonological loop alone. Dual-coding could theoretically support speech perception in noise independently of the amount of cognitive resources that are allocated to speech processing; however, additional research needs to be conducted to confirm precisely how visual speech facilitates speech perception in noise and to identify internal and external factors that can disrupt this process.

6.3 Limitations and Future Directions

6.3.1 Chapter 2

Although a range of sample sizes (i.e., approximately 12-120; Dryden, Allen, Henshaw, & Heinrich, 2017) have been used to investigate the relationship between cognition and speech perception in noise, it is possible that the relatively small sample tested for the experiment in Chapter 2 (i.e., 30 participants per age group) could have limited the adequacy of the correlational and DABEST analyses. This was recognised as a potential issue apriori. Despite this potential limitation, the DABEST analysis suggested that older adults who had lower SRTs for the speech recognition task were more physically active than older adults who had higher SRTs. Future research should follow up this finding by testing the effects of physical activity on speech perception in noise more systematically (i.e., by comparing different age groups with different physical activity levels) and with a larger sample. Future research should also continue to investigate the reasons why physical activity might facilitate speech perception in noise for older adults.

As physical activity has been positively associated with hearing sensitivity, it is possible that physical activity supports the functioning of the peripheral auditory system, which in turn facilitates speech perception in noise (Alessio, Hutchinson, Proce, Reinart & Sautman, 2002; Gipsen, Chen, Genther & Lin 2015; Loprinzi, Cardinal & Gilham, 2011). Physical activity has also been associated with increased activation of the attentional network regions of the brain (i.e., the frontal and parietal regions), which are recruited during cognitively demanding tasks such as speech perception in noise (Colcombe et al., 2004; Hull & Kerschen, 2010; Wong et al., 2009). A common cause of the auditory and cognitive benefits of physical activity could be the maintenance a healthy circulatory system. That is, by preventing atrophy and inflammation of the blood vessels in the cochlea, inner ear, and brain, physical activity likely helps to ensure that structures within the auditory-cognitive system receive an adequate blood supply. The nutrients within blood (e.g., glucose) could help to support the integrity of these structures, and in turn a listener's ability to perceive speech in noise (Colcombe et al., 2004; Han et al., 2016; Hutchinson, Alessio & Baiduc, 2010; Pei, Chen & Zheng, 2016).

6.3.2 Chapters 3, 4, and 5

A limitation of the experiments presented in Chapters 3 and 4 is that it is not possible to precisely differentiate between the contributions of gaze-switching and visual-spatial attention to the results. Although the case for the involvement of visual-spatial attention was presented in the discussions of each chapter, future research should employ eye tracking techniques to confirm the underlying cause of the differences observed between age groups. Similarly, comparing performance between younger and older adults with matched hearing ability could rule out the possibility that hearing loss was driving the differential effects of age observed for conditions with two talking faces rather than visual and/or cognitive factors.

Future studies should also investigate older adults' ability to get a visual speech benefit from visual speech presented at different eccentricities from fixation. Currently, such work has only evaluated how younger adults' perception of single syllables is affected by visual speech in the periphery (Kim & Davis, 2013; Paré et al., 2003). Extending this research by using continuous speech (sentences) and testing both younger and older adults would help to establish whether age-related changes in the "useful field of view" limits older adults' ability to process visual speech information (Sekuler, Bennett, Mamelak, 2000). Older adults with Glaucoma, for example, may have particular difficulty processing visual speech presented in the periphery, as the hallmark symptom of this prevalent disease is an impairment in peripheral vision (Gagné & Wittich, 2009). Impaired peripheral vision could potentially prevent attentional capture from visual distractors; however, it could also reduce the magnitude of the visual speech benefit by limiting older adults' ability to process relevant visual speech.

In order to carefully control the auditory processing demands of each task while the visual stimuli were manipulated, the experiments presented in Chapters 3, 4, and 5 presented energetic, non-spatialised auditory noise. Future studies should build on this work by developing auditory-visual speech perception tests that incorporate realistic auditory scenes. Some studies (e.g., Devesse, van Wieringen, & Wouters, 2019; Hendrikse, Llorach, Hohnmann, Grimm, 2019) have successfully integrated spatialised auditory scenes with virtual reality visual displays; however, participants have reported that the visual realism of these displays is poor. Presenting real-life video recordings on a panoramic screen (e.g., Google Liquid Galaxy) could potentially address this limitation of virtual visual displays and help participants to feel as if they are immersed in an auditory-visual listening scenario (Watson, Parker, Leahy, Piepers & Stevens, 2018).

Additionally, as the auditory-visual experiments in this thesis only tested the effect of one type of visual distractor (i.e., visual speech), future research should test how the presence of different types of visual information affect older and younger adults' ability to gain a visual speech benefit. Cohen & Gordon-Salant (2017), for example, found that when a video of a person performing a simple action was presented as a visual distractor, the visual speech benefit significantly reduced for both older and younger adults, even when the location of the target talker did not change throughout the experiment. This suggests that movement in general (not just movement from visual speech) could interfere with a listener's ability to gain a visual speech benefit in certain contexts. Of course, the level and type of auditory noise could affect a listener's ability to inhibit different types of visual distraction and should therefore also be manipulated in future studies.

Finally, as discussed in Chapter 2, an interesting avenue for future research would be to test the effects of age and talker familiarity on the visual speech benefit in noise when different types of visual distractors are within the visual field. That is, exposure to a particular speaker's auditory-visual speech (i.e., face and voice) may help older listeners to direct and control visual-spatial attention to that speaker in a multi-talker environment.

6.4 Conclusion

The experiments presented in this thesis suggest that a research approach that encompasses both auditory and visual speech processing may be necessary to fully understand the difficulties that older adults experience during day-to-day listening situations. That is, although seeing a talker's face is assumed to provide a 10-15 dB benefit for both older and younger adults, the presence of just one visual distractor reduced the visual speech benefit for older adults (and not younger adults) by approximately 50%, even when the location of the target talker was visually cued. The results from this thesis also indicate that this reduction in the visual speech benefit was likely due to age-related changes in how visual-spatial selective attention is deployed; however, additional research needs to be conducted to confirm this relationship and to understand the factors that condition visual and auditory distraction within a communicative setting.

The absence of a distraction effect for the experiment in Chapter 5 suggests that there is a need for future studies to systematically manipulate the components of speech perception in noise tasks (i.e., the stimuli presented and the response format) in order to identify when and why visual distractors affect the visual speech benefit, particularly for older adults. As previously discussed, a listener's ability to ignore a visual distractor may be influenced by the relevance of a visual distractor to a listening task. Continuing to study the auditory and visual components of communication could help to facilitate the development of new strategies and assistive technologies that support social engagement across the lifespan.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*(Sup.2), S53– S71. https://doi.org/10.1080/14992020802301142

Alessio, H. M., Hutchinson, K. M., Price, A. L., Reinart, L., & Sautman, M. J. (2002). Study finds higher cardiovascular fitness associated with greater hearing acuity. *The Hearing Journal*, *55*(8), 32-40. doi: 10.1097/01.HJ.0000293292.87003.3d

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839-843. https://doi.org/10.1016/j.cub.2005.03.046

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399-404. doi: 10.1007/s00221-007-1110-1

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: the McGurk effect revisited. *Multisensory Research*, *31*(1-2), 111-144. https://doi.org/10.1163/22134808-00002565

Alsius, A., & Soto-Faraco, S. (2011). Searching for audiovisual correspondence in multiple speaker scenarios. *Experimental Brain Research*, *213*(2-3), 175-183. https://doi.org/10.1007/s00221-011-2624-0

Andersen, T. S., Tiippana, K., Laarni, J., Kojo, I., & Sams, M. (2009). The role of visual spatial attention in audiovisual speech perception. *Speech Communication*, *51*(2), 184-193. https://doi.org/10.1016/j.specom.2008.07.004

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, *300*, 18–32. doi:10.1016/j.heares.2013.03.006.

Amos, N. E., & Humes, L. E. (2007). Contribution of high frequencies to speech recognition in quiet and noise in listeners with varying degrees of high-frequency sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, *50*(4), 819-834. https://doi.org/10.1044/1092-4388(2007/057)

Arlinger, S., Lunner, T., Lyxell, B., & Kathleen Pichora-Fuller, M. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology, 50*(5), 371-384. https://doi.org/10.1111/j.1467-9450.2009.00753.x

Aubanel, V., Davis, C., & Kim, J. (2015). Explaining the visual and masked-visual advantage in speech perception in noise: the role of visual phonetic cues. In the proceedings of *AVSP 2015*, 132-136.

Aubanel, V., Davis, C., & Kim, J. (2017). The MAVA corpus. Retrieved from: https://app.alveo.edu.au/catalog/mava

Bach, M. (2007). The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245*(7), 965-971. doi: 10.1007/s00417-006-0474-4

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *125*(5), 3358-3372. https://doi.org/10.1121/1.3110132

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., & Pichora-Fuller, M. K. (2015). Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies. *Ear and Hearing*, *36*(1), 24-41. doi: 10.1097/AUD.0000000000000096

Best, V., Keidser, G., Buchholz, J. M., & Freeston, K. (2016). Development and preliminary evaluation of a new test of ongoing speech comprehension. *International Journal of Audiology*, *55*(1), 45-52. https://doi.org/10.3109/14992027.2015.1055835

Best, V., Keidser, G., Freeston, K., & Buchholz, J. M. (2018). Evaluation of the NAL Dynamic Conversations Test in older listeners with hearing loss. *International Journal of Audiology*, *57*(3), 221-229. https://doi.org/10.1080/14992027.2017.1365275

Best, V., Streeter, T., Roverud, E., Mason, C. R., & Kidd Jr, G. (2016). A flexible question-and-answer task for measuring speech understanding. *Trends in Hearing*, *20*, 1-8. https://doi.org/10.1177/2331216516678706

Bialystok, E., & Craik, F. I. (Eds.). (2006). *Lifespan cognition: Mechanisms of Change*. New York: Oxford University Press.

Bojko, A., Kramer, A. F., & Peterson, M. S. (2004). Age equivalence in switch costs for prosaccade and antisaccade tasks. *Psychology and Aging*, *19*(1), 226-234. https://doi.org/10.1037/0882-7974.19.1.226

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), P223-P233. https://doi.org/10.1093/geronb/60.5.P223

Borchert, K., (2018). The integrated Visual and Auditory Continuous Performance Task. Millisecond Software. Retrieved from: https://www.millisecond.com

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-652. http://dx.doi.org/10.1037/0033-295X.108.3.624

Bradley, E. The Bootstrap and Modern Statistics. Journal of the American Statistical Association. *95*, 1293-1296. doi: 10.2307/2669773

Brett, C. H., & Machado, L. (2017). Manual versus saccadic assessment of cognitive inhibition and switching in young and older adults. *Psychological Assessment*, *29*(11), 1420-1425. http://dx.doi.org/10.1037/pas0000453

Brodeur, D. A., & Enns, J. T. (1997). Covert visual orienting across the lifespan. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *51*(1), 2-35. http://dx.doi.org/10.1037/1196-1961.51.1.20

Brown, V. A., & Strand, J. F. (2019). About face: Seeing the talker improves spoken word recognition but increases listening effort. https://doi.org/10.31234/osf.io/m7b8q

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

Cherry, K. E., Walker, E. J., Brown, J. S., Volaufova, J., LaMotte, L. R., Welsh, D. A., ... & Frisard, M. I. (2013). Social engagement and health in younger, older, and oldest-old adults in the Louisiana healthy aging study. *Journal of Applied Gerontology*, *32*(1), 51-75. https://doi.org/10.1177/0733464811409034

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing*, *23*(5), 439-449. doi: 10.1097/01.AUD.0000034781.95122.15

Cohen, J. I., & Gordon-Salant, S. (2017). The effect of visual distraction on auditory-visual speech perception by younger and older listeners. *The Journal of the Acoustical Society of America*, *141*(5), 470-476. doi: 10.1121/1.4983399

Colcombe, S. J., Kramer, A. F., Erickson, K. I., Scalf, P., McAuley, E., Cohen, N. J., ... & Elavsky, S. (2004). Cardiovascular fitness, cortical plasticity, and aging. *Proceedings of the National Academy of Sciences*, *101*(9), 3316-3321. https://doi.org/10.1073/pnas.0400266101

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331-335. https://doi.org/10.3758/BF03196169

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786. doi:10.3758/BF03196772

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414-427. https://doi.org/10.1121/1.2804952

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 62-101). Cambridge, UK: Cambridge University Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42-100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F.I.M Craik & S. Trehub, *Aging and cognitive processes* (191-211). New York: Plenum.

Craik, F. I., & Salthouse, T. A. (Eds.). (2011). *The handbook of aging and cognition*. New York: Psychology Press.

Damos, D. L. (1991). Dual-task methodology: Some common problems. In D. L. Damos, *Multiple-task performance* (101-119). London, UK: Taylor and Francis.

Daneman, M., and Carpenter, P. A. (1980). Individual-differences in working memory and reading. *Journal of Verbal Learning and Behaviour*. *19*(4), 450–466. doi: 10.1016/S0022-5371(80)90312-6

Davis, C., & Kim, J. (2004). Audio-visual interactions with intact clearly audible speech. *The Quarterly Journal of Experimental Psychology Section A*, *57*(6), 1103-1121. https://doi.org/10.1080/02724980343000701

Degeest, S., Keppler, H., & Corthals, P. (2015). The effect of age on listening effort. *Journal of Speech, Language, and Hearing Research*, *58*(5), 1592-1600. https://doi.org/10.1044/2015_JSLHR-H-14-0288

Denes, P.B., Pinson, E.N. (1993). *The Speech Chain: The Physics and Biology of Spoken Language*. New York, NY: W.H. Freeman and Company.

Devesse, A., & Wouters, J. (2019). AVATAR Assesses Speech Understanding and Multitask Costs in Ecologically Relevant Listening Situations. *Ear and Hearing*. [Epub ahead of print] doi: 10.1097/AUD.0000000000000778

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, *381*(6577), 66. https://doi.org/10.1038/381066a0

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends in Hearing*, *21*, 1-21. https://doi.org/10.1177/2331216517744675

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, *76*(1), 87-96. https://doi.org/10.1121/1.391011

Durlach N., (2006). Auditory masking: need for improved conceptual structure. *Journal of the Acoustical Society of America, 120*(4), 1787-90. https://doi.org/10.1121/1.2335426

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, *74*(3), 739-743. https://doi.org/10.1121/1.391011

Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association*, *95*(452), 1293-1296. https://doi.org/10.1080/01621459.2000.10474333

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, *51*(S1), S17-S26. https://doi.org/10.1086/650572

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23. https://doi.org/10.1111/1467-8721.00160

Erel, H., & Levy, D. A. (2016). Orienting of visual attention in aging. *Neuroscience & Biobehavioral Reviews*, *69*, 357-380. https://doi.org/10.1016/j.neubiorev.2016.08.010

Fan, J., Gu, X., Guise, K. G., Liu, X., Fossella, J., Wang, H., & Posner, M. I. (2009). Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition*, *70*(2), 209-220. https://doi.org/10.1016/j.bandc.2009.02.002

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340-347. https://doi.org/10.1162/089892902317361886

Folk, C. L., & Hoyer, W. J. (1992). Aging and shifts of visual spatial attention. *Psychology and Aging*, *7*(3), 453-465. http://dx.doi.org/10.1037/0882-7974.7.3.453

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments, & Computers, *35*(1), 116-124.

Fraser, S., Gagné, J.-P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research*, *53*(1), 18–33. https://doi.org/10.1044/1092-4388(2009/08-0140)

Frtusova, J. B., & Phillips, N. A. (2016). The auditory-visual speech benefit on working memory in older adults with hearing impairment. *Frontiers in Psychology*, *7*, 490. https://doi.org/10.3389/fpsyg.2016.00490

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. Y. (2005). Visual search for a target changing in synchrony with an auditory signal. *Proceedings* *of the Royal Society B: Biological Sciences*, *273*(1588), 865-874. https://doi.org/10.1098/rspb.2005.3327

Füllgrabe, C. (2013). Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss. *American Journal of Audiology*. *22*(2), 313–315. https://doi.org/10.1044/1059-0889(2013/12-0070)

Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, *6*, 347. https://doi.org/10.3389/fnagi.2014.00347

Füllgrabe, C., & Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology*. 7,1268. https://doi.org/10.3389/fpsyg.2016.01268

Gagné, J. P., Besser, J., Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm. *Trends in Hearing*, *21*, 1-25. https://doi.org/10.1177/2331216516687287

Gagné, J. P., & Wittich, W. (2009). Visual impairment and audiovisual speech perception in older adults with acquired hearing loss. In *Hearing Care for Adults: The Challenge of Aging. Proceedings of the Second International Conference* (pp. 165-177).

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet*, *366*(9491), 1111-1120.

Gatehouse, S., & Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *British Journal of Audiology*, *24*(1), 63-68. https://doi.org/10.3109/03005369009077843

Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology, 43*, 85-99. https://doi.org/10.1080/14992020400050014

Gayzur, N. D., Langley, L. K., Kelland, C., Wyman, S. V., Saville, A. L., Ciernia, A. T., & Padmanabhan, G. (2014). Reflexive orienting in response to short-and long-duration gaze cues in young, young-old, and old-old adults. *Attention, Perception, & Psychophysics*, *76*(2), 407-419. doi: 10.3758/s13414-013-0554-6

Gilmour, H. (2012). Social participation and the health and well-being of Canadian seniors. *Health Reports*, *23*(4), 23-32.

Gispen, F. E., Chen, D. S., Genther, D. J., & Lin, F. R. (2014). Association between hearing impairment and lower levels of physical activity in older adults. *Journal of the American Geriatrics Society*, *62*(8), 1427-1433. https://doi.org/10.1111/jgs.12938

Goldstein, E. B., & Brockmole, J. (2016). *Sensation and perception*. Cengage Learning.

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, *344*, 109-124. https://doi.org/10.1016/j.heares.2016.11.004

Gordon, M. S., & Allen, S. (2009). Audiovisual speech in older and younger adults: integrating a distorted visual signal with speech in noise. *Experimental Aging Research*, *35*(2), 202-219. https://doi.org/10.1080/03610730902720398

Gordon, M. S., Daneman, M., & Schneider, B. A. (2009). Comprehension of speeded discourse by younger and older listeners. *Experimental Aging Research*, *35*(3), 277-296. https://doi.org/10.1080/03610730902769262

Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear and Hearing*, *37*(5), 593-602. https://doi.org/10.1097/AUD.0000000000000316

Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. H. (2011). Auditory temporal processing and aging: implications for speech understanding of older people. *Audiology Research*, *1*(1), e4. doi: 10.4081/audiores.2011.e4

Gosselin, P. A., & Gagné, J.-P. (2011a). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, *50*(11), 786–792.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197-1208. https://doi.org/10.1121/1.1288668

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. https://doi.org/10.1121/1.422788

Greenwood, P. M., & Parasuraman, R. (2004). The scaling of spatial attention in visual search and its modification in healthy aging. *Perception & Psychophysics*, *66*(1), 3-22. doi: 10.3758/bf03194857

Grose, J. H., and Mamo, S. K. (2010). Processing of temporal fine structure as a function of age. *Ear and Hearing*, *31*(6), 755–760. doi: 10.1097/AUD.0b013e3181e627e7

Guerreiro, M. J., Murphy, D. R., & Van Gerven, P. W. (2013). Making sense of age-related distractibility: The critical role of sensory modality. *Acta Psychologica*, *142*(2), 184-194. https://doi.org/10.1016/j.actpsy.2012.11.007

Hajek, A., Brettschneider, C., Mallon, T., Ernst, A., Mamone, S., Wiese, B., ... & Stein, J. (2017). The impact of social engagement on health-related quality of life and depressive symptoms in old age-evidence from a multicenter prospective cohort study in Germany. *Health and Quality of Life Outcomes*, *15*(1), 140. https://doi.org/10.1186/s12955-017-0715-8

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2001). Evaluation of a cognitive test battery in young and elderly normal-hearing and hearing-impaired persons. *Journal of the American Academy of Audiology*, *12*(7), 357-370.

Han, C., Ding, D., Lopez, M. C., Manohar, S., Zhang, Y., Kim, M. J., ... & Tanokura, M. (2016). Effects of long-term exercise on age-related hearing loss in mice. *Journal of Neuroscience*, *36*(44), 11308-11319. https://doi.org/10.1523/JNEUROSCI.2493-16.2016

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193-225. https://doi.org/10.1016/S0079-7421(08)60041-9

He, N. J., Dubno, J. R., & Mills, J. H. (1998). Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects. *The Journal of the Acoustical Society of America*, *103*(1), 553-565. https://doi.org/10.1121/1.421127

Heinrich, A., Gagne, J. P., Viljanen, A., Levy, D. A., Ben-David, B. M., & Schneider, B. A. (2016). Effective communication as a fundamental aspect of active aging and well-being: paying attention to the challenges older adults face in noisy environments. *Social Inquiry into Well-Being*, *2*(1), 51–69. doi:10.13165/SIIW-16-2-1-05

Heinrich, A., Henshaw, H., & Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Frontiers in Psychology*, *6*, 782. https://doi.org/10.3389/fpsyg.2015.00782

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*, 234-242. https://doi.org/10.1044/jslhr.4002.432

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing*, *29*(1), 87-98. doi: 10.1097/AUD.0b013e31815d638b

Helfer, K. S., & Freyman, R. L. (2014). Stimulus and listener factors affecting age-related changes in competing speech perception. *The Journal of the Acoustical Society of America*, *136*(2), 748–759. doi:10.1121/1.4887463.

Hendrikse, M. M., Llorach, G., Hohmann, V., & Grimm, G. (2019). Movement and gaze behavior in virtual audiovisual listening environments resembling everyday life. *Trends in Hearing*, *23*, 1-29. https://doi.org/10.1177/2331216519872362

Hopkins, K., and Moore, B. C. J. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, *130*(1), 334–349. https://doi.org/10.1121/1.3585848

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology*, *52*(11), 753-761. https://doi.org/10.3109/14992027.2013.832415

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., & Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: A grounded theory to establish content validity for a new patient-reported outcome measure. *Ear and Hearing*, *39*(5), 922-934. doi: 10.1097/AUD.0000000000000553

Hull, R. H., & Kerschen, S. R. (2010). The influence of cardiovascular health on peripheral and central auditory function in adults: a research review. *American Journal of Audiology*. *19*(1), 9-16. https://doi.org/10.1044/1059-0889(2010/08-0040)

Humes, L. E. (2013). Understanding the speech-understanding problems of older adults. *American Journal of Audiology, 22*(2), 303–305. doi: 10.1044/1059-0889(2013/12-0066)

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., & Lee, L. (1994). Factors associated with individual differences in clinical measures of speech recognition among the elderly. *Journal of Speech, Language, and Hearing Research*, *37*(2), 465-474. https://doi.org/10.1044/jshr.3702.565

Hutchinson, K. M., Alessio, H., & Baiduc, R. R. (2010). Association between cardiovascular health and hearing function: Pure-tone and distortion product otoacoustic emission measures. *American Journal of Audiology*. *19*(1), 26-35. https://doi.org/10.1044/1059-0889(2009/09-0009)

Inquisit 5 [Computer Software]. (2016) Retrieved from https://www.millisecond.com.

James, W. (1890). The Principles of Psychology. Chapter XI. Retrieved from: http://psychclassics.yorku.ca/James/Principles/prin11.htm

James, B. D., Wilson, R. S., Barnes, L. L., & Bennett, D. A. (2011). Late-life social activity and cognitive decline in old age. *Journal of the International Neuropsychological Society*, *17*(6), 998-1005. https://doi.org/10.1017/S1355617711000531

JASP Team (2019). JASP (Version 0.10.2) [Computer software]. (BibTex)

Jerger, J. (1992). Can age-related decline in speech understanding be explained by peripheral hearing loss. *Journal of the American Academy of Audiology*, *3*(1), 33-38.

Jesse, A., & Janse, E. (2012). Audiovisual benefit for recognition of speech presented with single-talker noise in older listeners. *Language and Cognitive Processes*, *27*(7-8), 1167-1191. https://doi.org/10.1080/01690965.2011.620335

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*(10), 1995-2004. https://doi.org/10.1177/0956797613482467

Jones, J. A., & Munhall, K. G. (1997). Effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics*, *25*(4), 13-19. Retrieved from: https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1106

Jongman, A., Wang, Y., & Kim, B. H. (2003). Contributions of semantic and facial information to perception of nonsibilant fricatives. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1367-1377. https://doi.org/10.1044/1092-4388(2003/106)

Joses Ho, Tayfun Tumkaya, Sameer Aryal, Hyungwon Choi, Adam Claridge-Chang (2019) Moving beyond P values: Everyday data analysis with estimation plots. *Nature Methods*, 1548-7105. doi: 10.1038/s41592-019-0470-3

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Keidser, G., Best, V., Freeston, K., & Boyce, A. (2015). Cognitive spare capacity: evaluation data and its association with comprehension of dynamic conversations. *Frontiers in Psychology*, *6*, 597.

Keidser, G., Dillon, H., Mejia, J., & Nguyen, C. V. (2013). An algorithm that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function. *International Journal of Audiology*, *52*(11), 795-800. https://doi.org/10.3109/14992027.2013.817688

Kidd Jr., G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In Yost W.A., Popper A.N., Fay R.R. (eds) *Auditory Perception of Sound Sources*. Springer Handbook of Auditory Research, vol 29. Boston, MA: Springer.

Kiessling, J., Pichora-Fuller, M.K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., … & Rosenhall, U. (2003). Candidature for delivery and of audiological services: special needs of older people. International Journal of Audiology, *42*(sup2), 92-101.

Kim, J., & Davis, C. (2013). How far out? The effect of peripheral visual speech on speech perception. In Auditory-Visual Speech Processing (AVSP) 2013.

Kim, J., & Davis, C. (2014). How visual timing and form information affect speech and non-speech processing. *Brain and Language*, *137*, 86-90. https://doi.org/10.1016/j.bandl.2014.07.012

King, A., Hopkins, K., an48d Plack, C. J. (2014). The effects of age and hearing loss on interaural phase difference discrimination. *The Journal of the Acoustical Society of America*, *135*(1), 342–351. https://doi.org/10.1121/1.4838995

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557-583. http://dx.doi.org/10.1037/bul0000144

Krueger, K. R., Wilson, R. S., Kamenetsky, J. M., Barnes, L. L., Bienias, J. L., & Bennett, D. A. (2009). Social engagement and cognitive function in old age. *Experimental Aging Research*, *35*(1), 45-60. https://doi.org/10.1080/03610730802545028

Langley, L. K., Friesen, C. K., Saville, A. L., & Ciernia, A. T. (2011). Timing of reflexive visuospatial orienting in young, young-old, and old-old adults. *Attention, Perception, & Psychophysics*, *73*(5), 1546-1561. doi: 10.3758/s13414-011-0108-8

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467-477. doi: 10.1121/1.1912375

Lien, M. C., Ruthruff, E., & Cornett, L. (2010). Attentional capture by singletons is contingent on top-down control settings: Evidence from electrophysiological measures. *Visual Cognition*, *18*(5), 682-727. https://doi.org/10.1080/13506280903000040

Lincourt, A. E., Folk, C. L., & Hoyer, W. J. (1997). Effects of aging on voluntary and involuntary shifts of attention. *Aging, Neuropsychology, and Cognition*, *4*(4), 290-303. https://doi.org/10.1080/13825589708256654

Loprinzi, P. D., Cardinal, B. J., & Gilham, B. (2012). Association between cardiorespiratory fitness and hearing sensitivity. *American Journal of*

Audiology. *21*(1), 33-40. https://doi.org/10.1044/1059-0889(2011/11-0024)

Ludvigsen, C. (1974). Construction and evaluation of an audio-visual test (The Helen test). Scandinavian Audiology Supplementum, *4*, 67–75.

Luo, Y., Waite, L.J., 2005. The impact of childhood and adult SES on physical, mental, and cognitive well-being in later life. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*(2), S93-S101. https://doi.org/10.1093/geronb/60.2.S93

Lustig C, Hasher L, Zacks RT. (2007). Inhibitory deficit theory: Recent developments in a "new view". In Gorfein DS, MacLeod CM (Eds), *Inhibition and Cognition* (pp. 145–162). Washington, DC: American Psychological Association.

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, *21*(2), 131-141. https://doi.org/10.3109/03005368709077786

Madden, D. J., Connelly, S. L., & Pierce, T. W. (1994). Adult age differences in shifting focused attention. *Psychology and Aging*, *9*(4), 528-538. http://dx.doi.org/10.1037/0882-7974.9.4.528

Madden, D. J., & Monge, Z. A. (2019). Visual attention with cognitive aging. In *Oxford Research Encyclopedia of Psychology*. New York: Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.369

Madsen, S. M., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports, 9*. doi: 10.1038/s41598-019-46728-1

Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., & Olivetti Belardinelli, M. (2008). How the bimodal format of presentation affects working memory:

An overview. *Cognitive Processing*, *9*(1), 69–76. https://doi.org/10.1007/s10339-007-0195-6

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222-243. doi: 10.1037/a0017619

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal of Audiology*, *53*(7), 433-440. https://doi.org/10.3109/14992027.2014.890296

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

Meister, H., Rählmann, S., Lemke, U., & Besser, J. (2018). Verbal Response Times as a Potential Indicator of Cognitive Load During Conventional Speech Audiometry With Matrix Sentences. *Trends in Hearing*, *22*,1-11. https://doi.org/10.1177/2331216518793255

Mendes de Leon CF, Glass TA, Berkman LF (2003) Social engagement and disability in a community population of older adults: the New Haven EPESE. *American Journal of Epidemiology*, *157*(7), 633–642. https://doi.org/10.1093/aje/kwg028

Middelweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *The Journal of the Acoustical Society of America*, *82*(6), 2145-2147. doi: 10.1121/1.395659

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., & Lyxell, B. (2017). Objective Assessment of Listening Effort: Coregistration of Pupillometry and EEG. *Trends in Hearing*. https://doi.org/10.1177/2331216517706396

Millman, R. E., & Mattys, S. L. (2017). Auditory verbal working memory as a predictor of speech perception in modulated maskers in listeners with normal hearing. *Journal of Speech, Language, and Hearing Research*, *60*(5), 1236-1245. https://doi.org/10.1044/2017_JSLHR-S-16-0105

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, *21*(11), 1078-1085. https://doi.org/10.1002/gps.1610

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology. 9*(4), 399–406. doi: 10.1007/s10162- 008-0143-x

Moore, B. C. J., Glasberg, B. R., Stoev, M., Füllgrabe, C., and Hopkins, K. (2012). The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies. *The Journal of the Acoustical Society of America*, *131*(2), 1003–1006. https://doi.org/10.1121/1.3672808

Moore, B. C. J., Vickers, D. A., and Mehta, A. (2012). The effects of age on temporal fine structure sensitivity in monaural and binaural conditions. *International Journal of Audiology*, *51*(10), 715–721. https://doi.org/10.3109/14992027.2012.690079

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*(3), 351-362. https://doi.org/10.3758/BF03206811

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *15*(2), 133-137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x

National Health Services. (2013). General Practice Physical Activity Questionnaire. Retrieved from https://www.gov.uk/government/publications/general-practice-physical-activity-questionnaire-gppaq

Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech. *Information Fusion*, *11*(1), 4-11. https://doi.org/10.1016/j.inffus.2009.04.001

Naylor, G. (2016). Theoretical issues of validity in the measurement of aided speech reception threshold in noise for comparing nonlinear hearing aid systems. *Journal of the American Academy of Audiology*, *27*(7), 504-514. https://doi.org/10.3766/jaaa.15093

Netz, Y., Dwolatzky, T., Zinker, Y., Argov, E., Agmon, R., 2011. Aerobic fitness and multidomain cognitive function in advanced age. International Psychogeriatrics, *23*(1), 114-24. doi: 10.1017/S1041610210000797

Neumann, O. (1996). Theories of attention. In O. Neumann & W. Prinz (Eds.), *Handbook of Perception and Action*, Vol. *3*, (299-446). London; Academic Press.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085-1099. https://doi.org/10.1121/1.408469

Nishiwaki, Y., Breeze, E., Smeeth, L., Bulpitt, C. J., Peters, R., & Fletcher, A. E. (2004). Validity of the Clock-Drawing Test as a screening tool for cognitive impairment in the elderly. *American Journal of Epidemiology*, *160*(8), 797-807. https://doi.org/10.1093/aje/kwh288

Noble, W., Jensen, N. S., Naylor, G., Bhullar, N., & Akeroyd, M. A. (2013). A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: The SSQ12. *International Journal of Audiology*, *52*(6), 409-412. https://doi.org/10.3109/14992027.2013.781278

Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012. Musical experience offsets age-related delays in neural timing. *Neurobiology of Aging*, *33*(7), 1483.e1-4. http://dx.doi.org/ 10.1016/j.neurobiolaging.2011.12.015.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653-661. doi: 10.1097/AUD.0b013e3181b412e9

Parbery-Clark, A., Strait, D.L., Anderson, S., Hittner, E., Kraus, N., 2011. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE*, *6*(5): e18082. https://doi.org/10.1371/journal.pone.0018082

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics*, *65*(4), 553-567. https://doi.org/10.3758/BF03194582

Pei, K., Chen, F., & Zheng, D. (2016). Effect of autonomic cardiac modulation on speech perception in noise. In *2016 Computing in Cardiology Conference (CinC*, pp. 521-524). IEEE.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*(1), 73-89. doi: 10.1146/annurev-neuro-062111-150525

Phillips, N. A. (2016). The implications of cognitive aging for listening and the framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 44S-51S. doi: 10.1097/AUD.0000000000000309.

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology*, *42*(sup1), 59-67. https://doi.org/10.3109/14992020309074625

Pichora-Fuller, M. K., Alain, C., & Schneider, B. A. (2017). Older adults at the cocktail party. In *The auditory system at the cocktail party* (pp. 227-259). Springer, Cham.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., ... & Naylor, G. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*, 5S-27S. doi: 10.1097/AUD.0000000000000312

Pichora-Fuller, K., & MacDonald, E. (2007). Auditory temporal processing deficits in older listeners: From a review to a future view of Presbycusis. In *Proceedings of the International Symposium on Auditory and Audiological Research* (Vol. 1, pp. 291-300).

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*(2-3), 227–272. https://doi.org/10.1016/0001-6918(94)90004-3

Pollack, I. (1975). Auditory informational masking. *Journal of the Acoustical Society of America*, *57*(S1). https://doi.org/10.1121/1.1995329

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3-25. https://doi.org/10.1080/00335558008248231

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25-42.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160-174. http://dx.doi.org/10.1037/0096-3445.109.2.160

Redick, T. S., & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(5), 713-721. https://doi.org/10.1002/acp.1224

Reitan, R. M. (1992). *Trail Making Test: Manual for Administration and Scoring.* Tucson, AZ: Reitan Neuropsychology Laboratory.

Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1996). The structure of normal human attention: the test of everyday attention. *The Journal of the International Neuropsychological Society. 2*(6), 525–534. doi: 10.1017/S1355617700001697

Robertson, I. H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1994). *The Test of Everyday Attention: TEA.* Bury St. Edmunds, UK: Thames Valley Test Company.

Rönnberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, *58*(5), 247-261. https://doi.org/10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*, 31. doi: https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(sup2), S99-S105. doi:10.3389/fnsys.2013.00031

Rosemann, S., & Thiel, C. M. (2018). Audio-visual speech processing in age-related hearing loss: Stronger integration and increased frontal lobe recruitment. *NeuroImage*, *175*, 425-437. https://doi.org/10.1016/j.neuroimage.2018.04.023

Ross, B., Fujioka, T., Tremblay, K. L., and Picton, T. W. (2007). Aging in binaural hearing begins in mid-life: Evidence from cortical auditory evoked responses to changes in interaural phase. *Journal of Neuroscience*, *27*(42), 11172–11178. https://doi.org/10.1523/JNEUROSCI.1813-07.200

Rothauser, E.H., Chapman, W.D., Guttman, N., Hecker, M.H.L., Nordby, K.S., Silbiger, H.R., Urbanek, G.E., Weistock, M., McGee, V.E., Pachl, U. P., & Voiers, W.D. (1969). IEEE Recommended practice for speech quality measurements. *IEEE Trans. Audio Acoust*., 225-246.

Rudner, M., Mishra, S., Stenfelt, S., Lunner, T., & Rönnberg, J. (2016). Seeing the talker’s face improves free recall of speech for young adults with normal hearing but not older adults with hearing loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, *59*(3), 590–599.

Salthouse, T.A., 2006. Mental exercise and mental aging: evaluating the validity of the "use it or lose it" hypothesis. *Perspectives on Psychological Science*, *1*(1), 68-87. doi: 10.1111/j.1745-6916.2006.00005.x

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review. 103*(3), 403–428. doi: 10.1037/0033-295X. 103.3.403

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. Infancy, 2(4), 419-436. https://doi.org/10.1207/S15327078IN0204_02

Schneider, B. A., Daneman, M., Murphy, D. R., & See, S. K. (2000). Listening to discourse in distracting settings: the effects of aging. *Psychology and aging*, *15*(1), 110. http://dx.doi.org/10.1037/0882-7974.15.1.110

Schneider, B. A., and Hamstra, S. J. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. *The Journal of the Acoustical Society of America*, *106*(1), 371–380. doi: 10.1121/1.427062

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: implications for speech perception. *Seminars in Hearing*, 22(3), 227-240. doi: 10.1055/s-2001-15628

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

Schneider, B. A., Pichora-Fuller, K., & Daneman, M. (2010). Effects of senescent changes in audition and cognition on spoken language comprehension. In *The Aging Auditory System* (167-210). New York, NY: Springer.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., & Lamb, M. (1994). Gap detection and the precedence effect in young and old adults. *The Journal of the Acoustical Society of America*, *95*(2), 980-991. https://doi.org/10.1121/1.408403

Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in Aging Neuroscience*, *6*, 307. doi: 10.3389/fnagi.2014.00307

Schurman, J., Brungart, D., & Gordon-Salant, S. (2014). Effects of masker type, sentence context, and listener age on speech recognition performance in 1-back listening tasks. *The Journal of the Acoustical Society of America*, *136*(6), 3337-3349. https://doi.org/10.1121/1.4901708

Sekuler, A.B,, Bennett, P.J., Mamelak, M. (2000). Effects of aging on the useful field of view. *Experimental Aging Research*, *26*(2), 103-120. https://doi.org/10.1080/036107300243588

Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., & Farrell, C. (2012). Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception. *Ear and Hearing*, *33*(6), 709. doi: 10.1097/AUD.0b013e31825aab15

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, *12*(4), 283-299. https://doi.org/10.1177/1084713808325306

Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *The Journal of the Acoustical Society of America*, *124*(2), 1294-1305. https://doi.org/10.1121/1.2949399

Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2013). Time course and cost of misdirecting auditory spatial attention in younger and older adults. *Ear and Hearing*, *34*(6), 711-721. doi: 10.1097/AUD.0b013e31829bf6ec

Snell, K.B. (1997). Age-related changes in temporal gap detection. *The Journal of the Acoustical Society of America*, *101*(4), 2214-2220. https://doi.org/10.1121/1.418205

Snyder, J. S., and Alain, C. (2005). Age-related changes in neural activity associated with concurrent vowel segregation. *Cognitive Brain Research*, *24* (3), 492-499. https://doi.org/10.1016/j.cogbrainres.2005.03.002

Sommers, M. S., Hale, S., Myerson, J., Rose, N., Tye-Murray, N., & Spehar, B. (2011). Listening comprehension across the adult lifespan. *Ear and Hearing*, *32*(6), 775-781. doi: 10.1097/AUD.0b013e3182234cf6

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 and Hearing*, *26*(3), 263-275. doi: 10.1097/00003446-200506000-00003

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition*, *92*(3), B13-B23. https://doi.org/10.1016/j.cognition.2003.10.005

Souza, P., Arehart, K., & Neher, T. (2015). Working memory and hearing aid processing: Literature findings, future directions, and clinical applications. *Frontiers in Psychology*, *6*,

1894. https://doi.org/10.3389/fpsyg.2015.01894

Spankovich, C., Gonzalez, V. B., Su, D., & Bishop, C. E. (2018). Self reported hearing difficulty, tinnitus, and normal audiometric thresholds, the National Health and Nutrition Examination Survey 1999–2002. *Hearing Research*, *358*, 30-36. https://doi.org/10.1016/j.heares.2017.12.001

Stacey, P. C., Murphy, T., Sumner, C. J., Kitterick, P. T., & Roberts, K. L. (2014). Searching for a talking face: The effect of degrading the auditory signal. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2106. http://dx.doi.org/10.1037/a0038220

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., & Smith, J. (2018). Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures. *Journal of Speech, Language, and Hearing Research*, *61*(6), 1463-1486. https://doi.org/10.1044/2018_JSLHR-H-17-0257

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212-215. https://doi.org/10.1121/1.1907309

Swearer J. (2011) Visual Angle. In J.S. Kreutzer, J DeLuca., & B. Caplan (eds) *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-79948-3

Tales, A., Muir, J. L., Bayer, A., & Snowden, R. J. (2002). Spatial shifts in visual attention in normal ageing and dementia of the Alzheimer type. Neuropsychologia, *40*(12), 2000–2012. https://doi.org/10.1016/S0028-3932(02)00057-X

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *48*(3), 380–398.

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, *16*(3), 457-472. https://doi.org/10.1080/09541440340000268

To, W. M., & Chung, A. (2014). Noise in restaurants: Levels and mathematical model. *Noise and Health*, *16*(73), 368-373. Retrieved from: http://www.noiseandhealth.org/text.asp?2014/16/73/368/144412

Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E., Klein, R., Levy, S., ... & Cruickshanks, K. J. (2015). Self-reported hearing difficulties among adults with normal audiograms: The Beaver Dam Offspring Study. *Ear and Hearing*, *36*(6), e290-e299. doi: 10.1097/AUD.0000000000000195

Tun, P. A. (1998). Fast noisy speech: age differences in processing rapid speech with background noise. *Psychology and Aging*, *13*(3), 424-434. http://dx.doi.org/10.1037/0882-7974.13.3.424

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. http://dx.doi.org/10.1037/a0014802

Tun, P. A., Williams, V. A., Small, B. J., & Hafter, E. R. (2012). The effects of aging auditory processing and cognition. *American Journal of Audiology*, *21*(2), 344-350. https://doi.org/10.1044/1059-0889(2012/12-0030)

Tun, P. A., & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. Journals of

Gerontology Series B-Psychological Sciences and Social Sciences, *54B* (5), P317–P327. doi:10.1093/geronb/54B.5.P317.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing. *Ear and Hearing*, *28*(5), 656-668. doi: 10.1097/AUD.0b013e31812f7185

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and Aging*, *31*(4), 380-389. http://dx.doi.org/10.1037/pag0000094

Tyler, R. S., Summerfield Q., Wood, E. J., & Fernandes, M. A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *72*(3) 740-752. https://doi.org/10.1121/1.388254

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception & Psychophysics*, *79*(2), 396–403. https://doi.org/10.3758/s13414-016-1238-9

Van Rooij, J. C. G. M., Plomp, R., & Orlebeke, J. F. (1989). Auditive and cognitive factors in speech perception by elderly listeners. I: Development of test battery. *The Journal of the Acoustical Society of America*, *86*(4), 1294-1309. https://doi.org/10.1121/1.398744

van den Tillaart-Haverkate, M., de Ronde-Brons, I., Dreschler, W. A., & Houben, R. (2017). The influence of noise reduction on speech intelligibility, response times to speech, and perceived listening effort in normal-hearing listeners. *Trends in Hearing*, *21*, 1-13.

https://doi.org/10.1177/2331216517716844

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053-1065. http://dx.doi.org/10.1037/0096-1523.34.5.1053

Vermeire, K., Knoop, A., Boel, C., Auwers, S., Schenus, L., Talaveron-Rodriguez, M., ... & De Sloovere, M. (2016). Speech recognition in noise by younger and older adults: Effects of age, hearing loss, and temporal resolution. *Annals of Otology, Rhinology & Laryngology*, *125*(4), 297-302. https://doi.org/10.1177/0003489415611424

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index, *Journal of the American Academy of Audiology*, *23*(10), 779-788. https://doi.org/10.3766/jaaa.23.10.4

Wang, J., Tian, J., Wang, R., & Benson, V. (2013). Increased attentional focus modulates eye movements in a mixed antisaccade task for younger and older adults. *PLoS ONE*, *8*(4), e61566. https://doi.org/10.1371/journal.pone.0061566

Ward, K. M., Shen, J., Souza, P. E., & Grieco-Calub, T. M. (2017). Age-related differences in listening effort during degraded speech recognition. *Ear and Hearing*, *38*(1), 74-84. doi: 10.1097/AUD.0000000000000355

Wascher, E., Schneider, D., Hoffmann, S., Beste, C., & Sänger, J. (2012). When compensation fails: attentional deficits in healthy ageing caused by visual distraction. *Neuropsychologia*, *50*(14), 3185-3192. https://doi.org/10.1016/j.neuropsychologia.2012.09.033

Watson, K., Parker, D., Leahy, A., Piepers, D., & Stevens, K., (2018). Time travelling with technology. *Australian Journal of Dementia Care*, *4*(7), 16-17. Retrieved from: https://www.westernsydney.edu.au/__data/assets/pdf_file/0007/1615741/Issue_38_Aug-Sep_2018_Watson_et_al.pdf

Wayne, R. V., Hamilton, C., Jones Huyck, J., & Johnsrude, I. S. (2016). Working memory training and speech in noise comprehension in older adults. *Frontiers in Aging Neuroscience*, *8*, 49. https://doi.org/10.3389/fnagi.2016.00049

Wechsler, D. (1981) Adult Intelligence Scale- revised. New York: Psychological Corporation.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22,* 498.

Weisser, A., & Buchholz, J. M. (2019). Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *The Journal of the Acoustical Society of America*, *145*(1), 349-360. https://doi.org/10.1121/1.5087567

Weisser, A., Buchholz, J. M., Oreinos, C., Badajoz-Davila, J., Galloway, J., Beechey, T., & Keidser, G. (2019). The Ambisonic Recordings of Typical Environments (ARTE) Database. *Acta Acustica united with Acustica*, *105*(4), 695-713. https://doi.org/10.3813/AAA.919349

Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear and Hearing*, *37,* 35S-43S. doi: 10.1097/AUD.0000000000000310

Wingfield, A., Amichetti, N. M., & Lash, A. (2015). Cognitive aging and hearing acuity: modeling spoken language comprehension. *Frontiers in Psychology*, *6*, 684. https://doi.org/10.3389/fpsyg.2015.00684

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. https://doi.org/10.1111/j.0963-7214.2005.00356.x

Winneke, A. H., & Phillips, N. A. (2011). Does audiovisual speech offer a fountain of youth for old ears? An event-related brain potential study of age differences in audiovisual speech perception. *Psychology and Aging*, *26*(2), 427-438. 10.1037/a0021683

Wolfe, J. M. (2010). Visual search. *Current Biology*, *20*(8), R346-R349. https://doi.org/10.1016/j.cub.2010.02.016

Wong, P. C., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*(3), 693-703. https://doi.org/10.1016/j.neuropsychologia.2008.11.032

World Health Organization. (2015). *World report on ageing and health*. Geneva, Switzerland. World Health Organization Press.

Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., DeVries, S., & Oleson, J. (2019). Efficacy and Effectiveness of Advanced Hearing Aid Directional and Noise Reduction Technologies for Older Adults With Mild to Moderate Hearing Loss. *Ear and Hearing*, *40*(4), 805-822. doi: 10.1097/AUD.0000000000000672

Xia, J., Nooraei, N., Kalluri, S., & Edwards, B. (2015). Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *137*(4), 1888-1898. https://doi.org/10.1121/1.4916599

Yamasoba, T., Lin, F. R., Someya, S., Kashio, A., Sakamoto, T., & Kondo, K. (2013).

Current concepts in age-related hearing loss: epidemiology and mechanistic pathways. *Hearing Research*, *303*, 30-38. https://doi.org/10.1016/j.heares.2013.01.021

Zanto, T. P., & Gazzaley, A. (2014). Attention and ageing. In *The Oxford Handbook of Attention* (pp. 927-971). Oxford University Press, Oxford, UK.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, *32*(4), 498–510. doi:10.1097/ AUD.0b013e31820512bb.

Zendel, B.R., Alain, C., 2012. Musicians experience less age-related decline in central auditory processing. *Psychology and Ageing*, *27*(2), 410-417. doi: 10.1037/a0024816

Appendix A: Question-and-Answer Task Items

Category	Question	True Answer	False Answer
Days	What day comes after Friday?	Saturday	Thursday
	What day comes after Monday?	Tuesday	Sunday
	What day comes after Saturday?	Sunday	Friday
	What day comes after Sunday?	Monday	Saturday
	What day comes after Thursday?	Friday	Wednesday
	What day comes after Tuesday?	Wednesday	Monday
	What day comes after Wednesday?	Thursday	Tuesday
	What day comes before Friday?	Thursday	Saturday
	What day comes before Monday?	Sunday	Tuesday
	What day comes before Saturday?	Friday	Sunday
	What day comes before Sunday?	Saturday	Monday
	What day comes before Thursday?	Wednesday	Friday
	What day comes before Tuesday?	Monday	Wednesday
	What day comes before Wednesday?	Tuesday	Thursday
Months	What month comes after April?	May	March
	What month comes after August?	September	July
	What month comes after December?	January	November
	What month comes after February?	March	January
	What month comes after January?	February	December
	What month comes after July?	August	June
	What month comes after June?	July	May
	What month comes after March?	April	June
	What month comes after May?	June	April
	What month comes after November?	December	October
	What month comes after October?	November	September
	What month comes after September?	October	August

	What month comes before April?	March	May
	What month comes before August?	July	September
	What month comes before December?	November	January
	What month comes before February?	January	March
	What month comes before January?	December	February
	What month comes before July?	June	August
	What month comes before June?	May	July
	What month comes before March?	February	April
	What month comes before May?	April	June
	What month comes before November?	October	December
	What month comes before October?	September	November
	What month comes before September?	August	October
Colours	What colour are clouds?	white	green
	What colour are peas?	green	black
	What colour are raspberries?	red	white
	What colour are strawberries?	red	yellow
	What colour is a banana?	yellow	red
	What colour is a cucumber?	green	blue
	What colour is a frog?	green	silver
	What colour is a lemon?	yellow	black
	What colour is a lime?	green	silver
	What colour is a polar bear?	white	green
	What colour is broccoli?	green	white
	What colour is coal?	black	yellow
	What colour is corn?	yellow	silver
	What colour is grass?	green	blue
	What colour is milk?	white	red
	What colour is peanut butter?	brown	yellow

	What colour is snow?	white	brown
	What colour is the sky?	blue	silver
	What colour is vanilla ice cream?	white	red
	What colour is a 10 cent coin?[2]	silver	gold
	What colour is vegemite?[1]	brown	green
Opposites	What is the opposite of clean?	dirty	open
	What is the opposite of closed?	open	hot
	What is the opposite of cold?	hot	clean
	What is the opposite of dirty?	clean	up
	What is the opposite of down?	up	wet
	What is the opposite of dry?	wet	slow
	What is the opposite of fast?	slow	outside
	What is the opposite of high?	low	cold
	What is the opposite of hot?	cold	low
	What is the opposite of inside?	outside	high
	What is the opposite of low?	high	yes
	What is the opposite of no?	yes	on
	What is the opposite of off?	on	dirty
	What is the opposite of on?	off	closed
	What is the opposite of open?	closed	inside
	What is the opposite of outside?	inside	fast
	What is the opposite of slow?	fast	down
	What is the opposite of up?	down	dry
	What is the opposite of wet?	dry	no
	What is the opposite of yes?	no	off
Sizes	Which is bigger, a bear or a rat?	bear	rat
	Which is bigger, a butterfly or a giraffe?	giraffe	butterfly
	Which is bigger, a cat or a lion?	lion	cat

	Which is bigger, a dog or a horse?	horse	dog
	Which is bigger, a moose or a bee?	moose	bee
	Which is bigger, a mosquito or a donkey?	donkey	mosquito
	Which is bigger, a panda or a fly?	panda	fly
	Which is bigger, a pig or a cow?	cow	pig
	Which is bigger, a rabbit or an ant?	rabbit	ant
	Which is bigger, an elephant or a mouse?	elephant	mouse
	Which is smaller, a golf ball or a planet?	golf ball	planet
	Which is smaller, a house or a tent?	tent	house
	Which is smaller, a mountain or a tree?	tree	mountain
	Which is smaller, a pea or a watermelon?	pea	watermelon
	Which is smaller, a puddle or a lake?	puddle	lake
	Which is smaller, a spoon or a bed?	spoon	bed
	Which is smaller, a toy or a bus?	toy	bus
	Which is smaller, a tree or a leaf?	leaf	tree
	Which is smaller, a truck or a bike?	bike	truck
	Which is smaller, an adult or a child?	child	adult
	Which is bigger, a kangaroo or a koala?[1]	kangaroo	koala
Numbers	How many cents are there in a dollar?	100	24
	How many days are there in a week?	7	3
	How many ears do you have?	2	14
	How many eyes do you have?	2	60
	How many feet do you have?	2	15
	How many fingers do you have?	10	100
	How many hands do you have?	2	9
	How many hours are there in a day?	24	8
	How many items are there in a dozen?	12	17
	How many legs do you have?	2	14

	How many minutes are there in an hour?	60	2
	How many months are there in a year?	12	6
	How many seasons are there ?	4	16
	How many seconds are there in a minute?	60	4
	How many sides does a square have?	4	7
	How many sides does a triangle have?	3	18
	How many toes do you have?	10	13
	How many wheels does a bike have?	2	10
	How many wheels does a car have?	4	5
	What is 10 minus 2?	8	1
	What is 10 minus 4?	6	3
	What is 10 minus 5?	5	12
	What is 2 plus 1?	3	6
	What is 2 plus 2?	4	8
	What is 2 plus 3?	5	24
	What is 2 plus 4?	6	20
	What is 2 plus 5?	7	13
	What is 2 plus 6?	8	11
	What is 2 plus 7?	9	18
	What is 2 plus 8?	10	14
	What is 2 times 1?	2	18
	What is 2 times 10?	20	7
	What is 2 times 2?	4	16
	What is 2 times 3?	6	15
	What is 2 times 4?	8	3
	What is 2 times 5?	10	19
	What is 2 times 6?	12	7
	What is 2 times 7?	14	60

	What is 2 times 8?	16	4
	What is 2 times 9?	18	5
	What is 3 minus 2?	1	5
	What is 3 plus 1?	4	100
	What is 3 plus 2?	5	1
	What is 3 plus 3?	6	2
	What is 3 plus 4?	7	12
	What is 3 plus 5?	8	24
	What is 3 plus 6?	9	2
	What is 3 plus 7?	10	5
	What is 4 minus 2?	2	7
	What is 4 minus 3?	1	13
	What is 4 plus 1?	5	10
	What is 4 plus 2?	6	9
	What is 4 plus 3?	7	14
	What is 4 plus 4?	8	3
	What is 4 plus 5?	9	24
	What is 4 plus 6?	10	16
	What is 5 minus 2?	3	15
	What is 5 minus 3?	2	11
	What is 5 minus 4?	1	19
	What is 5 plus 1?	6	20
	What is 5 plus 2?	7	4
	What is 5 plus 3?	8	60
	What is 5 plus 4?	9	17
	What is 5 plus 5?	10	3
	What is 6 minus 2?	4	100
	What is 6 minus 3?	3	14

	What is 6 minus 4?	2	60
	What is 6 minus 5?	1	18
	What is 7 minus 2?	5	100
	What is 7 minus 3?	4	1
	What is 7 minus 4?	3	11
	What is 7 minus 5?	2	18
	What is 8 minus 2?	6	20
	What is 8 minus 3?	5	2
	What is 8 minus 4?	4	13
	What is 8 minus 5?	3	4
	What is 9 minus 2?	7	11
	What is 9 minus 3?	6	2
	What is 9 minus 4?	5	17
	What is 9 minus 5?	4	7
	What is half of 10?	5	8
	What is half of 12?	6	9
	What is half of 14?	7	10
	What is half of 16?	8	12
	What is half of 18?	9	15
	What is half of 2?	1	19
	What is half of 20?	10	6
	What is half of 4?	2	9
	What is half of 6?	3	8
	What is half of 8?	4	10
	What number comes after 1?	2	8
	What number comes after 10?	11	9
	What number comes after 11?	12	10
	What number comes after 12?	13	11

	What number comes after 13?	14	12
	What number comes after 14?	15	13
	What number comes after 15?	16	14
	What number comes after 16?	17	15
	What number comes after 17?	18	16
	What number comes after 18?	19	17
	What number comes after 2?	3	1
	What number comes after 3?	4	2
	What number comes after 4?	5	3
	What number comes after 5?	6	4
	What number comes after 6?	7	5
	What number comes after 7?	8	6
	What number comes after 8?	9	7
	What number comes after 9?	10	8
	What number comes after 19?	20	18
	What number comes before 10?	9	11
	What number comes before 11?	10	12
	What number comes before 12?	11	13
	What number comes before 13?	12	14
	What number comes before 14?	13	15
	What number comes before 15?	14	16
	What number comes before 16?	15	17
	What number comes before 17?	16	18
	What number comes before 18?	17	19
	What number comes before 19?	18	20
	What number comes before 2?	1	3
	What number comes before 20?	19	24
	What number comes before 3?	2	4

	What number comes before 4?	3	5
	What number comes before 5?	4	6
	What number comes before 6?	5	7
	What number comes before 7?	6	8
	What number comes before 8?	7	9
	What number comes before 9?	8	10