RELATION OF SHORT-TERM WORKING MEMORY AND SPEECH PERCEPTION:  
A CROSS-SECTIONAL STUDY

BY
LYNN M. BIELSKI

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Speech and Hearing Science in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Associate Professor Charissa R. Lansing, Chair
Associate Professor Ron D. Chambers
Associate Professor Cynthia J. Johnson
Instructor David M. Gooler
ABSTRACT

The current study examined the relation between short-term working memory and speech perception. Visual cues have been shown to improve speech understanding; therefore, visual cues from the talker’s face might also enhance short-term working memory. Word recall was examined as function of age (children, young adults, older adults), modality, (auditory-visual, auditory-only) and acoustic condition (quiet, noise). All participants had normal hearing bilaterally and good vision. A running memory task was utilized. Participants listened to strings of bisyalbic words and were instructed to repeat the last four words heard in any order. Additional measures included tests of lip-reading proficiency, receptive vocabulary knowledge, inhibition, subjective workload ratings and working memory. It was hypothesized that age would influence recall errors, with young adults having fewer errors compared to children and older adults. Fewer recall errors were expected when visual cues were present, and fewer recall errors were predicted in quiet than noise. Results revealed age group differences. Children showed a higher number of recall errors compared to adults. Inconsistent with what has been reported in the literature, older adult’s recall performance was similar to that of young adults. Few recall errors were present for the last word heard, indicating the stimuli were audible and intelligible. Recall errors increased for words further back in the string, indicating an effect of short-term working memory. All age groups demonstrated fewer errors when visual cues from the talker’s face were provided, especially in noise independent of lip-reading proficiency. Workload ratings collected from the adult groups showed greater perceived frustration in noise compared to quiet. With the exception of lip-reading proficiency and work load ratings, other additional measures did not help to
explain individual differences. An examination of types of recall errors revealed children had more part word and phoneme substitution errors compared to adults. Across age groups, fewer part word and phoneme substitution errors were present with visual cues compared to without. Future studies should explore the utility of visual cues in facilitating short-term working memory in listeners with hearing loss, signal-to-noise ratio loss, and loss of temporal/spectral resolution abilities.
ACKNOWLEDGEMENTS

I would like to thank my adviser, Dr. Charissa Lansing, for her many years of invaluable guidance. I would also like to thank the members of my committee, Dr. Ron Chambers, Dr. Cynthia Johnson, Dr. David Gooler for their insight and suggestions. Thanks to Dr. Robert Wickesberg for serving as a manuscript reader. I am very grateful to the participants who completed the study, and to the undergraduate research assistants who helped with data collection and coding. Finally, I would like to thank my family, friends and classmates for their love and support.
# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ........................................................................................................... 1

CHAPTER 2: LITERATURE REVIEW .......................................................................................... 12

CHAPTER 3: METHODS .............................................................................................................. 49

CHAPTER 4: RESULTS ................................................................................................................ 67

CHAPTER 5: DISCUSSION .......................................................................................................... 103

REFERENCES ............................................................................................................................ 123

APPENDIX A: PARTICIPANT QUESTIONNAIRES ..................................................................... 155

APPENDIX B: ERROR CODE DEFINITIONS ........................................................................... 157
CHAPTER 1

Introduction

Before a speaker’s message can be understood, a sequence of processes must be successfully executed. First, the desired signal has to be captured by the peripheral auditory system, and then transduced from mechanical energy to fluid energy, and finally neural impulses transmit the signal on to higher structures in the brainstem and cortex. In the brain, the acoustic signal and its linguistic components are integrated to create a cognitive framework, built by lexical representation (Marslen-Wilson, Brown & Tyler 1998) and previous knowledge of context (i.e., phonological, semantic, pragmatic; Boothroyd & Nittrouer, 1988). The most recently acquired information is manipulated and maintained by short-term memory, and is also linked to information in long-term memory stores, if appropriate (Baddeley & Hitch, 1974).

Short-term memory refers to memory for events that have occurred in the very recent past (Gathercole, 1998) and working memory is a subcomponent of short-term memory. Working memory serves to manipulate and maintain recent information, retrieving information from long-term memory stores and integrating information from many sources to formulate a response. Running memory refers to a situation in which the end is unknown; therefore information that has been maintained over a period of time may be discarded in order to allow new information to be accessed (Pollack, Johnson & Knaff, 1959).

Memory components, including short-term memory, allow the listener to maintain the most recent information and manipulate that information to formulate a response. This process can be made more difficult in the presence of noise or other cognitive functions that take
resources away from short-term memory (McCoy, Tun, Cox, Colangelo, Stewart & Wingfield, 2005; Rabbitt, 1968). Listening to speech in noise can be challenging for any listener, however, it has been noted that children and older adults tend to have more difficulty with speech perception in noise than young adults (Johnson, 2000; Larsby, Hällgren, Lyxell, & Arlinger, 2005; Pichora-Fuller, Schneider, & Daneman, 1995). Changes in short-term memory performance across the lifespan may contribute to these listeners’ difficulty with speech in noise.

Declines or deficits in auditory and visual sensitivity may also contribute to difficulties with perception of speech. Loss of auditory and visual acuity plays a role in problems understanding speech and cognitive function (Lindenberger & Baltes, 1994; Salthouse, 1996; Stenfelt & Rönnberg, 2009); however, reduced working memory capacity may also cause difficulties (Grady & Craik, 2000; Schneider & Pichora-Fuller, 2000). Loss of sensory acuity and declines in working memory result in older adults’ poorer speech perception performance in quiet settings compared to young adults, and noisy environments are even more detrimental to speech understanding (Gosselin & Gagné, 2011; Murphy, Craik, Li, & Schneider, 2000).

Cognitive factors, such as short-term working memory, also play a role in speech perception. Children may be developing short-term memory capacity (Gathercole, 1998), whereas in older adults difficulties may arise due to declines in this capacity. Compared to young adults, older adults’ speech perception may be negatively affected by factors such as loss of hearing and visual acuity and cognitive changes. The perception of speech in children may differ from that of young adults because their auditory and cognitive abilities have not fully matured (Sussman, 1993). Furthermore, compared to young adults, children have been shown to be less sensitive to acoustic cues such as voice-onset time and formant frequency transitions (Elliott,
Visual information plays an important role in speech perception. It can enhance perception of audible speech (Arnold & Hill, 2001; Reisberg, McLean & Goldfield, 1987). Adding visual cues to auditory speech can improve perception in noisy environments (Grant & Braida, 1991; Miller, Heise, & Lichten, 1951; Sumby & Pollack, 1954). Visual cues that are not matched to auditory stimuli are perceived as a fusion of both stimuli (McGurk & MacDonald, 1976). Evidence from neuroimaging studies demonstrate increased neural activation in the auditory cortex when an auditory signal is presented with a visual signal (Calvert, Brammer, Bullmore, Campbell, Iverson, & David, 1999).

**The Baddeley and Hitch Model of Short-term Memory**

A three component model of working memory was proposed by Baddeley and Hitch (1974) and recently has been revised to include a fourth component (Baddeley, 2000). The current model consists of the central executive and three slave systems: the phonological loop, visuospatial sketchpad, and episodic buffer. While there are other models of short-term working memory, the Baddeley and Hitch model is utilized as a framework for the present study because it contains specialized components for auditory and visual information.

The central executive is responsible for the control of attention and problem solving (Baddeley, 1986). It coordinates working memory operations and is responsible for the retrieval of information from long-term memory stores. The phonological loop is composed of the
phonological store and subvocal rehearsal, which maintain, rehearse and manipulate acoustic information. The visuospatial sketchpad has two subcomponents. The visual store maintains and manipulates physical features of objects such as color and shape (Logie, 1986). The spatial mechanism assists in motor movements, such as gesture and dance (Smyth & Pendleton, 1990). The episodic buffer is a limited capacity store, capable of holding information in a multi-dimensional code (Baddeley, 2000). It can integrate information from the phonological loop and visuospatial sketchpad. The central executive has access to the episodic buffer, and can focus attention selectively in order to restrict the information held and processed there.

Perhaps the four working memory components (Baddeley, 2000; Baddeley & Hitch, 1974) function as a dynamic mesh. The mesh would allow desired information to enter the cognitive processing system, while simultaneously keeping out undesired information for the function of speech perception. The mesh develops over time in children, with the holes narrowing as the system becomes more refined; while in older adults holes are widening, allowing unwanted signals into the system and letting out desired information. The central executive is responsible for the initial entry of information into the system. It can immediately block irrelevant information by directing attention only to specific stimuli for processing. From this point, the phonological loop and visuospatial sketchpad screen the information further, allowing the relevant information to be encoded, maintained and rehearsed. The episodic buffer would act as a very refined component of the mesh, because it is controlled by the central executive which thorough conscious awareness focuses on specific information to be processed. It also must screen the information it combines from the phonological loop and visuospatial sketchpad, carefully interpreting only relevant information in order to form a multi-dimensional representation.
Alternative Models of Short-Term Memory

There are several other models of short-term working memory. A few of the models are highlighted here. These models are discussed to illustrate how concepts from some models are incorporated into others, and to show the variety of ways short-term working memory has been conceptualized.

Control of the activated information in working memory is the focus of the working memory model proposed by Hasher and Zacks (1988). Inhibition of irrelevant information is vital to maintaining activation of information needed for the overall goal. At the time of encoding, inhibition is utilized to ensure only relevant information receives activation resources.

The importance of executive functions is emphasized by some models. Attentional focus as a means of maintaining the memory trace is suggested by Cowan’s (1999, 2005) model of short-term memory. Memory traces for information in short-term memory can be reactivated by the executive function, attention. Johnson (1992) included the concept that the memory trace can be refreshed by shifting attentional focus to information that was recently re-activated.

The ideas suggested by Cowan and Johnson were the foundations for the time-based resource-sharing (TBRS) model of short-term memory. The TBRS model (Barrouillet & Camos, 2007; Portrat, Camos, & Barrouillet, 2009) suggests that items in short-term memory decay when attentional resources are shifted from maintenance to processing. The spotlight of attention must be shifted from processing to maintenance for a very brief time period to activate the memory trace. The memory trace must be refreshed in order to maintain and later recall information.
A unitary model of working memory, capacity theory, has also been proposed (Just & Carpenter, 1992). In this model, working memory has a limited amount of activation which is distributed over all cognitive tasks. When working memory capacity has reached maximum, the amount of activation must be decreased before storage and retrieval resources can be allocated to new tasks.

The concept of limited capacity also plays a role in the effortfulness hypothesis proposed by Rabbitt (1968). The effortfulness hypothesis states that the amount of perceptual effort a person exerts to perceive a stimulus will affect other cognitive resources, such as working memory. A situation requiring a great deal of perceptual effort will result in fewer cognitive resources available for functions such as working memory. Rabbitt (1968) showed evidence for the effortfulness hypothesis, as digits presented auditorily were more accurately recalled when presented in quiet than in noise. In noise, the listeners were thought to exert greater effort to perceive the desired stimuli compared to quiet, thereby leaving fewer cognitive resources available for working memory.

Loss of hearing acuity also increases perceptual effort, decreasing available processing resources, as evidenced from experimental results reported by Rabbitt (1991). In the experiment, older adults with normal hearing and another group with mild hearing loss were asked to recall word lists. Audibility was controlled for by adding noise. Participants with normal hearing showed higher recall scores than those with normal hearing. Rabbitt interpreted these results to suggest the loss of peripheral sensitivity causes more perceptual effort to be exerted, therefore reducing available cognitive resources.
A recent model of short-term working memory, the Ease of Language Understanding model (ELU; Rönnberg, 2003a; Rönnberg, Rudner, Lunner & Zekveld, 2010), includes concepts found in several other models. The ELU model combines an element with functions similar to that of the episodic buffer in the Baddeley and Hitch model, processing capacity, processing speed and resource availability. In the ELU model multi-modal language input results in rapid, automatic, multi-modality binding of phonology (RAMBPHO, Rönnberg, Rudner, Foo & Lunner, 2008). Under optimal conditions, processing of the language input is implicit, and requires the recruitment of few cognitive resources. In the presence of background noise or other conditions resulting in degradation of the language input (e.g., peripheral hearing loss), processing is explicit, and additional cognitive resources are employed for language understanding.

**Evidence for the Development of Short-Term Memory**

The development of short-term working memory has been examined using the Baddeley and Hitch model. There is evidence to suggest the components of the model develop through childhood. Performance on working memory tasks, such as digit span, improves through childhood (Gathercole, 1998). The capacity of the central executive increases with age, as evidenced by increased listening span (Siegel, 1994). The subvocal rehearsal subcomponent of the phonological loop surfaces around age seven in children (Gathercole & Hitch, 1993). Evidence from investigations utilizing object and picture stimuli suggest children younger than seven years of age rely more on visual cues to remember stimuli (Hitch, Halliday, Schaafstal & Schraagen, 1988; Longoni & Scalisi, 1994). The visuospatial sketchpad also develops over time,
as children improve on recall of visual materials, and begin to rely more on the phonological loop to encode information (Hitch et al., 1988).

The decline of working memory in older adulthood has been examined using frameworks other than the Baddeley and Hitch model. Some of these experiments have looked at the performance of participants with decreased sensory acuity. Four groups of participants (young adult normal hearing, young adult hearing loss, older adult normal hearing, and older adult hearing loss) performed word and sentence recall tasks in an investigation by Gordon-Salant and Fitzgibbons (1997). For both word and sentence material, the young adult normal hearing group showed the highest recall performance, followed by the young adult hearing loss group, the older adult normal hearing group and finally, the older adult hearing loss group. Recall performance indicates that, while loss of sensory acuity affects working memory (young adult hearing loss group), age also has an impact.

Tasks requiring participants to answer questions about spoken discourse reveal younger adults achieve higher scores than older adults (Schneider, Daneman, Murphy, & See, 2000). This result was seen when discourse materials were presented in quiet and in noise, indicating noise negatively affects working memory for both groups of participants. Older adults’ poorer performance in quiet compared to young adults implies working memory declines play a part in poorer recall abilities.

Several reasons for the decline in working memory have been posited. Working memory declines could be caused by decreased comprehension of written conversation (Cohen, 1981; Light & Anderson, 1985; Splich, 1983) or decreased comprehension of spoken conversation in favorable listening conditions (Cohen, 1979, 1981; Light, Zelinski & Moore, 1982). Craik and
Byrd (1982) proposed that older adults have difficulty with working memory tasks because of a decline in cognitive processing resources. Decreased processing speed (Craik & Jennings, 1992; Salthouse, 1996), loss of ability to inhibit irrelevant information (Hasher & Zacks, 1988) and loss of peripheral sensory acuity (Baltes & Lindenberger, 1997; Salthouse, Hancock, Meinz, & Hambrick, 1996) have all been considered as causes of working memory decline in older adults.

The Present Study

The purpose of the present study was to investigate the nature of the relation between short-term working memory and recall accuracy in speech perception. This was examined at three different time points across the lifespan. Inhibitory control resulting in improved short-term memory performance and speech perception abilities may be a possible explanation for the development of abilities in children and later decline in older adults. As the ability to inhibit irrelevant information develops, the ability to select the salient information in the auditory signal may become more refined. Conceptually, this may be similar to the holes in a mesh screen closing, keeping out unwanted information and allowing in only desired information. Perhaps the four components (central executive, phonological loop, visuospatial sketchpad, and episodic buffer) in the multicomponent (Baddeley, 2000; Baddeley & Hitch, 1974) model of short-term memory are subject to the control of inhibition. It was hypothesized that young adults would demonstrate the fewest recall errors, followed by older adults and children.

The experimental task in the current study was a modification of a running memory task described by McCoy et al. (2005). Participants were presented with of strings of word stimuli in a running memory task. The words were presented in auditory-visual and auditory-only modalities. Participants were asked to repeat back the last 4 words heard at the end of the string.
The multicomponent model of short-term working memory was utilized as a theoretical framework (Baddeley, 2000; Baddeley & Hitch, 1974). The phonological loop was involved in creating a phonological representation of the auditory stimuli as well as rehearsing it, and the visuospatial sketchpad was responsible for maintaining and storing visual information about the speech cues presented by the talker’s face. The episodic buffer integrated information from the phonological loop and visuospatial sketchpad. Finally, the central executive was required to control attention during the task and manage the operation of the subcomponents.

The number and type of recall errors for the last four words heard in the string was examined. Lag 1 was defined as the last word heard in the string, lag 2 was the second back, lag 3 was the third word back and lag 4 was the fourth word back in the string. It was expected that participants would make the fewest number of recall errors for the last word heard (lag 1) because of the recency effect (Murdock, 1962). Correct recall of the last word heard also demonstrated that the stimuli were audible and intelligible. The number of recall errors for the other three target words was analyzed, in order to examine the effect of short-term working memory function. Words further back in the string would remain in working memory, requiring rehearsal and maintenance before recall. If the words further back in the string were incorrectly reported this would suggest a short-term working memory effect.

The young adult controls and older adults were expected to demonstrate fewer recall errors at all lag positions than the children, as adults have greater word knowledge and language experience (Miller, Stine-Morrow, Kirkorian, & Conroy, 2004). Across the adult participants, the younger adult controls were expected to have fewer recall errors than the older adults, as the cognitive load was expected to negatively affect older adults.
The stimuli were presented with and without the talker’s face present in order to test the hypotheses that visual cues mediate short-term working memory performance. Stimuli were presented with and without background noise to test the hypothesis that the presence of background noise would result in poorer short-term memory performance. It was expected that auditory-visual presentation would improve short-term memory performance by providing a visual representation the visuospatial sketchpad can then utilize. This would help to create a richer multi-dimensional representation in the episodic buffer when combined with information from the phonological loop. Fewer recall errors were expected in the quiet acoustic condition compared to conditions with noise, as noise interferes with the encoding and rehearsal of the desired signal.

Findings from the present study may provide data about the development of short-term memory for speech perception in children. The role of cognitive skills, such as short-term memory, has not been extensively examined in typically developing, normal hearing children (Choi, Lotto, Lewis, Hoover & Stelmachowicz, 2008). Examination of recall errors in different modalities and acoustic environments provides evidence for the use of strategies such as auditory-visual presentation to improve short-term working memory and speech perception. These strategies may be particularly helpful to those with immature or declining short-term memory skills. This would also provide evidence for multi-sensory presentation enhancing cognitive function. Increased signal redundancy provided by auditory-visual presentation may allow for greater use of context for word recognition. The results of the current investigation may serve to shape our understanding of the components of short-term memory and their function in speech perception. This also contributes to the understanding of both cognitive and peripheral functions and their role in an integrated understanding of speech perception.
CHAPTER 2

Literature Review

Speech perception is a multi-step process, from the peripheral auditory system to parts of
the cortex in the brain. Additional processes occur in the auditory and visual cortices, as auditory
information may be combined with visual information in the case of visible speech gestures.
Other processes also play a role in the perception of spoken speech, such as lexical
representation of speech, contextual knowledge, and long and short-term memory functions.

Role of Vision in Spoken Speech Perception

Vision plays an important role in the perception of speech. For example, McGurk and
MacDonald (1976) demonstrated that when participants were presented with the ambiguous
(information about the consonant is hidden) visual stimuli /ga ga/ paired with auditory stimuli /ba
ba/, the syllable perceived was /da da/. When presented with the highly visible visual stimuli /ba
ba/ paired with the acoustic stimuli /ga ga/, participants reported hearing a combination of both
consonants /gabga/. These data reveal visual cues can alter the perception of the auditory signal.

Presentation of visual information in addition to auditory information can also enhance
speech perception in unfavorable listening conditions, such as competing background noise.
Sumby and Pollack (1954) presented bisyllabic word stimuli to participants with and without
visual cues at 6 different signal-to-noise ratios. They concluded that the addition of visual
information to auditory presentation can result in up to a 15 dB improvement in signal-to-noise
ratio. Each 1 dB improvement in signal to noise ratio (SNR) can result in a 5-10% improvement
in speech intelligibility, depending on the speech materials utilized (Grant & Braida, 1991;
Miller, Heise & Lichten, 1951). Erber (1969) reported that word recognition scores of young adults improved from 20% when presented auditorily to 80% when presented auditorily and visually at -10 dB SNR.

It is important to note the testing procedures and the experimental task used when investigating of speech intelligibility when visual cues are present. Both Sumby and Pollack (1954) and Erber (1969) presented the word stimuli used in the experiment to the participants prior to testing. This prior exposure to the experimental stimuli may have increased the speech recognition scores because of speech reading, particularly at low signal-to-noise ratios (Ross, Saint-Armor, Leavitt, Javitt, & Fox, 2007).

Classic behavioral speech perception studies (Erber, 1969, 1975; Sumby & Pollack, 1954) and more recent objective studies illustrate speech intelligibility in background noise may be improved with visual cues present. These findings provide support for the principal of inverse effectiveness. The principal of inverse effectiveness was the product of a series of experiments examining multi-sensory integration done in the superior colliculus of cats by Meredith and Stein (1986). Their findings indicated multi-sensory enhancement was greatest when input from a single sense was weakest. Experiments such as Sumby and Pollack (1954) and Erber (1969) lend support for the principal of inverse effectiveness from a behavioral standpoint, in that they report the amount of enhancement from the addition of visual cues was inversely related to the signal-to-noise ratio presented.

Speech perception can also be enhanced by visual cues when speech is clear and audible, while the stimuli may be challenging to the listener for reasons other than noise. Arnold and Hill (2001) investigated the effect of auditory-visual presentation on speech comprehension in native
English speaking adults in 3 experiments. In the first experiment spoken passages were read in the second language the participants had learned (French). The participants in the second experiment were presented with passages spoken by a talker with a strong Glaswegian (Scottish) accent. In experiment three, the passages contained semantically and syntactically complex material. In all three experiments, a total of six passages were presented; a) three in the auditory-only modality; b) three in the auditory-visual modality. Participants were instructed to repeat back the passages and answer comprehension questions about them. Auditory-visual presentation resulted in a 15% improvement in speech tracking rate compared to auditory only presentation. On average, participants correctly answered two to three more comprehension questions correctly when the passages were presented in the auditory-visual condition compared to auditory only presentation. These findings reveal an advantage of auditory-visual presentation compared to auditory only presentation, not only in background noise, but also when the auditory signal is clear and audible.

**Lexical Representation of Speech**

In addition to the integration of visual information, speech perception requires the linguistic information (acoustic, phonetic, lexical, semantic, and syntactic) to be projected onto a mental model. Lexical representations provide a link between sound information and meaning by combining the phonological attributes of words with the semantic and syntactic attributes (Marslen-Wilson, Brown & Tyler, 1988). They also give the linguistic representation of the sound information a foundation. The structure of the linguistic representation can be mapped onto the mental model, allowing the information to be interpreted (Marslen-Wilson et al., 1988).
In order to investigate how quickly different lexical representations could affect processing, Marslen-Wilson et al. (1988) had participants complete a sentence monitoring task. They were instructed to press a button when they detected the target verb in a pair of sentences. The stimuli varied the relationship between the noun and the target verb in order to create sentences which violated the pragmatic, semantic, and categorical constraints derived from the argument frames connected with transitive and intransitive verbs. The sentences with pragmatic violations were possible, but not plausible in real life situations (e.g., “The boy buried the guitar.”). The semantic properties of the verb were not compatible with the semantic properties of the noun in the sentences with semantic violations (e.g., “The boy drank the guitar.”). Sentences with categorical violations contained an intransitive verb that could not be followed by a noun as a direct object (e.g., “The boy slept the guitar.”).

The results revealed response time was longest for categorical violations, followed by semantic and pragmatic violations. The long response time for categorical violations indicates the participant had difficulty interpreting lexical items that were not structurally interpretable. Semantic violations caused difficulty in interpreting the possible meaning of the sentence. Pragmatic violations were less disruptive because they are more plausible than semantic violations. Marslen-Wilson et al. (1988) concluded that these data provide evidence for immediate effects of these violations at the linguistic level and also to the understanding and analysis required to build the mental model of the utterance. In order to recognize the types of violations, the participant must utilize short-term memory to encode and recall the material presented.
Contextual Knowledge of Speech

Listeners also use previous contextual knowledge when building a mental model of speech stimuli. Phonological, semantic and pragmatic contextual knowledge may be used when constructing a mental model. These types of contextual knowledge can improve word and sentence recognition because they influence the probability of the stimulus pattern being recognized by the listener (Boothroyd & Nittrouer, 1988). As the probability of prior contextual knowledge of the stimuli increases, the probability of correct recognition also increases. Previous investigations have attempted to examine context effects through probability theory (Schiavetti, Sitler, Metz & Houde, 1984), utilizing one equation to relate the probability of speech units with and without context, and a second equation to relate the probability of recognizing the whole to the probabilities of the parts that make up the whole.

Boothroyd and Nittrouer (1988) developed a metric called the \( j \) factor to examine the effect of linguistic context on the correct recognition of speech units. The \( j \) factor score is a ratio of two recognition probabilities: the probability of correctly recognizing the whole word and the probability of recognizing the parts (phonemes) within the whole word. The \( j \) factor score is expressed as \( j = \log (P_w)/\log (P_p) \), where \( P_w \) represents the probability of recognizing the whole word, \( P_p \) represents the probability of recognizing a part of the whole word (phoneme), and \( j \) represents the number of independent channels of information needed for recognition. The number of independent channels needed to recognize the whole word can be from 1 to \( n \), where \( n \) represents the number of phonemes in the whole word. If there is no effect of context, the \( j \) factor score would equal the number of units that make up the whole. A smaller \( j \) factor score indicates a greater effect of lexical context on the recognition of the whole word.
Previous investigations have utilized the j factor to examine the effect of context on phoneme and syllable recognition for consonant-vowel-consonant words and non-sense words in children, young adults and older adults (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990). The findings from two studies (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990) showed children and older adults had higher j factor score compared to young adults. The authors speculated that a possible explanation for these results could be that children may be developing word knowledge compared to adults, and as a result may be less able to use semantic and phonetic context for word recognition. Older adults may have shown higher j factor scores for nonsense words compared to young adults because they impose meaning on word stimuli.

**Short-Term Memory: The Baddeley & Hitch Model**

Building a mental model of speech requires short-term working memory in order for the listener to encode, manipulate, and maintain a representation of the utterance. Short-term memory refers to memory for very recent events, where seconds or minutes separate the presentation of a stimulus to be remembered and its recall (Gathercole, 1998). Working memory is a subcomponent of short-term memory, where the manipulation and maintenance of information for recent events occurs. Running memory refers to a monitoring task in which the participant does not know the length of the task, and the participant must maintain old information and capture new information (Pollack, Johnson & Knaff, 1959). Over time old information is dropped from running memory maintenance and rehearsal to allow more recent information to be taken in. There are several models of short-term memory. Figure 1 illustrates the four component model and its subcomponents.
Functions of the central executive. The three component working memory model proposed by Baddeley and Hitch (1974) includes the central executive, the phonological loop, and the visuospatial sketchpad. Recently, the model has been revised to include a fourth component, the episodic buffer (Baddeley, 2000). The central executive is at the core of the model and is responsible for attention, control of attention, and problem solving (Baddeley, 1986). It functions as a coordinator of operations within working memory, retrieving information from long-term memory, and applying retrieval strategies. The central executive also is part of logical reasoning and mental arithmetic (Baddeley, 1986; Hitch, 1980).

The central executive is responsible for many functions. The function of control of attention can be separated into four capabilities: focus, divide and switch attention, and relate information in short-term memory to information in long-term memory (Baddeley, 1996). Evidence that the central executive has the capacity to focus attention comes from many task situations. Robbins et al. (1996) showed that playing chess is disrupted by a simultaneous visuospatial task and by a random digit generation task; however articulatory suppression (repetition of an irrelevant word) does not disrupt play. Category generation tasks (Baddeley, 1966c) and mental arithmetic (Logie, Gilhooly, & Wynn, 1994) are also disrupted by random digit generation tasks. These data reveal that the central executive plays a role in many cognitive tasks where focused attention is necessary.

The capacity to divide attention has been revealed through experiments with older adults with Alzheimer’s disease, who are presumed to have impairment of the central executive (Baddeley, Bressi, Della-Salla, Logie & Spinnler, 1991) due to their deficits in episodic long-term memory and attentional deficits (Perry & Hodges, 1999). The older adults were given a
dual task, requiring the repetition of strings of digits based on individual’s digit span (utilizing the phonological loop) and using a light sensitive pen to track a dot as it moves randomly on a screen (utilizing the visuospatial sketchpad). Task difficulty was manipulated so performance on the single task was at the same level for the older adults with Alzheimer’s and younger participants, who served as a control group. When the level of difficulty on one of the tasks was manipulated, the performance of the older adults with Alzheimer’s disease was not affected; however, dual task performance declined (Logie, Della-Salla, Wynn & Baddeley, 2000).

Baddeley, Chincotta, and Adlam (2001) utilized a dual task paradigm that incorporated articulatory suppression (repeating months of the year or days of the week) and the central executive (mental arithmetic), revealing that the central executive plays a role in attention switching. The results indicated that in some test conditions the effect of articulatory suppression (repetition of irrelevant words) was significant on attention switching; therefore the phonological loop may facilitate verbally-based tasks. Saeki and Saito (2004) replicated these results and also stated switch costs (poorer performance on task-switch trials compared to same-task trials) increased significantly if the dual-task was mental arithmetic and articulatory suppression (repetition of irrelevant words), but not with concurrent finger tapping. These results indicate that the process of switching attention may depend on many components of short-term memory. A multi-component model of task switching has been suggested by several investigators (De Jong 2000; Goschke, 2000; Rubenstein, Meyer & Evans, 2001; Saeki & Saito, 2004) and two capacities are presumed: maintenance of a task switching system which is a function of the phonological loop, and ability to activate the appropriate task which is attributed to the central executive.
The fourth capacity of the central executive is the ability to connect information in working memory to information in long-term memory (Baddeley, 1996). This capacity has been proposed to be the function of the episodic buffer, a recent addition to the working memory model. The episodic buffer is proposed to function as a go-between for working memory and long-term memory. It also integrates information from the phonological loop and the visuospatial sketchpad in a multi-dimensional code.

**Functions of the phonological loop.** The phonological loop is a slave system of the central executive and is composed of two subcomponents, the phonological store and the subvocal rehearsal process (Baddeley, 1986). The phonological store holds information in a phonological code; however, this is subject to quick decay, about 2 seconds, if not rehearsed (Baddeley, Thomson & Buchanan, 1975). Information enters the phonological store directly through auditory presentation of stimuli and it can also enter indirectly through non-auditory stimuli such as printed words, familiar visual objects, or phonetic gestures (lip reading). The indirect route generates an internal phonological code (Gathercole, 1998). The subvocal rehearsal process revives decaying information in the phonological store. It also serves as a gateway for information to the phonological store, recoding non-phonological information (e.g. pictures, printed words) into a phonological form (Gathercole, Pickering, Ambridge & Wearing, 2004).

The phonological store is sensitive to the internal phonologic code. Evidence for this internal code comes from the phonological similarity effect. Conrad and Hull (1964) demonstrated that recall of a sequence of phonemes that are similar sounding (e.g., /b/, /p/, /g/) is more difficult than recalling a dissimilar sequence of phonemes (e.g., /k/, /s/, /ʒ/) when the
phonemes are presented auditorily or visually. Similarity of sound has been shown to affect the number of words recalled; however, similarity of word meaning has little effect on recall (Baddeley, 1966a).

The word length effect provides evidence that subvocal rehearsal maintains information in the phonological store. Baddeley et al. (1975) showed more accurate recall of a sequence of words that were short in length (e.g., sat, sun, bat) compared to a sequence of longer words (e.g., banana, telephone, piano). This result was explained by the idea that shorter words require less rehearsal time, and therefore the memory trace has less time to decay compared to longer words which take more time to rehearse. An alternative explanation for this result is that word complexity affects recall rather than word length (Caplan, Rochon, & Waters, 1992; Service, 1998, 2000). Articulatory suppression, repeatedly articulating an irrelevant word during a word memory task, eliminates the word length effect (Baddeley, Lewis, & Vallar, 1984). This finding is evidence that subvocal rehearsal refreshes memory traces in the phonological store. Articulatory suppression greatly impairs the recall of words rather than eradicating it, suggesting that verbal information may also be stored in other ways. The episodic buffer is a possible way that verbal information might be stored during articulatory suppression (Repovš & Baddeley, 2006).

Functions of the visuospatial sketchpad. The visuospatial sketchpad is the component capable of maintaining and manipulating physical features and dimensions of events, like shape, color, and movement (Gathercole, 1998). It can further be separated into visual and spatial subcomponents, each with its own storage, maintenance and manipulation procedures (Repovš & Baddeley, 2006). The visual store subcomponent is where the physical characteristics of objects
and events are represented and the spatial mechanism subcomponent is used for planning movements (Logie, 1994). Evidence for 2 subcomponents of the visuospatial sketchpad comes from Della-Salla, Gray, Baddeley, Allamano and Wilson (1999). A spatial interference task (arranging pegs on a screen), the Corsi block tapping task (tapping out sequences on block arranged on a board), and a visual pattern task (re-creates a previously seen block pattern in a matrix) were utilized. Results indicated the spatial interference task interrupted performance on the spatial working memory task (Corsi block tapping), and the visual pattern task interrupted performance for visual working memory.

This result was later replicated through a series of experiments by Klauer and Zhao (2004), utilizing four tasks: a spatial memory task (memorizing the location of a dot on a screen), a visual memory task (memorize one of eight simple geometric forms), a visual interference task (discriminate colors red from blue) and a spatial interference task (identify one stationary target on a screen of 15 other moving targets). Data from these experiments provided further evidence for the division of the visual and spatial subcomponents. The authors also provide data to show the visuospatial rehearsal mechanisms are detached from the central executive. A random interval repetition task ([RIR]; Vandierendonck, De Vooght, & Van der Gooten, 1998) was added to engage the central executive without affecting the visuospatial sketchpad. The RIR task taps into the control of attention function of the central executive.

The Baddeley and Hitch model of short-term memory (Baddeley, 2000; Baddeley & Hitch, 1974) is able to explain how many types of auditory and visual information are processed and manipulated. The central executive is responsible for the control of attention, as well as overseeing the coordination of information from the slave systems. The phonological loop
encodes both auditory and visual information to form a phonologic code. The visuospatial sketchpad encodes a variety of dimensions of visual stimuli.

**The visuospatial sketchpad and visual cues in spoken speech.** Much of the evidence for the function of the visuospatial sketchpad comes from experiments utilizing pictures or objects rather than human faces. There are data to indicate that visual cues from human “talking faces” can be encoded and utilized in the same manner as auditory cues. The recency effect (improved recall for the final item in a list) and the decrease in the recency effect when a suffix is presented are two attributes of auditory short-term memory, and were hypothesized to be unique to auditory presentation (Crowder & Morton, 1969; Gardiner, 1983; Watkins & Watkins, 1980). These effects have been shown to also be present for stimuli presented in the auditory-visual and visual only (i.e., lip-read) conditions.

The first piece of evidence for the similarities between memory of acoustic lexical stimuli and lip read lexical stimuli came from Spoehr and Corin (1978). Lists of single syllable digits were presented in the auditory-visual condition in a recall task. At the end of the list either an auditory or visual suffix was presented. Results showed the recency effect was disturbed equally by an auditory suffix than by a visual suffix. Campbell and Dodd (1980) presented lists in the visual only modality (lip reading) in a recall task with a visual suffix. Data indicated the recency effect was disturbed by the visual suffix. These results provide evidence that when presented with a “talking face” the visual cues provided can be encoded and recalled when stimuli are presented in the auditory-visual or visual only conditions. It also strengthens the idea that the subcomponents of the Baddeley and Hitch short-term memory model (i.e., the visuospatial sketchpad and phonological loop) work together as an overall system of memory.
There is also objective evidence from neuroimaging studies illustrating the cross-modal effects of auditory-visual presentation. According to Calvert (2001) cross-modal (i.e., audition and vision) refers to tasks in which stimuli presented involve more than one sense. Activation of the auditory cortex was observed during a lip reading task in a study utilizing functional magnetic resonance imaging (fMRI), (Calvert, et al., 1997), and magnetoencephalography (MEG) (Sams, et al., 1991). A subsequent fMRI study by Calvert, Brammer, Bullmore, Campbell, Iverson, and David (1999) revealed that when stimuli were presented in the auditory-visual condition activation was observed in both cortices. In addition, the amount of activation in the auditory-visual condition was greater than that observed in the auditory-only or vision-only conditions, indicating cross-modal enhancement.

**Functions of the episodic buffer.** The central executive, phonological loop, and visuospatial sketchpad are useful in the explanation and interpretation of a great deal of data regarding working memory, however, the Baddeley and Hitch model is not without limitations. There are some phenomena the three component model is unable to account for. For example, the effect of articulatory suppression (i.e., verbal repetition of an irrelevant word while attempting to repeat back a sequence of numbers presented visually) would be expected to result in difficulty recalling the numbers because repetition of the irrelevant word suppresses the entry of the visual information in the phonological loop according to the model. While articulatory suppression does significantly decrease recall, the reduction in auditory memory is less than the model would predict, decreasing from about seven digits to about five digits (Baddeley, Lewis, & Vallar, 1984).
In patients with severely impaired phonological memory, visual presentation of digits can improve auditory memory (Baddeley, Vallar, & Wilson, 1987). This result is puzzling because, if phonological memory is impaired, the storage mechanism for the digits is unclear. One possibility is the visuospatial sketchpad; however, it is not well suited for serial recall tasks (Phillips & Christie, 1977). If the information was coded visually, it would be expected that articulatory suppression would make recall sensitive to visual similarity effects. This hypothesis has been investigated by Logie, Della Salla, Wynn and Baddeley (2000) and although visual similarity effects were found, they were small and were found in both suppressed and non-suppressed conditions. The effect of visual similarity in the non-suppressed conditions indicates that the phonological and visual information are somehow combined; however, the three component working memory model does not have a means of accounting for this finding. The central executive does not have the capability to store such information in the three component model. Page and Norris (1998) proposed a “back-up” store, capable of serial recall and of combining phonological and visual information.

Data from prose recall also highlights limitations of the phonological loop subcomponent in the working memory model. Participants will typically begin to make errors in recalling a sequence of unrelated words after the length of the sequence reaches five to six words, however if the words make up a meaningful sentence recall of 16 words or more has been recorded (Baddeley, Vallar & Wilson, 1987). This increase in the number of words recalled is referred to as “chunking” (Miller, 1956) in which information from long-term memory is utilized to integrate the sequence of words into smaller chunks. The capacity for recall is then set by the number of chunks instead of the total number of words. These data again raise the question of how the information from long-term memory and other sources are combined, in the
phonological loop or by some other means. Findings from a patient whose long-term memory was normal had a word recall span of one, was shown to have a sentence recall of five words (Vallar & Baddeley, 1984). This finding indicates that long-term memory is not solely responsible for chunking, and perhaps there is some back-up store or an interface between the phonological loop and long-term memory.

The data from articulatory suppression, prose recall, and from patients with short-term memory deficits point out limitations within the three component model of working memory. The three component model does not provide a means of combining information from the subcomponents (the phonological loop and visuospatial sketchpad) and long-term memory stores in order to form a temporary representation. The episodic buffer has been proposed as a fourth component in the working memory model as a means of addressing these issues.

The episodic buffer is a subcomponent of the central executive, similar to the phonological loop and visuospatial sketchpad. It is a limited-capacity system, able to store information in a multi-dimensional code (Baddeley, 2000), and functions as an interface between the phonological loop and the visuospatial sketchpad. This subcomponent is episodic because it holds information across episodes in space and perhaps time and acts as a buffer between subcomponents, each with their own codes (Baddeley, 2000). It is hypothesized to be limited in capacity because of the great demand placed on it in providing access to a large variety of codes (Hummel, 1999). Through conscious awareness, the central executive can access the episodic buffer and restrict the information in it by focusing attention on a source of information. The source of information could be the other subcomponents of working memory, long-term memory, or a perceptual source.
The entire short-term memory system could be conceptualized as a dynamic mesh, responsible for allowing information to enter working memory where it is encoded, manipulated, and maintained by the subcomponents. The system functions to screen irrelevant information and sort the information in the system into the correct component to process it. The components work together to transform information, connect it to information in long-term memory when appropriate, and to facilitate formulating responses to stimuli in real time.

**Alternative Models of Short-Term Memory**

**Capacity Theory.** Just and Carpenter (1992) proposed another model of working memory, capacity theory. In this model the processing and storage of information is determined by activation that can be distributed over cognitive tasks. The capacity for activation is variable from person to person. When all working memory capacity has been distributed, the level of activation must be decreased before any new task of processing or storage can occur. This model of working memory differs from that of Baddeley and Hitch (Baddeley, 2000; Baddeley & Hitch, 1974) in that it does not propose a system with distinct components.

**Inhibition.** The control of access to activated information is the focus of another model of working memory proposed by Hasher and Zacks (1988). The central feature of this model is the inhibition of tasks or actions that are not related to the goal at hand. Inhibition is in place during the encoding of stimuli, in order to activate specific parts of working memory and also during retrieval in order to determine the parts of long-term memory searched.

Data detailing the declines older adults experience in the areas of visual selective attention (Madden, 1983) memory (Winthrope & Rabbitt, 1988) and language (Gold, Andres,
Arbuckle, & Schwartzman, 1988) may be explained by the hypothesis that loss of inhibition for activation may be the cause for these problems. When inhibition is reduced, fewer stimuli will be excluded from the memory representation and activation at the time of input. Irrelevant stimuli will be allowed to be activated, and will receive more processing resources when inhibition is less effective. Evidence for this comes from data showing older adults are more likely to interpret meanings of reading material that are not pivotal to the overall meaning compared to young adults (Hamm & Hasher, 1992). Older adults are also more likely to retain interpretations that have been disproved or replaced by new information (Hamm & Hasher, 1992; Hartman & Hasher, 1991). Reduced inhibition will also allow irrelevant pathways to be activated at the time of retrieval (Hasher, Soltzfus, Zacks, & Rypma, 1991).

**Total processing space.** Case, Kurland, and Goldberg (1982) investigated the hypothesis that cognitive capacity might increase with age in children. Utilizing a counting span task, results showed counting span increased as age increased, and those who counted faster had a higher counting span. The framework suggested for this result was the concept of total processing space as a representation of cognitive capacity. The total processing space is composed of the operating space, which is required for the counting operation during the task, and the short term storage space, which is the residual space left over for storing the results. There is a tradeoff between processing (operating space) and storage (short term storage space) activities which compete for resources from a single, limited cognitive source (total processing space). It was suggested that the total processing space remains constant across age, while the age related increase in counting span was due to improved efficiency of the counting operation (operating space). This increase in efficiency results in lower demand on the resources from the total processing space, allowing
for more resources to be used by the short term storage space, and therefore a longer counting span.

**Time based models.** While other models include the role of time and memory decay, Towse and Hitch (1995; Towse, Hitch & Hutton, 2000) provided evidence that working memory span in adults and children was primarily dependent on processing time rather than resource limitation. The strength of the memory trace in the short term storage space might decrease as the interval between storage and recall increased. Older children might have longer counting spans because they count faster. The increased speed therefore decreases the amount of time the information has to be maintained. This has been called the memory decay hypothesis. It eliminates the operating space and the notion of sharing total processing space for processing and storage in order to account for the difficulty of counting span tasks.

In order to evaluate the memory decay hypothesis, a counting task was utilized. Task difficulty was manipulated; however, the time allowed for executing the task was kept constant. Results indicated the ability to store count totals did not reflect the amount of working space operations, but rather the time period over which the count totals may be forgotten. This result was observed in both young adults and children six, seven, eight and ten years of age. Towse and Hitch (1995) concluded that children may instead switch off between storage and processing operations instead of combining the two. Attentional resources are shifted between processing and storage functions and items in short-term working memory are subject to time related decay.

Improved efficiency alone for cognitive processes such as maintenance and recall may not fully explain the developments seen in children’s short-term working memory abilities. Differences in storage abilities may help to explain the variance in performance observed on
span tasks (Bayliss, Jarrold, Gunn, & Baddeley, 2003). Storage abilities primarily reflect differences in the refreshment of memory traces (Barrioullet, Bernadin, & Camos, 2004). This concept is part of the time based resource sharing (TBRS) model. The TBRS model states that regardless of task demands, continuous attention is rarely required. This allows attention to be switched to other thoughts (Barrouillet & Camos, 2001). The TBRS model suggests refreshment of memory traces takes place during very short pauses between processing and maintenance functions (Barrouillet, Vergauwe, Gaillard, Gavens, Gaillard, & Camos, 2009). Stronger memory traces lead to longer span performance observed as children age.

**Limited Capacity Models.** A limited capacity model of working memory called the effortfulness hypothesis was proposed by Rabbitt (1968). The effortfulness hypothesis states that the amount of perceptual effort a participant must exert in a situation affects other resources such as working memory. In two experiments Rabbitt (1968) showed that young adults with normal hearing had greater recall accuracy for strings of spoken digits presented in quiet than in noise. This was the case for a string of digits presented consecutively and for a string a divided by a silent pause. Rabbitt concluded that, when the digits were presented in noise, the listener exerted more perceptual effort in order to identify the digits; therefore, fewer processing resources were available for rehearsal of the digits in order to encode them in memory. Increased perceptual effort resulted in a negative impact on memory recall.

In a more recent experiment, Rabbitt (1991) tested the generalizability of the effortfulness hypothesis to the recall of words. Two groups of older adults were asked to repeat and later recall word lists. One group of participants had normal hearing and the other had mild hearing loss. Both groups were able to correctly repeat back the word lists at the same intensity
level, indicating that the speech was intelligible. Listeners with mild hearing loss showed poorer recall than listeners without hearing loss. These results support the conclusion that adults with hearing loss have to allocate more processing resources for correct identification of the word stimuli, and fewer processing resources are available to support rehearsal for encoding the words into memory.

The effect of perceptual effort on working memory resources was investigated by McCoy, Tun, Cox, Colangelo, Stewart, and Wingfield, (2005). The authors explored the possibility that Rabbitt’s (1991) findings for poorer word recall in older adult listeners with hearing loss is the result of decreased hearing acuity rather than increased perceptual effort. This means that because of peripheral hearing loss, the word stimuli are never correctly identified, and therefore cannot be rehearsed or encoded into memory.

McCoy et al., (2005) examined the effects of contextual constraint in English and perceptual effort on word recall. The amount of contextual constraint is determined by the order of approximation. For example, a first order approximation to connect speech is based on a single word out of context (e.g., “realizing most so the together home and for were wanted”), while the third order approximation in based on three words out of context (“family was large dark animal came roaring down the middle of my friends love books”). Stimuli were presented as word strings that differed in length and connected context (contextual constraint). A total of four sets of words were evaluated for which contextual constraints were manipulated. Participants completed a running memory task and were instructed to repeat back the last three words they heard, regardless of word string length. It would be expected that a word with a higher level of contextual constraint (e.g., third order approximation) would require less
perceptual effort to encode because it is less ambiguous than a word with lower constraint (e.g., first order approximation).

A group of older adults with normal hearing and a group with hearing loss participated. The authors hypothesized that participants with hearing loss would have to exert more perceptual effort than those without hearing loss for words with low constraint because they are more difficult to encode. McCoy et al., (2005) also wanted to test the assumption that a larger number of errors made by those with hearing loss were the result of increased perceptual processing demands. Both groups of participants were expected to be able correctly identify the words, as evidenced by correct recall of the final word in the 3-word set. The group without hearing loss would be able to correctly recall the first two words of the 3-word set with greater accuracy than the group with hearing loss.

Results showed that the final word of the 3-word set was recalled with high accuracy, 99.5% for the normal hearing participants and 98.2% for those with hearing loss. Accurate recall of the final word shows both groups could understand the words so they could repeat them back. This result also showed that the first two words of the set had been heard. Both the normal hearing and hearing loss groups had recall scores near ceiling levels for the first word of the 3-word set. Participants with hearing loss had significantly poorer recall scores for the second and third words in the 3-word set than those with normal hearing. McCoy et al., (2005) concluded that the additional perceptual effort needed for successful recognition of words in the presence of a mild to moderate hearing loss was sufficient to affect memory performance. Greater recall accuracy by the normal hearing group for the first and second word shows less perceptual effort was needed for correct identification of the final word in the set. Therefore, more resources were
available to those with normal hearing for encoding and rehearsal of the first two words compared to those with hearing loss.

Participants with hearing loss made more errors on the low contextual constraint condition; however, there was no difference in recall accuracy for the third word in the 3-word set between the participants with hearing loss and those without hearing loss. There was, however, a significant difference between the two groups for accurate recall of the first and second words in the set. Although both groups of listeners were able to identify the words, the consequence of the added perceptual effort exerted by those with hearing loss resulted in fewer processing resources available for working memory.

**Hybrid model.** The Ease of Language Understanding (ELU) model is based upon evidence that working memory for language processing is independent of the modality in which the language input is perceived, represented, and produced (Rönnberg, 2003a; Rönnberg, Rudner, Lunner & Zekveld, 2010). Its framework contains elements based on processing speed and capacity; however, it is unique in its conception of how cognitive resources are deployed when the language input is optimum or degraded. The model stems from evidence provided by experiments examining the neural correlates of speech understanding and cognitive performance (Rönnberg, 1995; Rönnberg et al., 2003b). These data revealed measures shown to be direct and indirect predictors of speech understanding performance. Verbal inference making (Lyxell & Rönnberg, 1989), word decoding (Lyxell & Rönnberg, 1991) and speed of lexical access (Larsby, Hälgren, & Lyxell, 2008; Rönnberg, et al., 1998) have been shown to be direct predictors of speech understanding, while visual evoked potentials, complex working memory span and verbal abilities (Lyxell & Rönnberg, 1989) are indirect predictors.
The ELU model is also based on evidence that working memory resources are general in nature across sensory and language modalities (Rönnberg, 2003a). This is supported by examination of visual-tactile speech tracking abilities (Andersson, Lyxell, Rönnberg, & Spens, 2001) and auditory-visual speech understanding performance in cochlear implant users (Lyxell, et al., 2008). These investigations revealed similar cognitive components were involved regardless of sensory modality used to present experimental stimuli. Further support for modality non-specific cognitive resources comes from imaging studies showing activation of the auditory cortex during lip reading (Calvert et al., 1997), shared variance between text-reception thresholds (Zekveld, George, Kramer, Goverts, & Houtgast, 2007) and speech reception thresholds.

Multi-modal language input enters the ELU model where rapid, automatic, multi-modality binding of phonology (RAMBPHO) takes place (Rönnberg, Rudner, Foo, & Lunner, 2008). The semantic, prosodic, linguistic, phonetic and syntactic information contained in the language input is bound together at this stage. RAMBPHO integrates multi-sensory language input with information in long term memory, similar to the function of the episodic buffer in the Baddeley and Hitch (Baddeley, 2000; Repvos & Baddeley, 2006) model. Implicit processing during RAMBPHO is determined by the clarity of the phonological information in working memory, the speed at which the representation can be processed, and the capacity available for processing. In quiet listening situations, RAMBPHO facilitates implicit access to the lexicon and connects the language input to phonological representations in long-term memory (Rönnberg, Rudner & Foo, 2008), ultimately resulting in language understanding.

If the language input provided is not clear due to signal degradation (e.g., hearing loss or background noise) (Näätänen, Pakarinen, Rinne, & Takegata, 2004), slowed processing speed,
reduced capacity for processing, or degraded phonological representation in long term memory (Andersson, Lyxell, Rönnberg, & Spens, 2001), a mismatch occurs. This mismatch results in explicit, effortful processing. Explicit storage and processing resources are employed. Resources specific to speech and/or visual information are also utilized during explicit processing.

The ELU model differs from other working memory models. It places greater emphasis on communicative skill compared to capacity theory (Daneman & Merikle, 1996) and, unlike the multi-component model proposed by Baddeley and Hitch (Baddeley & Hitch, 1974; Baddeley, 2000; Baddeley & Repvos, 2006) it suggests different roles for implicit and explicit processing of information. The explicit capacity component of the ELU model does share some similarities with the central executive in the multi-component model (Baddeley, 2000) and the suggestion of a supervisory attention system (Shallice, 2004).

**Developmental Changes in Short-Term Memory**

Models of working memory have also been applied to data gathered from children in order to understand and explain how short-term memory develops with age. The Baddeley and Hitch model is one which has been used to account for data regarding the development of short-term memory in children (Baddeley, 2000; Baddeley & Hitch, 1974). Performance on tests of phonological memory, such as digit span, improves considerably from the early to middle childhood years (Gathercole, 1998). For example, the maximum number of unrelated verbal items that can be recalled correctly in sequence increases from about two to three items at age four, and about six items at age 12 (Hulme, Muir, Thompson, & Lawrence, 1984).
**Development of the central executive.** There is some evidence to indicate the capacity of the central executive for complex working memory functions increases through childhood. Development of the central executive has been measured using span tasks, such as listening span. Children showed improved performance on listening span tasks from ages 6 to 15 (Siegel, 1994). These data could be accounted for by increased capacity of the central executive; however, they could also be interpreted in terms of the Just and Carpenter (1992) capacity model, stating that the amount of activation available increases as the child ages. Another possible interpretation is that, as the child ages, working memory becomes more efficient therefore, processing demands are reduced and the system can accomplish a greater number of tasks with the same number of resources available (Case, Kurland, & Goldberg, 1982).

Imaging studies have provided evidence that the frontal and pre-frontal cortices of the brain also develop, becoming more mylenated through childhood into late adolescence and early adulthood. Development of these cortices is important because functional imaging studies (Casey et al., 1997; Rubia et al., 2001; Rubia, Smith, Brammer, & Taylor, 2003) and lesion studies (Rakic, Bourgeois, & Goldman-Rakic, 1994) have found executive functions rely on the frontal lobes. Executive functions include inhibition of irrelevant stimuli, selective-attention, decision-making skills, and working memory (Blakemore & Choudhury, 2006). Sowell, Thompson, Tessner and Toga, (2001) conducted a magnetic resonance imaging study (MRI) of participants ages 7 to 30 and concluded that white matter increased in the frontal lobe through childhood and this increase accelerated during the early 20’s up until age 30.

**Development of the phonological loop.** There is evidence to suggest that the phonological store component of the phonological loop is present in young children, however,
the subvocal rehearsal subcomponent of the phonological loop does not surface until about 7 years of age (Gathercole & Hitch, 1993). Performance on serial recall tasks from children 3 to 5 years of age has been shown to be sensitive to phonological similarity and word length effects, as long as the test items are presented auditorily and do not need to access the phonological loop via subvocal rehearsal (Ford & Silber, 1994; Gathercole & Adams, 1994; Hitch & Halliday, 1983; Hulme, Muir, Thompson, & Lawrence, 1984). Instead of recoding picture stimuli into phonological code in order for the information to access the phonological loop, children younger than 7 years of age appear to remember such stimuli based on their visual properties (Hitch, Halliday, Schaaftal & Schraagen, 1988; Longoni & Scalisi, 1994).

**Development of the visuospatial sketchpad.** Changes also take place in the visuospatial sketchpad subcomponent of the 4 component working memory model. As age increases, children rely less on the visuospatial sketchpad and instead rely more on the phonological loop for immediate memory of visual material. Older children are able to recode visual information into phonological form to utilize the phonological loop for a visual memory task (Hitch et al., 1988; Hitch, Wooden, & Baker, 1989). Evidence for this comes from Hitch et al., (1988) where the recall of lists of words with similar features (e.g., fork, comb, pen) was compared to the recall of lists of words with few similar features (e.g., doll, glove, bath). Results indicated that 5 year olds are more reliant on the visuospatial sketchpad for recall of visual material compared to 10 year olds or adults. The 10 year olds had poorer recall performance for pictures with longer names (e.g., umbrella, kangaroo, banana) compared to adults, indicating performance continues to improve as age increases.
Short-term Memory and Word Recognition in Children

Choi, Lotto, Lewis, Hoover & Stelmachowicz (2008), examined the short-term memory abilities of normal hearing children for word stimuli presented with background noise. Normal hearing children ages 7-14 participated. The children were asked to perform two single tasks first. They were asked to repeat monosyllabic words from the phonetically balanced kindergarten (PBK) list three, presented with speech shaped noise at a signal-to-noise ratio (SNR) of +8 dB. Then in a single task of serial recall, five digits were visually presented on a computer screen. After ten seconds the children were asked to repeat the digits. If the child was not able to recall five digits with 100% accuracy, they were then given a three digit recall task.

A dual task paradigm was also utilized. First, the children saw a string of either three or five digits on a computer screen. Then, five randomly selected words from the (PBK) lists one and four were presented with speech shaped noise at an SNR of +8 dB. The children were instructed to repeat back the word immediately after it was presented. When all five words had been presented, the children were asked to repeat back the digits that had been shown on the screen at the beginning of the trial.

The results indicated that children showed poorer digit recall accuracy in the dual task paradigm compared to the serial recall task; however this was not seen for word recall in noise. The difference in accuracy on the serial recall and dual task recall of three digits became smaller as age increased, although this result was not seen for five digits. These data provide evidence that limited short-term memory capacity might result in children’s poor speech perception in background noise, and that this improves with age.
Cognitive Changes in Older Adulthood

There is considerable evidence from behavioral and imaging studies documenting structural and functional changes associated with aging. Behavioral evidence reveals that some cognitive functions such as verbal knowledge, mainly vocabulary (Park, Lautenschlager, Hedden, Davidson, Smith & Smith, 2002) and procedural memory (Howard, Howard, Dennis, LaVine, & Valentino, 2008), are not greatly impacted by aging. Memory functions that rely on familiarity also do not show large declines with age (MacDaniel, Einstein, & Jacoby, 2008).

Declines in cognitive processes such as increased processing and reaction time are seen on a variety of tasks (Salthouse, 1996). Loss of sensory acuity has been associated with a decline in cognitive performance (Baltes & Lindenberger, 1997; Lindenberger, Scherer, & Baltes, 2001). Older adults with reduced visual and auditory acuity show lower performance on cognitive tasks. This suggests a dedifferentiation of cognitive performance with age caused by the lower quality sensory information being supplied.

Brain structure and function show changes with age. Reduced brain volume (Raz et al., 2005), cortical thickness (Salat et al., 2004) and white matter (Sullivan, Adalsteinsson, & Pfefferbaum, 2006) have been found in older adults. White matter reductions may be related to cognitive changes such as decreased processing speed (Batzokis et al., 2010; Sullivan, Rohlfing, & Pfefferbaum, 2010) and facial perception abilities (Thomas et al., 2008). Older adults also show decreased dopamine receptors in the frontal areas of the brain (Kaasinen et al., 2000) and this may result in poorer performance on working memory tasks (Li & Sikström, 2002).
Evidence from imaging studies show brain activation in the frontal lobe areas is greater in older adults compared to young adults on working memory tasks (Park & Reuter-Lorenz, 2009; Reuter-Lorenz & Cappell, 2008). Broader areas of activation across the right and left hemispheres of the brain in older adults have been seen on a variety of cognitive tasks. This has been termed HAROLD, hemispheric asymmetry reduction in older adults (Cabeza, 2002). Increased bilateral activation may help older adults during cognitive tasks (Cabeza, Anderson, Locantore, & McIntosh, 2002; Rosen et al., 2003). Facilitation has been seen on response time tasks (Reuter-Lorenz et al., 2000) as well as tasks requiring episodic memory encoding and retrieval (Cabeza, Anderson, Locantore, & McIntosh, 2002).

It has been suggested that these behavioral, structural, and functional changes may be understood through a scaffolding framework. In older adults, the brain utilizes compensatory actions in response to decreased white-matter, brain shrinkage, and declining dopamine receptors (Park & Reuter-Lorenz, 2009) and is called the scaffolding theory of aging and cognition. According to the scaffolding theory of aging and cognition (STAC), the broader areas of brain activation seen in older adults are a way of creating protective scaffolds needed to compensate for structural changes. The idea of scaffolding within the brain can be applied to children as well. Children who are acquiring new skills and knowledge utilize scaffolding to build new neural circuits, while in older adults scaffolding helps to maintain existing circuits and build new ones in novel situations. The STAC model may be useful in guiding programs and interventions designed to promote healthy cognitive aging (Goh & Park, 2009).

**Peripheral changes and cognition in older adulthood.** Changes in sensory receptors may result in reduced audibility; however, speech perception difficulties may be related to
changes in both peripheral and central auditory processes (Walton, 2010). The damaged peripheral auditory system encodes the speech signal with reduced fidelity. This degraded speech signal is sent along to the central auditory pathways and may result in speech perception difficulties. The effects of central processing deficits are especially apparent in challenging listening situations such as background noise (Gordon-Salant & Fitzgibbons, 1995; Snell & Frisina, 2000). Declines in cognitive processing abilities are a possible reason for difficulty perceiving speech in noise related to aging (Tremblay, Piskosz, & Souza, 2002; Tun, Wingfield, & O’Kane, 2002). Older adults suffer from greater distraction from the semantic content of the background noise compared to young adults (Tun, Wingfield, & O’Kane, 2002). Older adults also show broader areas of brain activation in the frontal areas, possibly as a compensatory measure in response to sensory declines (Goh & Park, 2009; Wong, Jin, Gunasekera, Abel, Lee & Dhar, 2009). Those who experience diminished memory and/or attention are particularly impacted by decreased sensory perception (Shinn-Cunningham & Best, 2008).

The ability to encode the temporal features of speech is crucial for speech understanding. Temporal cues are crucial for speech recognition because listeners are able to achieve good speech recognition without spectral cues (Rosen, 1992; Shannon, Zheng, Kamath, Wygonski, & Ekelid, 1995). Changes to the peripheral neural systems responsible for temporal encoding due to age may play a role in the decrease in speech recognition seen in older adult listeners with and without hearing loss (Dubno, Horwitz, & Ahlstrom 2003; Gordon-Salant & Fitzgibbons, 1993). Gap detection is a temporal resolution paradigm frequently used to examine the neural correlates of temporal processing. It is used to assess static temporal acuity. Gap detection consists of placing a silent gap between two carriers in order to model the silent pauses found in
spoken speech, such as voice onset time. A series of gap durations are presented in order to find the shortest gap the listener is able to detect, or gap threshold. Speech contains intensity fluctuations in the temporal envelope, known as amplitude modulation (AM). Speech cues needed for discrimination of phonemic features such as voicing and vowel perception are provided by AM characteristics. Parameters of AM speech may be valuable in auditory scene analysis, or the listener’s ability to separate the talker's message from background signals (Grimault, Bacon, & Micheyl, 2002). Older adult listeners, with and without hearing loss, have shown poorer performance on tasks assessing temporal acuity, such as gap detection and sinusoidal amplitude modulation, compared to young listeners (Galsberg, Moore, & Bacon, 1987; Moore & Glasberg, 1988; Moore, Peters, & Glasberg, 1992; Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994; Snell & Frisina, 2000).

Age related temporal processing deficits may be the result of changes in brainstem structures, such as the cochlear nucleus (Schatteman, Hughes, & Caspary, 2008) and the inferior colliculus (Walton, Frisina & O’Neill, 1997; Wang, Turner, Ling, Parrish, Hughes, & Caspary, 2009). The cochlear nucleus contains octopus cells which detect gaps in noise (Oertel, Bal, Gardner, Smith & Joris, 2009), while the inferior colliculus contains cells that respond to stimulus onset (Eggermont, 1999; Walton, Frisina, & O’Neill, 1997). Decline of these specialized cells in the inferior colliculus may result in age related changes in gap encoding, amplitude and frequency modulation encoding (Eggermont, 1999).

Declines in temporal processing abilities have also been shown through the use of objective subcortical measures. Older adults show offset timing delays in response to syllable stimuli on the auditory brainstem response (Vander Werff & Burns, 2011). They also have
impaired perception of voice onset time contrasts, providing evidence that some of the
difficulties may be the result of decreased temporal resolution on the central auditory system
(Tremblay, Piskosz, & Souza, 2003). Previous experiments have illustrated the importance of
accurate representation of subcortical temporal information, known to contribute to speech
perception in noise in young adults and children (Kraus, McGee, Carrell, King, Tremblay, &
Nicol, 1995; Skoe, Nicol, & Kraus, 2011).

Spectral processing abilities are also impacted with age. Older adults have greater
difficulty processing spectral information from the speech signal, such as pitch cues (Helfer &
Vargo, 2009). Pitch information provides cues for speaker identification. In the presence of
background noise an older listener may have difficulty separating one talker’s voice from many
talkers’ voices (Oxenham, 2008; Shinn-Cunningham & Best, 2008). The frequency following
response (FFR) has been shown to be reduced in older adults as well as increases in frequency
difference limens (Clinard, Tremblay, & Krishnan, 2010), providing objective evidence for
deprees in spectral encoding abilities which in turn impact speech perception.

The central auditory brain stem response (cABR) is an objective tool that can be used to
evaluate auditory processing associated with perception of speech in noise (Anderson & Kraus,
2010). The temporal and spectral components of the cABR have been shown to be related to
speech in noise perception both in children (Anderson, Skoe, Chandrasekaran, Zecker & Kraus,
2010) and adults (Song, Skoe, Banai, & Kraus, 2010). Those with better speech in noise
performance have shown greater fundamental frequency magnitude encoding in cABR
recordings (Anderson, Skoe, Chandrasekaran, Zecker & Kraus, 2010; Song, Skoe, Banai, &
Kraus, 2010). Those with poor speech in noise abilities showed reductions in peak timing and
poor waveform morphology on the cABR when speech stimuli were presented with background noise (Anderson, Skoe, Chandrasekaran, & Kraus, 2010, Anderson, Skoe, Chandrasekaran, Zecker & Kraus, 2010; Parbery-Clark, Skoe, & Kraus, 2009).

Anderson, Parbery-Clark, Yi and Kraus (2011) examined the cABR responses in older adults. Their speech in noise perception abilities were assessed using the Hearing in Noise Test (HINT; Bio-logic Systems Corp, Mundelin, IL). The participants were separated into 2 groups, those who scored above the median on the HINT and those below the median. Each participant in the high performing group was matched with one in the low performing group based on audiometric thresholds.

The cABR responses showed older adults in the high performing HINT group had greater subcortical representation of fundamental frequency in the speech syllable stimuli, as measured by RMS amplitude of the waveform. The neural timing of those in the high performing group was also less affected by noise, as reflected by a higher quiet-to-noise correlation between responses elicited in quiet compared to noise.

The experiment also revealed a strong relationship between the cABR response measures and HINT score. The encoding of speech syllable fundamental frequency, amount of change in waveform morphology, and timing were predictive of how well the participant performed on the HINT. This provides another piece of evidence to strengthen the findings that audiometric thresholds do not account for speech in noise performance (Killion & Niquette, 2000; Souza, Boike, Witherell, & Tremblay, 2007).
Short-term Memory Declines in Older Adulthood. Although this evidence indicates a development of working memory through the childhood years, there is also a great deal of evidence demonstrating the decline of working memory in older adulthood. Gordon-Salant and Fitzgibbons (1997) investigated the word and sentence recall performance of younger and older adults. There were 4 groups of participants, a) young adults without hearing loss, b) young adults with hearing loss, c) older adults without hearing loss and d) older adults with hearing loss. Stimuli were from the Revised Speech in Noise (R-SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) which contains sentence materials with high and low predictability word endings. The R-SPIN was modified to have silent intervals of 0, 400, 800, 1200, and 1600 milliseconds inserted between each word in the sentence. This created sentences with 5 different inter-word intervals (IWI). The task was sentence and final word recall.

Word recall for the low predictability stimuli revealed the following ranking from highest to lowest word correct scores: a) normal hearing young adults; b) normal hearing elderly adults; c) young adults with hearing loss; d) elderly adults with hearing loss. As the IWI increased, word recall scores declined for all participants, however the ranking of word correct scores associate with each participant group remained the same. A similar pattern of results was seen for word recall scores for low predictability sentences. For both the high predictability word and sentence stimuli, hearing loss and IWI had a significant effect on correct recall, however age did not. Overall, elderly adults without hearing loss showed poorer working memory performance compared to younger adults without hearing loss.

Schneider, Daneman, Murphy, and See (2000) evaluated the speech understanding of spoken discourse in quiet and noise for groups of younger and older adults. Participants were
asked to answer integrative comprehension questions and provide details about the discourse. The younger adults answered more questions correctly than the older adults, although this was the case for detail related questions rather than integrative questions. Both the younger and older adults answered more questions of both types correctly when the discourse was presented in quiet than in noise. The difficulty for the recall and/or comprehension of spoken stimuli could be attributed to decreased working memory capacity (Cohen, 1987; Wingfield, 1996; Wingfield & Tun, 2001) or loss of inhibition (Hasher & Zacks, 1988).

Declines in working memory processing and storage functions related to aging have also been investigated. Deficits in these functions may be related to age related declines in comprehension of written conversation (Cohen, 1981; Splich, 1983; Light & Anderson, 1985) and comprehension of spoken conversation in favorable listening conditions (Cohen, 1979, 1981; Light, Zelinski & Moore, 1982). Comprehension of spoken conversations may be more challenging than written conversations for an older listener. The listener must store and retrieve spoken information and is unable to go back and review conversational information if something is unclear (Pichora-Fuller, Schneider & Daneman, 1995).

Age related declines in working memory may be the result of decreased processing resources, as proposed by Craik and Byrd (1982). Older adults have fewer cognitive resources available for the functions of encoding, manipulation, maintenance, and retrieval. Evidence to support this hypothesis comes from investigations in which older adults were given survey questions in written format (answer choices are visible), and presented auditorily (answer choices are presented sequentially). Data reveal older adults answer the questions differently based on the format they in which are presented. The written format reduces the processing
resources required to complete the survey, as the participant can go back and read the question and possible answer choices. Auditory presentation requires encoding, rehearsal and maintenance of the question and answer choices, functions that require greater processing resources (Park & Gutchess, 2000).

Conclusions

The components and functions of short-term memory have been modeled and examined in several ways. The development of the four components of the Baddeley and Hitch model of short-term memory has been investigated (Baddeley, 2000; Baddeley & Hitch, 1974). The relation between short-term memory and spoken speech perception has not been examined. This relation has not been examined as a function of age. Previous investigations have shown that visual cues enhance auditory speech perception; however, the effect of auditory-visual presentation on short-term memory has not been investigated. Further examination of short-term memory and spoken speech perception will result in a greater understanding of peripheral and cognitive processes, as well as how they develop and decline over the lifespan. This may lead to the development of strategies for enhancing short-term memory and spoken speech perception through auditory-visual presentation.
Figure 1. A diagram of the four component short-term memory model. Adapted from Baddeley, 2000 (page 421). The arrows indicate bidirectional sharing of information between components (central executive, episodic buffer, phonological loop and visuospatial sketchpad) and their subcomponents (phonological store, subvocal rehearsal, visual store, spatial mechanism).
CHAPTER 3

Methods

Three different age groups were targeted and a cross-sectional design was used to evaluate potential difference in short-term working memory processes associated with age. School-age children were chosen to represent an age at which short-term memory functions might still be developing. Young adults were chosen to represent an age group in which short-term memory functions are mature. Finally, older adults over age 60 were chosen to represent an age group that may be experiencing a decline in short-term memory function.

A running memory task was used to test the hypothesis that short-term working memory performance would be enhanced if auditory information and visual speech production cues observable on the talker’s face were presented concurrently compared to auditory-only presentation. Stimuli were presented in either the auditory-visual or auditory-only modality, and under two acoustic conditions, quiet and in the presence of background noise. It was hypothesized that the visual cues would be especially useful for short-term working memory enhancement in the presence of background noise. Recall errors, j factor analysis for context effects (ability to identify the whole word based on the parts) for auditory-visual and auditory-only presentation, perceived workload, and error types were evaluated to explain differences in recall performance within or across groups.

Participants

Community dwelling participants composed the three participant groups in the study, 32 participants in each group (16 males per group). The first group consisted of children ages 10
and 11, the second of adults ages 26 to 30 and the third of adults over age 60. A priori calculation of power based on 96 participants was 0.78, indicating high power (Faul, Erdfelder, Lang & Buchner, 2007).

Written informed consent was obtained from the adult participants and from the child’s parent or guardian. Written assent was obtained from each child. All experimental protocols were approved by the Institutional Review Board of the University of Illinois. Participants were compensated $10 upon completion of the experiment and given coupons to local businesses. Children were also given a small prize, such as a brightly colored pencil.

Consideration for age –range groups. Children ages 10 and 11 were chosen to participate in the experiment. Previous investigations indicated that age seven is the youngest age at which the subvocal rehearsal process surfaces (Gathercole, 1998), and by age 14 short-term memory skills are adult-like (Choi, et al., 2008; Gathercole, Pickering, Ambridge, & Wearing, 2004). Pilot testing was conducted with children seven to eight years of age; however they were unable to maintain attention on the experimental task, and those able to complete the experiment achieved very low accuracy. Therefore, pilot testing was conducted with children ages 10 and 11 who demonstrated the ability to maintain attention and complete the experimental task with greater accuracy. As a result, children ages 10 and 11 were chosen to represent the age group with immature short-term working memory skills.

The age range for the young adult group was chosen to represent participants with mature short-term memory skills to serve as an experimental control group. It was also chosen to ensure maturity of frontal lobe functions have been reached (e.g., executive functioning and attentional
spotlight) (Fuster, 2002), as previous evidence suggests frontal lobe development continues through late adolescence and early adulthood (Sowell et al., 2001; Yurgelun-Todd, 2005).

The age range for the older adult group was over age 60. The reason this age range was chosen to represent participants who may have declines in short-term working memory function (Stenfelt & Rönnberg, 2009). Many older adults also have loss of hearing sensitivity (Gopinath et al., 2009); therefore, the hearing screening was crucial in determining the participant had normal hearing thresholds and that potential differences in performance were not the result of hearing loss.

**Recruitment strategy.** Flyers advertising the study were posted in the community and information was broadcast online to the University community (http://publicaffairs.illinois.edu/resources/eweek.html) to alert potential participants to the opportunity to participate. General eligibility criteria (described in detail, below) were reviewed with those individuals who initiated email or telephone contacts with the experimenter and any questions about the nature of the experiment were answered. Adults and parents of children who reported that they were interested in participating and met the general inclusion criteria were invited into the laboratory for further assessment of their eligibility and possible participation.

**Eligibility criteria.** A detailed questionnaire (see Appendix A) was completed by the adult participant or by the parent or guardian of the child participant to document the self-report of the following inclusion criteria: (a) negative history of heart attack, stroke, neurological problems, vision problems, noise exposure, ototoxic medications, attention problems, speech or language problems, learning disability, tinnitus or hearing aid use; and (b) American English as a first language. Additionally, to be eligible: adults were required to have at least eight years of
education; and all participants were required to have no history of speech, language, reading or attention difficulties; and negative history of chronic otitis media.

Finally, participants were required to pass measures of visual acuity, hearing sensitivity, and understanding of speech in noise, administered and scored by the experimenter. Also, the adult participants were required to pass a screening measure for general cognitive function. A brief description of each measure follows.

To meet the visual acuity criteria, participants were required to achieve a near visual acuity score of 20/30 or better, as measured by the Snellen near vision test for corrected vision (http://www.disabledworld.com/artman/publish/eye chart.shtml). To meet the hearing sensitivity criteria, participants were required to demonstrate behavioral air and bone conduction thresholds of 25 dB HL or better for pure tones bilaterally at 250-4000 Hz (ANSI 1996; 2004), tested by the experimenter, a licensed audiologist. The Quick Speech in Noise (SIN; Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004) was conducted in order to evaluate the participant’s speech in noise abilities in order to ensure that performance on the test stimuli presented in noise was not impacted by atypical difficulty with speech perception in noise. If a participant demonstrated low performance on the Quick SIN they would have been deemed ineligible for the study; however, none of the participants screened were deemed ineligible for the study because of performance on the Quick SIN. To meet the criteria for general cognitive function, adult participants were required to achieve a passing score on the Mini-Mental Status Examination (MMSE), a screening for major deficits in memory or cognitive function (Folstein, Folstein, & McHugh, 1975).
Test Materials

Experimental stimuli were bisyllabic words selected from the Medical Research Council database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.html) with a written frequency rating (Kucera & Francis, 1967) of 25 or greater and a familiarity rating (Gilhooly & Logie, 1980; Toglia & Battig 1978) of 100 or greater. Stimulus words were chosen to contain visible phonetic cues with viseme ratings (Ickes, 1980) of 1.0 to 0.25 for consonants to ensure they were visible on the talker’s face.

Audio and video recording. Word stimuli were spoken by one male and one female talker. They were video recorded with a Canon XL-H1 high definition camera to capture as much facial detail as possible and audio recorded with the onboard stereo condenser microphone to record the speech stimuli. The audio recorded speech signal was 24 dB above the ambient room noise, which was measured to be 41 dBZ on average using Praat software (Boersma, & Weenink). The talkers were native speakers of American English with no detectable dialect. They had natural, observable lip movements while speaking and symmetrical facial features. The talkers were instructed to practice saying the words naturally and fluently prior to recording.

Audio and video editing. The audio and video footage was edited using Virtual Dub software (www.virtualdub.org). A total of 480 words were separated into eight sets, called sets A through H. Each set contained six strings of words. The strings had 5, 7, 9, 11, 13 or 15 words in them. The words were ordered so that consecutive words were semantically unrelated and did not co-occur in everyday phrases to avoid priming effects. The order of the strings within each set was randomized. Two sets of word strings were presented in each of the stimuli conditions.
The recordings were edited to create four conditions of modality and acoustic condition: a) auditory-visual, quiet (AVQ); b) auditory-visual, noise (AVN); c) auditory-only, quiet (AOQ); and d) auditory-only, noise (AON). In the auditory-visual condition the stimuli consisted of the talker’s face being present along with the audio (65 dB SPL). In the noise conditions, the stimuli were edited to include continuous white Gaussian noise (+5 dB SNR) that was simultaneous with the word stimuli. This background noise was created using MatLab (The Mathworks, Inc., Natick, Massachusetts). An inverse Fast Fourier Transform was used to produce white Gaussian noise with a flat spectrum, and this noise was scaled to create a +5 dB signal-to-noise ratio relative to the word stimuli. Pilot testing using the Quick Speech in Noise (Quick SIN; Etymotic Research, 2001) revealed at the level of +5 dB SNR older adult listeners were able to recognize the speech materials and performance was not at ceiling.

**Stimuli Presentation**

The participants were seated inside a sound treated booth (Industrial Acoustics Company, model 1023), positioned approximately 1 meter away from the monitor (Viewsonic VP2365wb 23.5 inch monitor with a vertical refresh rate of 60 Hz at a rate of 30 frames per second) used to present visual stimuli (video of the talker’s face). The monitor also presented a large green dot that appeared immediately after the last word was presented, indicating the list was complete. The center of the monitor was 44 inches from the floor.

An ER-3A insert earphone (www.etymotic.com) was placed in the participant’s right ear for presentation of the auditory stimuli. Stimuli were presented monaurally in order to avoid variation in performance as the result of individual differences in binaural summation (Hawkins, Prosek, Walden & Montgomery, 1987). Verbal responses for word recall were captured with a
headset microphone (DPA 4088) positioned at a 45° angle near the participant’s mouth and recorded using CuBase 4 recording software (www.steinberg.net).

An RME FireFace 800 mixer with Total Mix software (www.rme-audio.de) was used to deliver the auditory and visual stimuli. The stimuli were played using the VLC media player (www.videolan.org). The order in which the stimuli conditions were presented was partially counterbalanced. They were organized into 16 playlists. The participants were assigned to a playlist so equal numbers of males and females saw the male talker and the female talker.

**Running Memory Task**

The experimental task was an adaptation of a running memory task used by McCoy et al., (2005). In that experiment, participants listened to word stimuli from taken from Miller and Selfridge (1950). In their study, the word stimuli were arranged into 16 lists, with 15 words per list. The 16 lists were divided based on the degree of contextual constraint, as indicated by the order of approximation to English. Order of approximation indicates the amount of context provided by the words in the list available. For example, a list with a zero order approximation to English would be composed of random words with no contextual meaning to one another (better, write, catch, native). There were four lists of words with low contextual constraint (zero and first order approximation) and 12 lists of words with high contextual constraint (2nd to 9th order approximation to English). Participants were told the task was to listen to lists of words. The list was stopped after 5, 7, 8, 10, 12, 13, 14 or 15 words. The end of the list was signaled by three large asterisks appearing on a screen in front of the participant. Participants were instructed to repeat the last three words they heard in the list out loud, regardless of list length.
**Pilot Testing.** The speech intelligibility of the stimuli heard in the presence of white noise (+5 dB SNR) was verified. This was done in order to confirm that errors made by participants were related to short-term memory failures, rather than poor intelligibility of the stimuli in noise. Six young adults with normal hearing were recruited. They were seated in the sounds treated booth. Auditory stimuli were presented to the right ear using an ER-3A insert earphone (www.etymotic.com). Participants were instructed to repeat the word they heard out loud as soon as they heard it. They were instructed to guess if they were unsure of a word.

Three participants (1 male) heard the female talker and three participants (1 male) heard the male talker. Each participant was presented with all 480 words that appeared in the experiment. Mean speech intelligibility for the female talker was 94% correct and 96% correct for the male talker. Across participants, 53 of the errors made out of a total 1440 words were no responses (0.02%), indicating that in very few cases the participant was not able to repeat back the word heard at all.

A small number of words (22) were reported incorrectly by more than 1 participant. There were 16 words that were reported incorrectly by 2 participants, and 6 words reported incorrectly by 3 participants. Fourteen of the 22 words had a verbal frequency rating (Kucera & Francis, 1967) less than 100 and 10 words had a verbal frequency rating less than 50. It is possible that these words were reported incorrectly because they were not frequently used words.

**Procedure**

Upon arrival in the lab, the participant or the participant’s parent/guardian (for the 10 and 11 year old children) completed the consent form and questionnaire. All participants then
completed the vision screening and the young and older adults completed the Mini-Mental Status exam (Folstein, Folstein, & McHugh, 1975). Next, the participants were seated in the sound treated booth. The hearing screening and the Quick Speech in Noise task were then conducted. The 10 and 11 year olds were able to complete the Quick SIN, although it is not typically utilized with children.

A practice list of bisyllabic words not used in the experiment was read aloud to participants in order to familiarize them with the task prior to beginning the experiment. An interval of 15 seconds separated words lists during the experiment. Participants were told they would hear strings of words that varied in length. They were told that regardless of list length, the task was to repeat the last four words they heard in the list out loud (McCoy et al., 2005). The words could be said in any order. Participants were encouraged to guess if they were unsure of a target word, and no limit was place on how much time they could take to respond. They were instructed to keep their eyes open throughout the experiment in order to see the large green dot that would appear when the list was complete. The green dot was the cue for the participants to recall the last 4 words they heard. Participants were given a break at the completion of a set of word strings. Both the young adult and older adult groups were notified when they completed half of the experimental trials.

The children were given the option of either collecting stickers on a cartoon picture after attempting to repeat the last four words in the string, or pushing game pieces from one side of a game board from the start line to the finish line. The number of stickers available and the number of game pieces corresponded to the number of experimental trials. Children were given a sticker
or game piece after completing each trial so they could see how many stickers or game pieces they had collected and how many remained.

A team of four undergraduate students from the Speech and Hearing Science department assisted with a portion of the data collection. Students were trained to follow the experimental protocol. During data collection for the running memory task, a student sat near the participant and wrote down verbal responses to provide a comparison to the audio recorded responses. If the participant was a child, a second student experimenter provided praise and motivation if needed. Students assisted in data collection for 63 of the 96 participants who completed the experiment (28 children, 16 young adults, 19 older adults). During data collection for the running memory task the experimenter was seated outside the sound treated booth to operate the computer programs used to present the stimuli. The experimenter monitored the participant’s recall inside the sound booth using circumaural headphones (Sony MDR-7509HD) and also wrote down the participants’ verbal word recall responses.

After completing an experimental condition, participants in the young adult and older adult groups were instructed to complete a portion of the Modified Subjective Workload Assessment (Luximon & Goonetilleke, 2001). Subjective workload was evaluated on a subset of three scales: time demand, concentration, and frustration. This was repeated for all four experimental conditions. Participants were asked to rate the subjective workload on a five point Likert-like scale for the task they had just completed. The workload ratings were completed for the four experimental conditions to determine if auditory-visual presentation reduced perceived workload, or if noise increased perceived workload. A significant difference in the workload...
ratings for the four experimental conditions would support the notion that auditory-only presentation and background noise increase perceptual effort.

**Additional Test Measures**

After the experiment, participants completed additional measures chosen in order to help explain possible sources of variance within and/or across age-range groups. Participants completed the following test measures: a) the digit span backwards task (Bromley, 1958) to examine short-term working memory; b) the Stroop color-word task (Stroop, 1935) for inhibition of irrelevant stimuli; c) a lip-reading task (sentence list from Utley, 1946) to evaluate visual speech perception; d) the Peabody Picture Vocabulary Test, Fourth Edition (Pearson, 2007); e) a target word receptive vocabulary measure; and f) a target word meaning task to examine vocabulary proficiency. A brief explanation of each measure follows.

**Digit span backwards task.** The experimenter read a string of digits aloud to the participant at a rate of one word per second. Participants were instructed to repeat back the digits out loud in the opposite order in which they were presented. If the string was repeated correctly, the next string of digits (one digit longer than the previous string) was presented. If the string was incorrectly repeated, the experimenter read a string of the same length composed of different digits. If the second string of the same length was also incorrectly repeated the task stopped and the longest span of digits correctly repeated backwards was recorded.

The digit span backwards task (Bromley, 1958) was administered to obtain a measure of participant’s ability to manipulate information in short-term verbal memory. It was hypothesized that a longer digit span backwards score might be correlated with fewer errors on the experimental task. Low backwards digit span (for the participants’ age) and poor performance on
the experimental task might indicate an immature or declining subvocal rehearsal process. As children age, the mean number of items correctly recalled on the digit span backwards task has been shown to increase, and has a mean of 3.8 items at age 10 (Isaacs & Varghs-Khadem, 1989). Older adults have been shown to have a slightly lower mean digit span backward score (about 5.34 items) compared to young adults who have a mean score of about 5.88 items (Bopp & Verhaeghen, 2005). A high backwards digit span score, poor performance on the vocabulary measures and poor performance on the experimental task might indicate the errors on the experimental task were not due to short-term memory but rather to limited language experience.

**Stroop task.** The Stroop task (Stroop, 1935) was created using E-Prime 1.1 software (www.pstnet.com). Two test blocks were created, one in which participants were instructed to identify the word on the screen and the other in which they were instructed to identify the color of the ink of the printed word. There were six colors in the task: blue, green, purple, yellow, red and black. The presentation was either congruent, (e.g., the word “blue” printed in blue ink), incongruent (e.g., the word “blue” printed in red ink), or neutral (the word “black” printed in black ink). The neutral presentation was the same for both blocks. The order in which the test conditions (color or word) were presented was randomly selected. The participant was instructed to either indicate the word presented or the color of the printed word by pressing a button on the keyboard. Identification accuracy and reaction time was recorded for each block. Performance on the Stroop color-word task was obtained to examine the participant’s ability to inhibit irrelevant information. A low score on the Stroop task and a large number of errors resulting from the recall of a word heard earlier in the list may indicate immature executive function in the children, or a decline of function in the older adults.
**Lip-reading task.** The male and female talkers used in the experiment were video and audio recorded saying the sentences created by Utley (1946). Recording conditions were identical to those used to record the experimental word stimuli. Participants were told to watch a video showing the talker from the experiment saying sentences. They were instructed to repeat aloud what they thought the talker said. Sentences were presented on the computer monitor used to present the experimental stimuli. The talker’s face was in full view. The sentences were presented with the audio recordings turned off. Participants were encouraged to guess if they were unsure. Each participant’s verbal responses were written down by the experimenter in the sound booth in addition to being audio recorded. The audio recordings were used if the experimenter had any question as to what the participant said. The lip reading task was completed to investigate the hypothesis that participants with higher lip reading proficiency may be better able to use visual cues to perceive the stimuli in the auditory-visual condition than those with low lip reading proficiency.

**Peabody Picture Vocabulary Test.** The Peabody Picture Vocabulary Test Fourth Edition (PPVT-4, Pearson, 2007) was administered to all participants. Testing began after verifying baseline vocabulary with the portion of the test appropriate for the participant’s age. The participant was instructed to look at the page consisting of four pictures and point to the picture said by the experimenter. Testing stopped when the participant made eight errors in one section of test materials, or when all materials had been presented in the case of the adults. The PPVT-4 was administered to identify participants who did not have an age appropriate vocabulary level and the scores were used to test the hypothesis that high scores in the receptive vocabulary measure might be associated with fewer errors on the experimental running memory
task. Fewer errors made on the experimental running memory task might be the result of larger vocabulary and/or greater language experience rather than short-term memory ability.

**Target word receptive vocabulary measure.** The participants were asked to complete a receptive vocabulary measure consisting of target words from the experimental task. This measure was conducted to determine if the word stimuli presented in the experimental running memory task were part of the participant’s receptive vocabulary. Photos representing 28 of the 56 target words were taken. One of the photos representing a target word was presented on a page along with three photos of foils. Participants were instructed to point to the photo of a target word. This measure was administered to determine if the participants who had these words in their receptive vocabularies might also make fewer errors on the experimental running memory task because of greater language experience rather than short-term memory ability.

**Target word meaning task.** Finally, participants were asked to complete a target word meaning identification task. They were shown a page with a target word presented in the experiment at the top and three foils at the bottom. For example, the target word “chlorine” appeared at the top of the page, and the three foils at the bottom were “pool, chemical, and water”. The target word and the foils were read to the participant. They were instructed to choose a word from the three foils that was similar in meaning to the target word. Each participant was presented with 21 target words. This task was included to determine if participants who knew the meaning of the target words might make fewer errors on the experimental task.
Data Analysis

When the experiment was complete, the responses written down by the experimenter were compared to those collected by the student inside the booth. This was done in order to confirm the participant’s response. The audio recording was replayed only if the responses written down by the student and the experiment did not agree. If there were any disagreement between the response written down by the student in the booth and the experimenter, the audio recording of the response was replayed. If the response on the audio recording did not match the responses written down by either the student or the experimenter, the response on the audio recording was determined the participants’ response.

Data were also analyzed to determine if significant differences existed in the following areas: a) speech intelligibility of last word heard in the string; b) overall word recall errors; c) j factor analysis (to examine context effects); d) workload ratings (from the young and older adult groups); e) error type (See Appendix B); f) Stroop task accuracy and reaction time; and g) lip-reading proficiency.

Overall word recall errors. The total number of word recall errors made in the four experimental conditions (AVQ, AVN, AOQ, and AON) X three age groups (children, young and older adults) were analyzed in a repeated measures ANOVA to determine if recall performance differed by age group. It was hypothesized that a greater number of recall errors might indicate immature or declining abilities in one or all of the components of short term memory, or loss of inhibition for irrelevant stimuli. The number of word recall errors made in the quiet acoustic condition was compared to those made in noise in order to determine if noise had a detrimental effect on short-term working memory performance.
Speech intelligibility of last word heard. The recall accuracy of the last word heard in the string (lag 1) was examined. The last word heard served as a measure of speech intelligibility, and correct recall indicated the word stimuli was audible and intelligible to the participant. A word was deemed an error if the response given was not the exact word spoken by the talker. For example, if the target word was “nation” and the participant recalled “nations” the word recall would be considered incorrect. Incorrectly recalled word were later categorized and examined (see Appendix B). A check of speech intelligibility was needed to distinguish it from perception that is loaded by working memory demands.

Lag errors. Recall errors for the last four target words presented in each string (lag 1, 2, 3, 4) X two presentation modalities (auditory-visual, auditory-only) X two acoustic conditions (quiet, noise) X eight word sets were evaluated in a repeated measures ANOVA. Lag 1 was defined as the last word heard in the string, lag 2 was the second word back, lag 3 was the third word back, and lag 4 was the fourth word back in the string. In order to determine if visual cues mediated short-term memory performance, the number of lag errors at each position in the auditory-visual modality was compared to those made in the auditory-only modality. A mixed-design ANOVA was used with the within subjects factors of lag, presentation modality, and acoustic condition, and between subjects factors of age group, condition order, and lip-reading proficiency.

Context effects. In addition to analysis of recall errors, a j factor analysis was conducted. The j factor was developed from a ratio of the recognition probabilities for the whole item and for parts within the whole and can be expressed as $j = \log (P_w) / \log (P_p)$, where $P_w$ represents the probability of recognizing the whole word, $P_p$ represents the probability of recognizing a part of
the whole word (phoneme) and $j$ represents the number of independent channels of information needed for recognition (Boothroyd & Nittrouer, 1988). It provides a way to quantify the relationship between the recognition of a whole item (i.e., word, syllable) and the recognition of its parts (Benkí, 2003). Previous investigations have shown that the $j$ factor as an effective measure of the tendency to perceive parts or chunks of an entire unit of stimuli (Boothroyd & Nittrouer, 1988) and to evaluate lexical bias (Benkí, 2003). The $j$ factor analysis provided information about the units of the target word recalled, giving insight into the lexical representation of the word. It was also used to examine the effects of context compared in recall. If the effects of context were high when recall performance was also high, this may indicate that context was associated with STM processes in spoken speech perception.

The number of phonemes correctly recalled from each response was coded by the experimenter. This was done for all 96 participants. The data from each participant was later coded by one of the four trained undergraduate research assistants. Each research assistant coded the number of phonemes correctly recalled for eight participants (4 males, 4 females) from the three age groups (children, young adults, and older adults) for a total of 24 participant data files per assistant. The participant data files to be coded were randomly assigned.

**Error type.** The errors were analyzed to determine if the incorrect word recalled was a word heard earlier in the string, word heard earlier in the experiment, was not related to the target word, contained part of the target word, or if the participant gave no response. Detailed descriptions of the error types can be found in Appendix B. Errors from all 96 participants were first coded by the experimenter, and then coded by a trained undergraduate research assistant. Each research assistant coded the error type for eight participants (4 males, 4 females) from the
three age groups (children, young adults, and older adults) for a total of 24 participant data files per assistant. It is important to note that the research assistant coded the error types and number of phonemes correctly recalled for the same set of participants.

Errors resulting from recalling a word heard earlier in the string provided information about the phonological loop, immature or declining inhibition abilities in the subvocal rehearsal process, and the phonological store. A no response error may have indicated failure to encode the information in the phonological loop, the visuospatial sketchpad or both. Incorrect responses categorized as unrelated to the target word, or containing part of the target word, were examined to determine if the error was the result of phoneme substitution, addition, or deletion (See Appendix B). The information about part word errors provided insight into errors resulting from incomplete phonetic representation, as the incorrectly recalled word might contain some of the phonemes found in the target word. One possible cause of a part word error might be the incorrect or incomplete representation of the target word being rehearsed in the subvocal rehearsal process.
CHAPTER 4

Results

Participants

The experiment was completed by 96 participants: 32 (16 male) from each of 3 age groups, with an equal number of males and females in each group. The age groups consisted of children ages 10-11, young adults ages 26-30, and adults 60 years and older. Initially, 194 potential participants were screened for eligibility over the phone or via email (37 children, 35 young adults, 122 older adults). Eligible participants were invited to the lab to complete the hearing screening and the participant questionnaire (See Appendix A) to determine final eligibility. After the initial screening, 32 children, 35 young adults, and 59 older adults were invited to the lab. Some participants were deemed ineligible after screening in the lab: 3 young adults (2 male) and 31 older adults (25 male). The most common reason for participants to be deemed ineligible for the experiment was hearing loss. The largest proportion of ineligible participants was older adult males. As displayed in Table 1, the males and females in the group of children and young adults were similar in age. The older adult males were more broadly distributed in age compared to the older adult females.

All participants had pure tone hearing sensitivity of 25 dB HL or less from 250 Hz to 4000 Hz, bilaterally. An omnibus analysis of variance (ANOVA) using a mixed-model repeated measures design with ear (right, left), frequency (250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz) as within subjects factors, and age group (children, young adults, older adults), and participant gender (male, female) as between subjects factors was conducted. Significant main effects were observed for frequency $F (4, 360) = 17.59, p < 0.01$ and age group $F (2, 90) = 43.19, p < 0.01$. 
There were no interactions (Ear X Frequency $F(4, 90) = 1.99$, $p = 0.15$; Ear X Frequency X Age Group $F(8, 360) = 3.51$, $p = 0.71$; Ear X Frequency X Gender $F(8, 360) = 1.85$, $p = 0.83$; Ear X Frequency X Age Group X Gender $F(8, 360) = 1.59$, $p = 0.20$).

Paired comparisons showed group differences in hearing sensitivity. There were no differences in mean thresholds between male and female participants across groups. Mean thresholds for the children ($M = 4.92$; $SD = 0.81$) were lower than those of the older adults ($M = 9.75$; $SD = 2.46$), $p < 0.01$. The mean thresholds for the young adults ($M = 4.66$; $SD = 1.05$) were lower than the older adults ($M = 9.75$; $SD = 2.46$), $p < 0.01$. As shown in Table 2, all participants had symmetrical hearing sensitivity in both ears, although only the right ear was used for presentation of experimental stimuli. The older adults had higher mean thresholds compared to the children and younger adults; however, the older adults’ thresholds were still within normal clinical limits, and eligibility criteria.

All groups also completed the Quick SIN (Killion, Niquette, Gudmendsen, Revit, & Banerjee, 2004). A score of 25.5 to 22.5 indicates the listener has little to no signal to noise loss, while a score between 22.5 and 18.5 indicates the listener has mild signal to noise loss. Mean scores from the group of children indicated mild signal to noise loss, while the young and older adults showed little to no signal to noise loss, as seen in Table 3. Group data also showed no difference between scores from male and female participants in each age group, also displayed in Table 3.

**Overall Word Recall Errors**

The overall numbers of word recall errors were determined for each of the four experimental conditions by age group. The number of errors made was converted to percent error. The percent error scores were transformed to arcsine units before statistical analysis.
An omnibus analysis of variance (ANOVA) using a mixed-model repeated measures design with modality (auditory-visual, auditory-only) and acoustic condition (quiet, noise) as within subjects factors, and age group (children, young adults, older adults) between subjects factor was carried out on the transformed scores. The ANOVA showed significant main effects of modality, $F(1, 84) = 123.54, p < 0.01$; acoustic condition, $F(1, 84) = 335.64, p < 0.01$; and age group $F(2, 84) = 31.57, p < 0.01$. There were no interactions (Modality X Age Group $F(1, 84) = 1.25, p = 0.29$, Acoustic Condition X Age Group $F(2, 84) = 5.94, p = 0.40$).

Pre-planned pairwise comparisons (Sidak confidence interval adjustment, $p = 0.05$) were carried out to examine the independent variables showing significant main effects. The transformed error scores pooled across modality showed significantly fewer errors were made in the auditory-visual modality ($M = 14.92; SEM = 0.55$) compared to the auditory-only condition ($M = 19.25; SEM = 0.57$), $p < 0.01$. There were also fewer errors made in quiet ($M = 13.70; SEM = 0.51$) compared to noise ($M = 20.46; SEM = 0.61$), $p < 0.01$. The transformed overall word recall error scores pooled across the experiment differed by group. The number of errors made by children ($M = 22.83; SEM = 0.92$) was significantly higher than the number of errors made by the young adults ($M = 12.79; SEM = 0.92$) or by the older adults ($M = 15.49; SEM = 0.91$), $p < 0.01$. The young adults’ performance ($M = 21.79, SEM = 0.92$) was not significantly different than that by the older adults ($M = 15.49; SEM = 0.92$), $p = 0.11$.

In order to investigate age group differences, the overall word recall errors in each condition were examined. Figure 2 shows the arcsine transformed error scores for each group by condition, comparing quiet conditions to noise conditions, and auditory-visual conditions to auditory-only conditions. Significant differences ($p = 0.05$) are indicated with a star. The error
bars represent standard error of the mean (SEM) in order to show the uncertainty in how well the sample mean represents the population mean (Nagele, 2003). As seen in Figure 2, when visual cues were present, fewer word recall errors were made in quiet compared to noise (AVQ compared to AVN; AOQ compared to AON). This was seen across age groups. In the presence of noise, the addition of visual cues resulted in fewer word recall errors (AVN compared to AON), also displayed in Figure 2. It is important to note that the number of errors made in the experimental conditions was not similar across groups, while pattern errors made were similar. The children made more errors across all conditions ($M = 22.85; SEM = 1.30$) than either the young ($M = 12.84; SEM = 0.92$) or older adults ($M = 15.78; SEM = 0.98$), $p < 0.01$. Another point of interest is that the number of errors made by the older adults ($M = 15.78; SEM = 0.98$) was not significantly different from that of the young adults ($M = 12.84; SEM = 0.92$), $p = 0.11$.

**Speech Intelligibility**

The number of errors made in the lag 1 position (last word heard) was examined as a check of speech intelligibility. As displayed in Table 4, low percentage of error for the lag 1 position across groups indicates the word stimuli were audible and intelligible. The children had a higher percentage of errors compared to the adult age groups, although the children also made more errors overall across conditions. The percentage of error was highest for the AON condition. Across groups, the fewest lag 1 errors were present in the AVQ condition, and the highest number of recall errors was made in the AON condition.

**Lag Errors**

The number of errors made in each lag position (i.e., lag 1 = last word heard, lag 2 = second back, lag 3 = third back, lag 4 = fourth back) was examined in order to explore short-term
working memory effects. The number of errors made in each lag position was converted to a percent error score. This percent score was then arcsine transformed. An omnibus analysis of variance (ANOVA) using a mixed-model repeated measures design with modality (auditory-visual, auditory-only), acoustic condition (quiet, noise), and lag (1, 2, 3, 4) as within subjects factors, and age group (children, young adults, older adults), as a between subjects factor was carried out on the transformed scores. The analysis of variance showed significant main effects of modality $F(1, 84) = 98.65, p < 0.01$; acoustic condition $F(1, 84) = 344.88), p < 0.01$; lag $F(1, 84) = 451.39, p < 0.01$; and age group $F(2, 84) = 33.83, p < 0.01$. There were no interactions (Modality X Age Group $F(2, 84) = 7.47, p = 0.48$; Acoustic Condition X Age Group $F(2, 84) = 6.13, p = 0.05$; Lag X Age Group $F(6, 252) = 11.01, p = 0.05$).

Pre-planned pairwise comparisons (Sidak confidence interval adjustment, $p = 0.05$) of the transformed scores revealed significantly fewer lag errors in the auditory-visual modality ($M = 15.70; SEM = 0.60$) compared to the auditory-only modality ($M = 19.95; SEM = 0.60$), $p < 0.01$. The errors in quiet ($M = 14.30; SEM = 0.56$) were significantly fewer compared to those in noise ($M = 21.36; SEM = 0.63$), $p < 0.01$. The errors in the lag 1 position ($M = 7.28; SEM = 0.52$) were significantly fewer compared to the errors in the lag 2 position ($M = 10.50; SEM = 0.60$), lag 3 position ($M = 19.40; SEM = 0.81$), and the lag 4 position ($M = 34.13; SEM = 0.97$), $p < 0.01$.

Finally, the number of lag errors made by the children ($M = 24.14; SEM = 0.97$) was more than those made by young adults ($M = 13.16; SEM = 0.97$) and older adults ($M = 16.17; SEM = 0.97$), $p < 0.01$. The number of lag errors made by the young adults ($M = 13.16; SEM = 0.97$) was not significantly different than those made by older adults ($M = 16.17; SEM = 0.97$), $p < 0.10$.

The recall errors made in each lag position were examined to determine if a short-term memory effect was present. Lag 1 was compared to lag 2, lag 3, and lag 4. Lag 2 was compared
to lag 3 and lag 4. Lag 3 was compared to lag 4. This was done in the 4 experimental conditions for each group. Paired samples \(t\) tests (Bonferroni corrected, \(p = 0.05\)) of the arcsine transformed lag errors between the lag positions revealed similar trends across groups. Comparisons between lag positions showed the recall errors made in each position were different from one another. All contrasts were significant \(p < 0.01\), with a few exceptions. In the AON condition the difference between the transformed lag 1 errors \((M = 21.76; SD = 12.21)\) and lag 2 errors \((M = 23.49; SD = 11.24)\) were not significant \((p = 0.42)\) for the group of children. The difference between the transformed lag 1 errors \((M = 3.00; SD = 4.01)\) and lag 2 errors \((M = 4.80; SD = 5.34)\) were not significant in the AVQ condition, \((p = 0.04)\) and the AON condition \((lag 1 M = 13.37; SD = 12.44; lag 2 M = 13.93; SD = 9.26; p = 0.78)\) for the group of young adults. In the group of older adults, there was no significant difference in the transformed lag 1 errors \((M = 16.00; SD = 10.30)\) and lag 2 errors \((M = 15.03; SD = 8.93)\) in the AON condition \(p = 0.97\).

The lag errors made in the quiet and noise acoustic conditions were compared for each group because of the significant main effect of acoustic condition and are displayed in Figure 3. Paired samples \(t\) tests (Bonferroni corrected, \(p = 0.05\)) revealed significantly fewer lag errors were made in quiet compared to noise in the auditory-only modality (i.e., without visual cues). As shown in Figure 3, fewer lag errors were made in the AOQ condition compared to the AON condition across groups. Again, note that the pattern of errors was similar across age groups though the children had more errors compared to the adult groups.

The lag errors made in the auditory-visual and auditory-only conditions were examined by group. Paired \(t\) tests showed the differences between the lag positions in the AVN and AON conditions were significantly different at several lag positions. As shown in Table 5, fewer lag
errors were made when visual cues were present compared to auditory-only conditions in the presence of noise.

**Workload Ratings**

The perceived workload ratings (Luximon & Goonetilleke, 2001) completed by the young and older adults were examined. The three workload rating categories were time, concentration, and frustration and they were rated on a five point Likert-like scale. A time rating of 1 indicated the experimental task took little time to complete and the participant felt they had a great deal of time between their response and the next word list, while a rating of 5 indicated the participant felt they had little time between their response and the presentation of the next word list. A concentration rating of 1 indicated the participant felt the experimental task required little concentration to complete, while a rating of 5 indicated the task required a great deal of concentration. A frustration rating of 1 indicated the experimental task produced little frustration or anxiety, while a rating of 5 indicated the task produced a great deal of frustration or anxiety.

An overall mean value and standard deviation were calculated using a participant’s ratings in three categories (time, concentration, frustration) for the four experimental conditions (AVQ, AVN, AOQ, AON). A standardized $z$ score was calculated by subtracting a raw score from the participant’s overall mean and then dividing that number by the participant’s overall standard deviation. The raw scores were converted to standardized $z$ scores in order to allow for the comparison of scores originating from different normal distributions.

The ratings in conditions with visual cues were compared to those without visual cues and the ratings in conditions in quiet were compared to those in noise. Wilcoxon matched-pairs tests were used to determine if the difference between the median rating for the conditions with and without visual cues was significant and for the conditions with and without noise.
Overall, the young adults’ workload ratings indicate less time, concentration and frustration in quiet conditions compared to noise. The young adults’ mean ratings of the perceived time needed to complete the experimental task was lower in the AOQ condition \((M = -0.28, SEM = 0.13)\) compared to the AON condition \((M = 0.01, SEM = 0.17), p = 0.02\). They also rated less perceived concentration was required to complete the task in the AVN condition \((M = 0.56; SEM = 0.13)\) compared to the AON condition \((M = 1.04; SEM = 0.17), p = 0.02\). The concentration ratings in the AOQ condition \((M = 0.34, SEM = 0.14)\) were lower than the AON condition \((M = 1.04, SEM = 0.08), p = 0.01\). The perceived frustration ratings in the AVQ condition \((M = -0.66, SEM = 0.14)\) were lower than the AVN condition \((M = -0.21, SEM = 0.18), p = 0.02\). Finally, the frustration ratings in the AOQ condition \((M = -0.79, SEM = 0.13)\) were lower compared to the AON condition \((M = 0.01, SEM = 0.17), p < 0.01\).

The workload ratings from the older adults indicated the experimental task required less perceived concentration and was less frustrating in quiet conditions compared to noise. Concentration ratings were lower in the AVQ condition \((M = 0.66, SEM = 0.10)\) compared to the AVN condition \((M = 0.84, SEM = 0.08), p = 0.04\). Frustration ratings were lower in the AVQ condition \((M = -0.81, SEM = 0.16)\) compared to the AVN condition \((M = -0.11, SEM = 0.15), p < 0.01\). Finally, frustration ratings in the AOQ condition \((M = -0.73, SEM = 0.15)\) were lower than the AON condition \((M = 0.14, SEM = 0.15), p < 0.01\).

**Digit Span Backwards**

Performance on the digit span backwards task showed the mean score for the children \((M = 4.41; SD = 0.76)\) was lower than the young adults \((M = 5.34; SD = 1.00)\) and older adults \((M = 5.00; SD = 1.08)\) for the older adults. A one-way analysis of variance was performed. The within subjects factor was digit span backwards score and the between subjects factors were age group
and participant gender. There was a significant effect of age group $F(2, 90) = 7.64; p = 0.001$. Participant gender was not significant $F(1, 90) = 0.44, p = 0.83$. There was no Age Group X Gender interaction $F(2, 90) = 0.77, p = 0.93$. Planned comparisons (Sidak corrected, $p = 0.05$) revealed that the scores from the children ($M = 4.41, SEM = 0.24$) were significantly lower than the young adults ($M = 5.34; SEM = 0.24$) $p = 0.001$ and the older adults ($M = 5.00; SEM = 0.24$) $p = 0.04$. Scores from the young adults ($M = 5.34; SEM = 0.24$) were not significantly different than those from the older adults ($M = 5.00; SEM = 0.24$), $p = 0.41$.

**Lip-reading**

The lip-reading scores were calculated in percent correct for each group. A sentence based lip-reading task was administered (Utley, 1946). Percent correct scores were determined by dividing the number of words correctly identified by the participant by the total number of words presented in the lip-reading task (125 total words). The children showed the lowest percent correct scores ($M = 11.47, SD = 9.87$) while the young adults ($M = 24.53, SD = 22.15$) were similar to the older adults ($M = 24.75, SD = 21.31$). Within each group, participants showed a wide range of scores. The children showed a smaller range of scores (median score = 5, minimum score = 3, maximum score = 37) compared to the young adults (median score = 17.5, minimum score = 3, maximum score = 78) and older adults (median score = 15.5, minimum score = 3, maximum score = 67). A wide range of individual variation in lip-reading ability has been previously reported in children and adults (Bernstein, Demorest, & Tucker, 1991; Demorest, Bernstein & DeHaven, 1996; Demorest & Bernstein, 1992).

Enhancement scores were calculated from the total number of errors made in the AVN and AON conditions. This was done to examine the benefit the addition of visual cues in the presence of noise, and to examine the possibility that participants with higher lip-reading scores
would experience facilitated recall when visual cues were present. The number of errors made in
the AVN and AON conditions were converted to the number of correct responses. Enhancement
scores wer then calculated from the formula AVN-AON/(AVN+AON), where the number of
correct responses without visual cues were subtracted from the number of correct responses with
visual cues and were divided by the number of correct responses with visual cues added to the
number of correct responses without visual cues (Holmes, 2007; Ross, Saint-Amour, Leavitt,
Javitt, & Foxe, 2007). The enhancement scores were not calculated for each lag position for the
AVN and AON conditions, because the order in which the conditions were presented was
partially counterbalanced; therefore, participants were not presented with the same lists of words
in each condition.

Figure 4 shows scatterplots of lip-reading scores against enhancement scores. The dashed
line represents the zero value for enhancement score, or no auditory visual enhancement. Scores
above the dashed line indicate the participant had a higher number of words correctly recalled
when visual cues were present (AVN condition), while scores below the dashed line indicate
participants who had a higher number of words correctly recalled without visual cues (AON
condition). As shown in Figure 4, many participants across groups did not have high lip-reading
scores; however, across groups, most participants were also above the dashed line, indicating a
lower mean number of recall errors in the AVN condition compared to the AON condition. This
suggests and overall benefit from the addition of visual cues across groups. Some participants
were below the dashed line, indicating they did not benefit from the addition of visual cues. The
scatterplots in Figure 4 also show individual differences within each age group. Some participants
who had low lip-reading scores had high enhancement scores, while others who had high lip-
reading scores did not show enhancement (below the dashed line).
Stroop Task

Performance accuracy on the Stroop task (Stroop, 1935) was high across groups, as shown in Table 6. This was the case when the participants were instructed to press the key corresponding to the ink color of the printed word (color condition), when they were told to press the key corresponding to the word on the screen (word condition), and when the ink color and word on the screen matched (neutral condition). Table 6 shows the mean percent correct for the congruent, incongruent, and neutral conditions for color and word.

Reaction time (in milliseconds) on the Stroop task was examined by age group. There was a significant main effect condition (congruent, incongruent or neutral) in the Stroop task reaction time $F(2, 186) = 19.77, p < 0.01$. There was not a significant main effect of instruction (color, or word) $F(2, 93) = 0.72, p = 0.40$. There were no interactions (Instruction X Age Group $F(2, 93) = 2.31, p = 0.11$; Condition X Age Group $F(4, 186) = 1.37, p = 0.05$; Instruction X Condition $F(4, 93) = 0.16, p = 0.34$; Instruction X Condition X Age Group $F(4, 93) = 0.16, p = 0.92$). Pairwise comparisons revealed the reaction times in the congruent condition ($M = 1302.80$; $SEM = 30.77$) were slower than the incongruent condition ($M = 1229.88$; $SEM = 31.51$) and the neutral condition ($M = 1160.66$; $SEM = 29.43$), $p < 0.01$. Reaction times in the incongruent condition ($M = 1229.88$; $SEM = 31.51$) were slower compared to the neutral condition ($M = 1160.66$; $SEM = 29.43$), $p < 0.01$.

There was also a significant difference in reaction time between groups. Figure 6 shows the comparison of mean reaction time by group. As shown in Figure 5, children ($M = 1518.56$; $SEM = 47.90$) had significantly slower reaction times than young adults ($M = 1042.82$; $SEM = 39.95$), $p < 0.000$ and older adults ($M = 1132.05$; $SEM = 49.90$), $p < 0.000$. Older adults ($M =$
1132.05; $SEM = 49.90$) had slower reaction times compared to young adults ($M = 1042.82; SEM = 39.95$), however, this was not statistically significant, $p = 0.19$.

The overall word recall errors made in each experimental condition (AVQ, AVN, AOQ, AON) were compared to the Stroop task reaction times (color-congruent, color-incongruent, color-neutral, word-congruent, word-incongruent, word-neutral). The pattern of correlations between Stroop task reaction times and overall recall errors differed by age group. As seen in Table 7, there were several significant correlations between Stroop reaction times and overall errors in the AVN condition for the group of children. For the group of young adults, there were significant correlations between Stroop reaction times and overall recall errors in the AVQ and AVN conditions, as displayed in Table 7. The group of older adults showed a significant correlation between Stroop reaction time and overall errors in the AVN condition, as shown in Table 7.

The number of errors made in lag 3, lag 4 and pooled lag 3 and 4 errors from the 4 experimental conditions were compared to the mean Stroop task reaction times. First, the mean number of lag errors in the lag 3, lag 4 and pooled lag 3 and 4 positions were calculated. The mean number of errors in the lag 3, lag 4 and pooled lag 3 and 4 positions were correlated with some of the reaction times on the Stroop task. As shown in Table 8, some Stroop reaction times were significantly correlated to lag 3, lag 4 and pooled lag 3 and 4 errors.

**Peabody Picture Vocabulary Test (PPVT)**

Performance on the Peabody Picture Vocabulary Test Fourth Edition (PPVT-4, Pearson, 2007) showed all participants had scores appropriate for their chronological age. PPVT-4 standardized scores from the children fell into the average to moderately high score categories. Performance from the adults was at ceiling, with standardized scores in the moderately high to
extremely high score categories. A one way ANOVA with the within subjects factor of standardized PPVT score and the between subjects factors of participant gender and age group showed a significant main effect of age group $F(2, 90) = 571.18, p = 0.000$ (Sidak corrected, $p = 0.05$). There was no significant main effect of participant gender $F(1, 90) = 0.71, p = 0.97$. There were no interactions (Gender X Group $F(2, 90) = 1.04, p = 0.36$. A priori planned comparisons showed the standardized scores from the children ($M = 101.69; SD = 8.27$) were significantly lower than the young adults ($M = 154.56; SD = 5.16$) and the older adults ($M = 140.81; SD = 5.49$), $p = 0.000$. The scores from the young adults ($M = 154.56; SD = 5.16$) were significantly higher than those from the older adults ($M = 140.81; SD = 5.49$), $p = 0.000$.

The PPVT scores had been obtained in order to test the hypothesis that participants with large receptive vocabularies and/or greater language experience might make fewer errors on the experimental task. Scores across groups indicated all participants had excellent vocabulary skills for their age, regardless of performance on the experimental task.

**Target Word Receptive Vocabulary**

Stimuli for the target word receptive vocabulary task were photos representing 28 of the 56 target words. A photo representing a target word was presented on a page with three photos of foils. This task was administered to test the hypothesis that participants with the target words in their receptive vocabularies might make fewer errors on the experimental task because of greater language experience. There were 4 items missed by 82% of the participants across age groups: *portion*, *sugar*, *outside*, and *morning*. The high error rate across groups for these items suggests the photo presented to represent the target word was unclear; therefore they were removed from the analysis. Mean percent correct scores from the children ($M = 88.41; SD = 5.37$), were slightly
lower compared to the young adults ($M = 94.92; SD = 2.75$) and older adults ($M = 95.57; SD = 3.66$).

A one way ANOVA with the within subjects factor of arcsine transformed target word receptive vocabulary score and the between subjects factors of participant age group showed a main effect of age group $F (2, 93) = 23.67, p = 0.000$ (Sidak corrected, $p = 0.05$). Planned comparisons of the transformed scored showed the scores from the children ($M= 62.80; SD = 6.48$) were significantly lower compared to the young adults ($M = 73.05; SD = 7.25$), $p = 0.00$ and the older adults ($M = 75.78; SD = 9.74$) $p < 0.000$. There was no statistically significant difference in scores between the young ($M = 73.05; SD = 7.25$) and older adults ($M = 75.78; SD = 9.74$), $p = 0.45$.

**Target Word Meaning Task**

The target word meaning task consisted of a target word printed on a page with three foils underneath. Participants were instructed to point to the foil that had the same meaning as the target word. This task was administered to test the hypothesis that participants who knew the meaning of the target words might make fewer errors on the experimental task. There were four items on the target word meaning task that were missed by 75% of the children and 40% of the young and older adults who completed the task. The words were: *carbon, naval,* *merit,* and *patent.* These items were removed from the analysis. Mean percent correct scores from the children ($M = 78.86; SD = 13.35$) were lower than the young adults ($M = 92.38; SD = 6.24$) and older adults ($M = 90.81; SD = 6.33$).

A one way ANOVA with the within subjects factor of arcsine transformed score on the target word meaning task and the between subjects factor of participant age group showed a significant main effect of age group $F (2, 93) = 26.03, p = 0.000$ (Sidak corrected, $p = 0.05$).
priori planned comparisons showed scores from children ($M = 54.97; SD = 12.77$) were significantly lower than the young adults ($M = 70.79; SD = 12.44$), $p = 0.000$ and the older adults ($M = 68.11; SD = 12.00$), $p = 0.000$. The number of errors in the young adults ($M = 70.79; SD = 12.44$) was not significantly different compared to the older adults ($M = 68.11; SD = 12.00$), $p = 0.77$.

**J Factor Analysis**

The $j$ factor score (Boothroyd & Nittrouer, 1988) was calculated to examine the effect of context. The $j$ factor score was calculated to determine if the addition of visual cues would provide greater signal redundancy and richer context, particularly in noise, facilitating word recall. A lower $j$ factor score indicates a greater effect of context. The number of phonemes correctly recalled from each response, and the number of words correct was coded by the experimenter and one trained undergraduate research assistant. Inter-rater reliability was high for number of phonemes correct and number of words correct. Cronbach’s alpha (Cronbach, 1951) for phonemes correct and words correct were 0.93 and 0.99 respectively.

Word recall errors in the data were examined and categorized prior to $j$ factor score calculation. Errors categorized as earlier in string, no response, repeated another target, or earlier in experiment were removed from the analysis (See Appendix B). Errors categorized as no response were removed from the analysis because they did not provide phonological information, while the earlier in string, earlier in experiment and repeated another target errors were clearly whole words that had been incorrectly recalled. All the errors made in an experimental condition fell into one of those error categories, for some of the participants, leaving no data available to calculate a $j$ factor score. As seen in Table 9, this resulted in several
missing values for each condition by group. Note that there were no missing values in the AON condition for the children, young or older adult groups.

Participants’ j factor scores in the AVN condition was plotted against their lip-reading score (in percent correct) to investigate the hypothesis those with higher lip-reading score would be able to utilize visual contextual cues, enhancing word recall. Figure 6 shows scatterplots of lip-reading scores in percent correct and AVN j factor score for each group. The regression line on the plot represents an estimation of the relationship between the lip-reading score and AVN j factor score. Note that the scatterplots are missing values for participants for whom a j factor score could not be calculated (children = 15.63% missing data, young adults = 34.38 % missing data, older adults = 28.13% missing data). As seen in Figure 6, participants in the group of children and young adults show a slight trend of higher lip-reading scores associated with lower j factor scores. A lower j factor score indicates a greater effect of context. This suggests a trend that participants higher lip-reading scores may have been better able to use contextual visual cues for correct word recall in the AVN condition compared to those with lower lip-reading scores.

The j factor scores in the AVN condition were compared to those from the AON condition to investigate the hypothesis that the addition of visual cues would provide more contextual cues facilitating recall performance, especially in noise. In the group of children, 53% had a lower j factor score in the AVN condition than the AON condition, while 65% of the young adults and 45% of the older adults had a lower score in the AVN condition compared to the AON condition. Lower scores in the AVN condition compared to the AON condition might indicate the addition of visual cues may have provided contextual cues that facilitated higher recall accuracy.
**Error Types**

The word recall errors were examined by type. See Appendix B for a full description of how error types were categorized. Earlier in string errors, earlier in experiment, and repeated another target errors were examined because they were thought to occur due to inability to inhibit irrelevant stimuli. Unrelated word and part word errors were thought to occur when the participnt was trying to guess the target word, or perhaps was able to recall certain parts of the word. The analyses reveal the pattern of errors differed by modality, acoustic condition and by group. Intra-rater reliability was calculated using Cronbach’s alpha (Cronbach, 1951) for the experimenter and the four trained undergraduate assistants, and was found to be 0.98, 0.96, 0.95, 0.96, and 0.95, respectively. Inter-rater reliability was high for earlier in string, earlier in experiment, no response, unrelated word, part word, semantically related, nonsense word, repeated another target, phoneme substitution, phoneme deletion and phoneme addition errors. Cronbach’s alpha (Cronbach, 1951) for earlier in string, earlier in experiment, no response, unrelated word, part word, semantically related, nonsense word, repeated another target, phoneme substitution, phoneme deletion, and phoneme addition errors were 0.99, 0.98, 0.99, 0.92, 0.94, 0.98, 0.99, 0.93, 0.97 and 0.98 respectively.

The error types were analyzed using the Kruskall-Wallis one way analysis of variance (Kruskal & Wallis, 1952), because the distribution of the errors across type and group was not homogenous. The distribution of errors across modality (auditory-visual, auditory-only) was examined for each group to see if more errors were present when visual cues were absent. As seen in Table 10, more errors were present when visual cues were absent for the part word and phoneme substitution error categories across age groups.
The distribution of errors across acoustic condition (quiet, noise) was examined for each group to see if more errors were present in noise compared to quiet. Table 11 shows the categories showing a significant difference ($p = 0.05$) in the distribution of errors across acoustic conditions. Across groups, more errors were present in noise for the unrelated word, part word and phoneme substitution error categories.

The number of recall errors from each group was compared by error category to determine if group performance differed. A one way ANOVA with the within subjects variables of error type (earlier in string, earlier in experiment, part word, unrelated word, and no reponse) and the between subjects variable of age group (children, young adults, older adults) showed a main effect of group for the error types of earlier in string $F(2, 381) = 44.82, p = 0.02$; and no response $F(2, 381) = 36.43, p = 0.000$. Planned comparisons showed no significant difference in the number of earlier in string errors between the group of children ($M = 5.75; SD = 4.67$) and the young adults ($M = 4.95; SD = 2.61$), $p = 0.18$, or between the children ($M = 5.75; SD = 4.67$) and older adults ($M = 6.11; SD = 2.65$), $p = 0.79$. The number of earlier in string errors was significantly different between the young adults ($M = 4.95; SD = 2.61$) and the older adults ($M = 6.11; SD = 2.65$), $p = 0.02$. In the no response error category the group of children had significantly more errors ($M = 5.20; SD = 4.90$) compared to the young adults ($M = 1.84; SD = 2.86$), $p = 0.000$ and the older adults ($M = 1.86; SD = 2.68$), $p = 0.000$. There was no significant different between the young adults ($M = 1.84; SD = 2.86$) and older adults ($M = 1.86; SD = 2.68$), $p = 1.00$, in the no response error category.

The number of phoneme substitution, addition, and deletion errors were compared from each age group. Paired samples $t$ tests (Bonferoni corrected, $p = 0.05$) showed the children had significantly more phoneme substitution errors ($M = 19.63; SD = 17.49$) compared to the young
adults ($M = 8.38; SD = 10.00$), $p = 0.000$ and compared to the older adults ($M = 10.52; SD = 11.50$) $p = 0.000$. There was no significant difference between the young adults ($M = 8.38; SD = 10.00$) and the older adults ($M = 10.52; SD = 11.50$), $p = 0.50$ in the phoneme substitution error category.

The older adults may have been unable to inhibit some irrelevant words heard previously in the string, as shown by more earlier in string errors compared to the children and young adults. The children may have been less willing to guess if they were unsure of a target word, or the target word may not have been encoded, or maintained in STWM, resulting in more no response errors compared to the young and older adults. The children also had more phoneme substitution errors compared to the young and older adults, suggesting perhaps the encoding of the target word was inaccurate, or the target word was incorrectly rehearsed or maintained in short-term working memory.
Table 1

*Years of Age for Participants Completing the Experiment*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>10.50 (0.52)</td>
<td>27.88 (1.59)</td>
<td>65.38 (4.84)</td>
</tr>
<tr>
<td>Female</td>
<td>10.38 (0.50)</td>
<td>27.06 (1.44)</td>
<td>65.50 (2.88)</td>
</tr>
</tbody>
</table>
Table 2

Age Group Comparisons of Hearing Thresholds (dB HL) Mean (SD)

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RE (SD)</td>
<td>RE (SD)</td>
<td>RE (SD)</td>
</tr>
<tr>
<td>250 Hz</td>
<td>5.87 (2.87)</td>
<td>5.63 (2.46)</td>
<td>7.66 (3.81)</td>
</tr>
<tr>
<td>500 Hz</td>
<td>4.38 (4.16)</td>
<td>2.97 (3.78)</td>
<td>7.50 (4.02)</td>
</tr>
<tr>
<td>1000 Hz</td>
<td>5.16 (3.47)</td>
<td>5.47 (2.87)</td>
<td>8.44 (4.66)</td>
</tr>
<tr>
<td>2000 Hz</td>
<td>5.47 (3.20)</td>
<td>5.63 (3.30)</td>
<td>10.47 (4.81)</td>
</tr>
<tr>
<td>4000 Hz</td>
<td>4.06 (3.69)</td>
<td>4.06 (3.22)</td>
<td>14.38 (4.71)</td>
</tr>
</tbody>
</table>

Note. RE = Right Ear, LE = Left Ear
Table 3
*Age Group Comparisons of Quick Speech in Noise (SIN) Score, Mean (SD)*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Children</th>
<th></th>
<th>Young Adults</th>
<th></th>
<th>Older Adults</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RE</td>
<td>LE</td>
<td>RE</td>
<td>LE</td>
<td>RE</td>
<td>LE</td>
</tr>
<tr>
<td>Male</td>
<td>22.56 (0.77)</td>
<td>22.56 (0.93)</td>
<td>23.50 (0.37)</td>
<td>23.38 (0.50)</td>
<td>22.94 (0.73)</td>
<td>22.88 (0.72)</td>
</tr>
<tr>
<td>Female</td>
<td>22.63 (0.62)</td>
<td>22.19 (0.79)</td>
<td>23.38 (0.50)</td>
<td>23.31 (0.71)</td>
<td>22.75 (0.86)</td>
<td>22.79 (0.76)</td>
</tr>
</tbody>
</table>

*Note. Quick Speech in Noise (Quick SIN, Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). The Quick SIN is not normed on children. Right Ear = RE, Left Ear = LE*
Table 4

*Lag 1 Percent Error, Mean (SD) by Condition and Age Group*

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Experimental Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AVQ</td>
</tr>
<tr>
<td>Children</td>
<td>0.61 (0.95)</td>
</tr>
<tr>
<td>Young Adults</td>
<td>0.43 (0.58)</td>
</tr>
<tr>
<td>Older Adults</td>
<td>0.24 (0.38)</td>
</tr>
</tbody>
</table>

*Note.* 12 words were possible in each experimental condition. AVQ = Auditory-visual, quiet; AVN = Auditory-visual, noise; AOQ = Auditory-quiet, noise; AON = Auditory-only, noise
Table 5
*Age Group Comparison of Lag Errors, Mean (SD) in Noise*

<table>
<thead>
<tr>
<th>Lag Position</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AVN</td>
<td>AON</td>
<td>AVN</td>
</tr>
<tr>
<td>Lag 1</td>
<td>1.91 (1.73)*</td>
<td>4.34 (2.27)</td>
<td>0.78 (1.18)*</td>
</tr>
<tr>
<td>Lag 2</td>
<td>2.84 (2.57)*</td>
<td>4.69 (2.07)</td>
<td>1.28 (1.08)*</td>
</tr>
<tr>
<td>Lag 3</td>
<td>5.81 (2.84)</td>
<td>6.22 (2.42)</td>
<td>2.59 (1.56)*</td>
</tr>
<tr>
<td>Lag 4</td>
<td>8.16 (2.30)*</td>
<td>9.25 (1.34)</td>
<td>5.56 (2.69)</td>
</tr>
</tbody>
</table>

*Note.* For all comparisons degrees of freedom = 31. Bonferroni corrected utilized. * = $p < 0.01$. 
Table 6
*Age Group Comparison of Percent Correct Stroop Task Scores, Mean (SD)*

<table>
<thead>
<tr>
<th>Stroop Condition</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>100.00 (0.00)</td>
<td>100.00 (0.00)</td>
<td>100.00 (0.00)</td>
</tr>
<tr>
<td>CI</td>
<td>96.47 (4.60)</td>
<td>95.47 (8.36)</td>
<td>95.64 (7.10)</td>
</tr>
<tr>
<td>CN</td>
<td>100.00 (0.00)</td>
<td>100.00 (0.00)</td>
<td>100.00 (0.00)</td>
</tr>
<tr>
<td>WC</td>
<td>97.65 (6.64)</td>
<td>98.13 (5.92)</td>
<td>100.00 (0.00)</td>
</tr>
<tr>
<td>WI</td>
<td>99.12 (1.96)</td>
<td>98.91 (2.76)</td>
<td>97.35 (5.89)</td>
</tr>
<tr>
<td>WN</td>
<td>98.82 (4.85)</td>
<td>98.75 (4.92)</td>
<td>98.82 (4.85)</td>
</tr>
</tbody>
</table>

*Note.* CC = Color Congruent; CI = Color Incongruent; CN = Color Neutral; WC = Word Congruent; WI = Word Incongruent; WN = Word Neutral
Table 7
Age Group Comparisons of Pearson Correlations Between Overal Recall Errors and Stroop Reaction Times

<table>
<thead>
<tr>
<th>Stroop Condition</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AVQ</td>
<td>AVN</td>
<td>AOQ</td>
</tr>
<tr>
<td>CC</td>
<td>0.10</td>
<td>0.42*</td>
<td>0.24</td>
</tr>
<tr>
<td>CI</td>
<td>0.26</td>
<td>0.52**</td>
<td>0.32</td>
</tr>
<tr>
<td>CN</td>
<td>0.12</td>
<td>0.35*</td>
<td>0.21</td>
</tr>
<tr>
<td>WC</td>
<td>0.15</td>
<td>0.47**</td>
<td>0.28</td>
</tr>
<tr>
<td>WI</td>
<td>0.11</td>
<td>0.30</td>
<td>0.18</td>
</tr>
<tr>
<td>WN</td>
<td>0.37</td>
<td>0.37*</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Note. * = \( p < 0.05 \), ** = \( p < 0.01 \) Bonferroni correction utilized. CC = Color Congruent; CI = Color Incongruent; CN = Color Neutral; WC = Word Congruent; WI = Word Incongruent; WN = Word Neutral. AVQ = Auditory-visual, quiet; AVN = Auditory-visual, noise; AOQ = Auditory-only, quiet; AON = Auditory-only, noise.
Table 8
Age Group Comparisons of Pearson Correlations Between Lag Errors and Stroop Reaction Times

<table>
<thead>
<tr>
<th>Stroop Condition</th>
<th>Age Group</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Pooled</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Pooled</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory-visual, quiet</td>
<td>Children</td>
<td>-0.19</td>
<td>0.28</td>
<td>0.67</td>
<td>-0.03</td>
<td>0.13</td>
<td>0.08</td>
<td>-0.00</td>
<td>-0.12</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.00</td>
<td>0.30</td>
<td>0.19</td>
<td>0.12</td>
<td>0.19</td>
<td>0.20</td>
<td>0.11</td>
<td>-0.02</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>CN</td>
<td>-0.17</td>
<td>0.29</td>
<td>0.08</td>
<td>0.18</td>
<td>0.07</td>
<td>0.15</td>
<td>0.05</td>
<td>-0.24</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>-0.13</td>
<td>0.23</td>
<td>0.07</td>
<td>0.14</td>
<td>0.13</td>
<td>0.17</td>
<td>0.05</td>
<td>-0.18</td>
<td>-0.10</td>
</tr>
<tr>
<td></td>
<td>WI</td>
<td>-0.10</td>
<td>0.02</td>
<td>-0.05</td>
<td>0.00</td>
<td>-0.10</td>
<td>-0.07</td>
<td>0.07</td>
<td>-0.17</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>WN</td>
<td>0.02</td>
<td>0.26</td>
<td>0.17</td>
<td>-0.03</td>
<td>0.22</td>
<td>0.17</td>
<td>0.10</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>Auditory-visual, noise</td>
<td>Children</td>
<td>0.33</td>
<td>0.21</td>
<td>0.33</td>
<td>0.51</td>
<td>0.17</td>
<td>0.37</td>
<td>0.15</td>
<td>-0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>CI</td>
<td>0.39*</td>
<td>0.24</td>
<td>0.39*</td>
<td>0.49</td>
<td>0.18</td>
<td>0.37</td>
<td>0.21</td>
<td>0.22</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>CN</td>
<td>0.37*</td>
<td>0.15</td>
<td>0.32</td>
<td>0.44</td>
<td>0.29</td>
<td>0.44</td>
<td>0.01</td>
<td>-0.19</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>WC</td>
<td>0.37*</td>
<td>0.19</td>
<td>0.34</td>
<td>0.27</td>
<td>0.03</td>
<td>0.15</td>
<td>0.01</td>
<td>-0.19</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>WI</td>
<td>0.26</td>
<td>0.10</td>
<td>0.23</td>
<td>0.16</td>
<td>-0.14</td>
<td>-0.04</td>
<td>0.05</td>
<td>-0.18</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>WN</td>
<td>0.40*</td>
<td>-0.01</td>
<td>0.23</td>
<td>0.30</td>
<td>-0.24</td>
<td>-0.05</td>
<td>0.23</td>
<td>-0.06</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Table 8, cont’d.  
*Age Group Comparisons of Pearson Correlations Between Lag Errors and Stroop Reaction Times*

<table>
<thead>
<tr>
<th>Stroop Condition</th>
<th>Age Group</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Pooled</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Pooled</th>
<th>Lag 3</th>
<th>Lag 4</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Auditory-only, quiet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>Children</td>
<td>0.13</td>
<td>0.17</td>
<td>0.17</td>
<td>0.41</td>
<td>0.19</td>
<td>0.21</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.08</td>
</tr>
<tr>
<td>CI</td>
<td>Young Adults</td>
<td>0.14</td>
<td>0.21</td>
<td>0.21</td>
<td>0.16</td>
<td>0.05</td>
<td>0.12</td>
<td>0.23</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>CN</td>
<td>Older Adults</td>
<td>0.18</td>
<td>0.07</td>
<td>0.14</td>
<td>0.03</td>
<td>0.26</td>
<td>0.20</td>
<td>0.33</td>
<td>-0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>WC</td>
<td></td>
<td>0.02</td>
<td>0.17</td>
<td>0.12</td>
<td>-0.09</td>
<td>0.10</td>
<td>0.03</td>
<td>0.12</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>WI</td>
<td></td>
<td>-0.01</td>
<td>0.05</td>
<td>0.03</td>
<td>-0.16</td>
<td>0.04</td>
<td>-0.06</td>
<td>0.14</td>
<td>0.35*</td>
<td>0.36*</td>
</tr>
<tr>
<td>WN</td>
<td></td>
<td>-0.09</td>
<td>-0.01</td>
<td>-0.06</td>
<td>-0.15</td>
<td>0.70</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Auditory-only, noise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>Children</td>
<td>0.29</td>
<td>0.10</td>
<td>0.27</td>
<td>0.06</td>
<td>0.83</td>
<td>0.09</td>
<td>0.19</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>CI</td>
<td>Young Adults</td>
<td>0.25</td>
<td>0.08</td>
<td>0.24</td>
<td>0.13</td>
<td>0.21</td>
<td>0.21</td>
<td>0.18</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>CN</td>
<td>Older Adults</td>
<td>0.28</td>
<td>0.06</td>
<td>0.25</td>
<td>0.20</td>
<td>0.09</td>
<td>0.19</td>
<td>0.03</td>
<td>-0.19</td>
<td>-0.11</td>
</tr>
<tr>
<td>WC</td>
<td></td>
<td>0.36*</td>
<td>0.20</td>
<td>0.36*</td>
<td>0.19</td>
<td>0.25</td>
<td>0.28</td>
<td>0.15</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td>WI</td>
<td></td>
<td>0.39*</td>
<td>0.23</td>
<td>0.41*</td>
<td>-0.11</td>
<td>0.21</td>
<td>0.06</td>
<td>0.12</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>WN</td>
<td></td>
<td>0.25</td>
<td>0.02</td>
<td>0.21</td>
<td>0.18</td>
<td>0.12</td>
<td>0.19</td>
<td>0.14</td>
<td>0.16</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*Note.*  
* = p < 0.05, ** = p < 0.01  
Bonnferroni correction utilized.  
CC = Color Congruent; CI = Color Incongruent; CN = Color Neutral; WC = Word Congruent; WI = Word Incongruent; WN = Word Neutral.
Table 9
*Percent Missing Data by Condition and Age Group*

<table>
<thead>
<tr>
<th>Age group</th>
<th>AVQ</th>
<th>AVN</th>
<th>AOQ</th>
<th>AON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>15.63</td>
<td>6.25</td>
<td>6.25</td>
<td>0.00</td>
</tr>
<tr>
<td>Young Adults</td>
<td>34.38</td>
<td>9.38</td>
<td>25.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Older Adults</td>
<td>28.13</td>
<td>9.38</td>
<td>21.88</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note.* There were 32 values possible per condition. AVQ = Auditory-visual, quiet; AVN = Auditory-visual, noise; AOQ = Auditory-only, quiet; AON = Auditory-only, noise.
Table 10

*Age Group Comparisons of Error Type Contrasts, Mean (SD) by Condition*

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AV</td>
<td>AO</td>
<td>AV</td>
</tr>
<tr>
<td>NR</td>
<td>4.91 (5.00)</td>
<td>5.48 (4.82)</td>
<td>1.48 (2.21)</td>
</tr>
<tr>
<td>UW</td>
<td>2.34 (2.48)*</td>
<td>4.20 (3.63)</td>
<td>1.15 (1.39)</td>
</tr>
<tr>
<td>PW</td>
<td>1.05 (1.12)*</td>
<td>2.03 (1.80)</td>
<td>0.69 (0.86)*</td>
</tr>
</tbody>
</table>

*Note. * = p < 0.05 Bonferroni correction utilized. AV = Auditory-visual, AO = Auditory-only, NR = no response, UW = unrelated word, PW = part word*
Table 11  
Age Group Comparisons of Error Type Contrasts, Mean (SD) by Acoustic Condition

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Children</th>
<th>Young Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>Condition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIE</td>
<td>1.77 (2.16)</td>
<td>2.38 (2.33)</td>
<td>0.63 (0.72)*</td>
</tr>
<tr>
<td>NR</td>
<td>4.03 (3.72)*</td>
<td>6.36 (5.64)</td>
<td>1.42 (2.28)</td>
</tr>
<tr>
<td>UW</td>
<td>1.89 (2.09)*</td>
<td>4.66 (3.58)</td>
<td>0.60 (0.88)*</td>
</tr>
<tr>
<td>PW</td>
<td>0.98 (1.15)*</td>
<td>2.09 (1.74)</td>
<td>0.66 (0.80)*</td>
</tr>
<tr>
<td>NS</td>
<td>0.09 (0.34)*</td>
<td>0.33 (0.62)</td>
<td>0.00 (0.00)</td>
</tr>
</tbody>
</table>

Note. * = p < 0.05 Bonferroni correction utilized. Q = Quiet, N = Noise EIE = earlier in experiment, NR = no response, UW = unrelated word, PW = part word, NS = nonsense word
Figure 2. Mean overall word recall arcsine transformed error scores made in each condition for (a) children, (b) young adults and (c) older adults. Auditory-visual, quiet (AVQ) = solid white, Auditory-visual, noise (AVN) = dots, Auditory-only, quiet (AOQ) = solid gray, Auditory-only, noise (AON) = stripes. Error bars indicate SEM. * = p < 0.05 Sidak confidence interval adjustment utilized.
Figure 3. Mean arcsine transformed lag errors in the Auditory-only, quiet condition (solid gray) and the Auditory-only, noise condition (stripes) for (a) children, (b) young adults and (c) older adults. * = p < 0.05. Error bars represent SEM. Bonferroni correction utilized.
Figure 4. Scatterplots of lip-reading score in percent correct and auditory-visual enhancement scores (Holmes, 2007; Ross, Saint-Armor, Leavitt, Javitt & Foxe, 2007) in noise for (a) children, (b) young adults and (c) older adults calculated from the formula (AVN-AON)/(AVN+AON). The talker was identical in both tasks. The dashed line represents the enhancement zero value.
Figure 5. Mean reaction times for children, young adults, and older adults in milliseconds. Horizontal bars indicate significant comparisons. Error bars indicate SEM. Young adults are represented by white bars, older adults by gray bars and children by stripes. YA = young adults, OA = older adults, C = children. * = p < 0.05
Figure 6. Scatterplots of lipreading score (in percent correct) and j factor score in the Auditory-visual, noise condition for (a) children, (b) young adults and (c) older adults. Solid line indicates linear trend.
CHAPTER 5

Discussion

The goal of the present investigation was to examine the relation, if any, between short-term working memory and recall errors in spoken speech perception. Three different age groups composed of school age children, young adults, and older adults with normal hearing and vision were recruited to investigate short-term working memory performance across three different times in development. Participants completed a running memory task in 4 experimental conditions: auditory-visual, quiet (AVQ); auditory-visual, noise (AVN); auditory-only, quiet (AOQ); and auditory-only, noise (AON). The experimental paradigm for the present investigation was modeled after the task utilized by McCoy et al., (2005) in order to separate speech intelligibility and target word recall. It was hypothesized that the number of recall errors in the group of children would be higher than the young and older adults, and that the older adults would have a higher number of recall errors compared to the young adults. It was hypothesized that the addition of visual cues might facilitate short-term working memory, reflected by fewer recall errors, compared to conditions without visual cues. The presence of noise was expected to result in more recall errors compared to quiet conditions.

A framework for short-term working memory is the multi-component model (Baddeley, 2000; Baddeley & Hitch, 1974). The Baddeley and Hitch model is well suited as a framework for examining short-term working memory and spoken speech perception because it contains specialized components for auditory and visual stimuli, the phonological loop and the visuospatial sketchpad. It is important to note that short-term working memory also serves to connect very recent events to information in long term memory as well. In the Baddeley and Hitch multi-component model of short-term working memory, the central executive is
responsible for this function. The central executive manages working memory functions, as well as retrieving information held in long-term memory. The episodic buffer is the fourth component in the Baddeley and Hitch model, which coordinates functions between the phonological loop, visuospatial sketchpad, and the central executive.

**Overall Target Word Recall Errors**

The overall number of recall errors made in each experimental condition was compared across age groups. Across groups, the number of overall target word recall errors was highest in the AON condition and lowest in the AVQ condition. Children showed a significantly higher number of overall target word recall errors compared to the young and older adults. The overall target word recall errors by the young and older adults were not statistically different.

The difference in performance by the children had been expected, as previous findings have concluded that children’s short-term working memory skills are still developing until about the age of 14 (Gathercole, Pickering, Ambridge, & Wearing, 2004). This result has also been demonstrated for speech perception tasks involving short-term working memory, such as the one used by Choi et al., (2008), in which children saw a string of digits on a computer screen, and then they listened to a list of five (PBK) words. They were instructed to repeat the word out loud. At the end of the five word list, the children were instructed to recall the digits that had been presented. Children’s accuracy for recall of the digits improved as a function of age, suggesting short-term working memory capacity increases with age.

The relation between age and short-term working memory in the present experiment may help to explain the higher number of overall recall errors by children in conditions with visual cues present. This result may be because, in children, auditory-visual speech perception skills continue to develop with age compared to adults. Additional support for this possibility comes
from Ross et al., (2011), who presented monosyllabic words to children and adults in quiet and in pink noise at a range of signal-to-noise ratios (SNR). Stimuli were presented with and without visual cues (video of the talker’s face). Findings from Ross et al., (2011) indicated the children showed improved scores when visual cues were present; however, their performance was significantly poorer compared to the adults, especially as the SNR increased. Within the group of children, scores improved as a function of age, suggesting the auditory-visual speech processing abilities change up through later adolescence.

Although the ages of the children in the Ross et al., (2011) study were similar to those in the present study, it is important to note that the stimuli, the experimental task, and the SNRs differ. Consistent with the Ross et al. (2011) study, the children in the present study showed more errors compared to adults when visual cues were present. It is plausible that the children in the current study may have been continuing to develop their auditory-visual speech perception skills; however, the current experiment does not allow auditory-visual speech perception skills to be disentangled from short-term working memory skills.

Children showed more overall recall errors in conditions with noise compared to young and older adults. It is important to point out that the Quick Speech in noise task (Quick SIN; Etymotic Research, Elk Grove Village, IL; Killion, Niquette, Gudmendsen, Revit, & Banerjee, 2004) administered was not normed on children. In this case it was used as a screening for those who may have had a greater difficulty with speech perception in noise. Children have been shown to have greater difficulty with speech perception in noise compared to adults (Johnson, 2000). This may have contributed to children’s higher number of recall errors in noise compared to young and older adults.
It had been expected that the older adults would have a higher number of overall target word recall errors compared to the young adults. This expectation was based on previous findings suggesting potential deficits from several areas in older adults with normal hearing sensitivity. For example, difficulties listening in noise (Gordon-Salant & Fitzgibbons, 1995; Oxenham, 2008; Shinn-Cunningham & Best, 2008; Snell & Frisina, 2000), as well as deficits in cognitive processing in the presence of background noise (Tremblay, Piskosz, & Souza, 2002; Tun, Wingfield, & O’Kane, 2002), have been shown in older adults with normal hearing sensitivity. Objective measures have shown older adults with normal hearing sensitivity may have declines in temporal (Schatteman, Hughes, & Caspary, 2008; Vander Werff & Burns, 2011) and spectral resolution abilities (Clinard, Tremblay, & Krishnan, 2010; Helfer & Vargo, 2009), resulting in reduced accuracy of speech signal.

**Evidence for a Short-term Working Memory Effect**

A key finding from the present study was the presence of a short-term working memory effect. A short-term working memory effect was defined as increase in the number of lag errors with an increase in lag position. The last word heard in the string was defined as lag 1, second word back lag 2, third word back lag 3, and fourth word back lag 4. Comparisons of the number of lag errors made in the four experimental conditions showed a short-term working memory effect for nearly all comparisons analyzed.

Evidence for the increase in the number of lag errors as a function of increasing lag position is consistent with the findings of McCoy et al., (2005). Unlike the present study, McCoy et al. (2005) included older adults with clinically normal hearing and with hearing loss while all participants in the current study had clinically normal hearing. The present study utilized semantically unrelated bisyllabic word stimuli, while McCoy et al., (2005) presented words that
varied in order of approximation to English. The experimental task utilized by McCoy et al., (2005) instructed participants to recall the final three words heard in the string while the present study instructed participants to recall the last four words heard in the string.

McCoy et al., (2005) hypothesized the fewest recall errors would be seen for the last word presented in the string. Fewer recall errors for the last word heard would suggest the stimuli were audible and intelligible. It would also suggest that recall errors for words further back in the string were the result of short-term working memory issues, rather than poor speech intelligibility. A second hypothesis was that the group of older adults with hearing loss would show poorer recall performance (more errors) compared to those with normal hearing. This result would not be due solely to reduced hearing sensitivity but to higher demand placed on processing resources needed to accurately perceive the word stimuli, reducing cognitive resources available for encoding and rehearsal of the stimuli. Reduced cognitive resources would then be reflected in poorer recall accuracy for the second and third words back in the string. The hypotheses suggested by McCoy et al. (2005) are in line with the effortfulness hypothesis (Rabbitt, 1968), which states exertion of greater perceptual effort results in fewer available resources for cognitive functions such as short-term working memory.

The results of the McCoy et al. (2005) study supported their hypotheses. Few recall errors were made for the last word heard, indicating the stimuli were audible and intelligible. Recall errors increased with increasing lag position, indicating the higher number of recall errors for words further back in the string were the result of short-term working memory issues. Additionally, the older adults with hearing loss showed more recall errors compared to those
with normal hearing, suggesting those with hearing loss had to exert greater perceptual effort; therefore, fewer resources were available for cognitive functions.

Similar to the results of the McCoy et al. (2005) study, the participants in the present study showed the fewest errors for the last word heard (lag 1), indicating the lower recall accuracy for the target words further back in the string (lag 2, 3 and 4) was not due to an inability to understand the spoken words. Few recall errors in the lag 1 position indicate the spoken words were intelligible, indicating the increased number of errors with increasing lag position was the result of short-term working memory rather than intelligibility issues.

Although the increased number of lag errors as a function of lag position also supports the effortfulness hypothesis, the Baddeley and Hitch model may also be able to provide an explanation for these results. Recall errors for words further back in the string might be the result of greater cognitive load on the central executive’s attentional switching and focus functions. The greater cognitive load for words further back in the string might also strain the episodic buffer’s ability to coordinate functions between the central executive and the components specialized for auditory (phonological loop) and visual (visuospatial sketchpad) information.

**Perceived workload.** The young and older adults were asked to rate the amount of perceived time, concentration and frustration they felt while completing the task after the completion of an experimental condition using the modified subjective workload ratings scale (Luximon & Goonetilleke, 2001). Both the young and older adults rated conditions with noise as requiring higher levels of concentration and producing higher levels of frustration or anxiety compared to quiet conditions. Rabbitt’s (1968) effortfulness hypothesis may help to explain the differences in perceived workload ratings in quiet and noise. In the presence of noise, the listener
must exert greater perceptual effort in order to encode the desired stimuli, leaving fewer resources available for cognitive processes such as short-term working memory, resulting in greater levels of frustration. The increased work load ratings in noise might also be explained by the multi-component model (Baddeley & Hitch, 1974, Baddeley, 2000) as the presence of background noise may reduce the short-term working memory resources available, increase the cognitive load on the central executive and episodic buffer, and perhaps result in greater perceived frustration when recalling the target words.

The increased perceived workload in the presence of noise may have resulted in increased lag errors at all positions compared to quiet conditions. Participants may have had fewer short-term working memory resources available in one or more components in the Baddeley and Hitch model (Baddeley, 2000; Baddeley & Hitch, 1974), or have exerted greater perceptual effort (Rabbitt, 1968) in the presence of noise, for all of the word stimuli heard. The reduction in resources and/or increased effortfulness may have resulted in a higher number of lag errors at all lag positions.

**Addition of visual cues.** The number of lag errors in noise conditions with visual cues was compared to those without visual cues as a function of lag position. This was done to investigate the hypothesis that the addition of visual cues would facilitate short-term working memory functions. There were fewer lag errors in the noise condition with visual cues (AVN) compared to the noise condition without visual cues (AON). The pattern of results differed by group. The children showed significantly fewer errors for lags 1, 2 and 4 in the AVN condition compared to the AON condition. The young adults and older adults showed significantly fewer errors for lags 1, 2 and 3 in the AVN condition compared to the AON condition. In the group of young and older adults, perhaps the presence of visual cues could not provide further benefit at
lag 4 because of the increased cognitive processing in the central executive and episodic buffer required for words further back in the string.

**Auditory-visual enhancement.** Auditory-visual enhancement scores Holmes, 2007; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007) were calculated from the lag errors in the AVN and AON conditions (AVN-AON)/(AVN+AON). An enhancement score greater than zero indicates the participant had fewer lag errors in the AVN condition compared to the AON condition. Many participants across groups showed enhancement scores above zero, suggesting the addition of visual cues in noise facilitated target word recall.

Enhancement scores were plotted against scores from the lip-reading task to determine if those with higher lip-reading scores showed greater benefit from the addition of visual cues. Each age group contains participants who had a low lip-reading score, and still demonstrated auditory-visual enhancement. Alternatively, there were also participants with had high lip-reading scores who did not demonstrate auditory-visual enhancement.

The visual cues in the AVN condition may have benefitted participants who did not have high lip-reading scores by providing information ahead of the auditory information. Previous evidence has shown talkers begin facial articulation movements about 100-300 milliseconds before speech (Grant & Seitz, 2000). Previous data support the suggestion that the visual information that occurs before speech provides an estimation of when the auditory information will occur, thus boosting sensitivity and supporting processing in the brain (Arnal, Morillon, Kell, & Giraud, 2009; Arnal, Wyart, & Giraud, 2011). The hypothesis that these early visual cues serve as way in which the listener’s attention can be directed to a time point when events relevant to the speech signal will occur has been called “Attention in Time” (Jones, Johnston, &
Puente, 2006; Large & Jones, 1999; Nobre & Coull, 2010; Nobre, Correa, & Coull, 2007). A recent study completed utilizing MEG (Golumbic, Cogan, Schroeder, & Poeppel, 2013) provides objective evidence suggesting in the presence of noise, early visual cues may act as way of directing the attentional spotlight to relevant stimuli. Therefore, participants in the present study who did not have high lip-reading scores may have benefitted from the early visual cues in the AVN condition, as they may have acted as a spotlight on relevant stimuli when background noise was present.

Previous findings from Brault, Gilbert, Lansing, McCarley, & Kramer (2010) showed older adults with lip-reading scores that fell below the median within the experimental sample benefitted from the addition of visual cues when noise was present on a running memory task. Similar to the present study, participants in the Brault, et al. (2010) experiment showed fewer recall errors on a running memory task even if their lip-reading proficiency was at the median within the sample group. The authors speculate that less proficient lip-readers might be able to make use of coarse visual cues that may have been helpful in the presence of background noise. This speculation is supported by previous findings from Jordan, McCotter and Thomas (2000) who found point-light images of consonant-vowel syllables presented with auditory stimuli improved recognition in background noise. Point-light images convey only highly prominent information from visual speech (Rosenblum & Saldaña, 1998). Consistent with these findings, some participants in the present study with low lip-reading scores showed auditory-visual enhancement, suggesting the addition of visual cues facilitated short-term working memory performance, as demonstrated by fewer recall errors.
Summary of Running Memory Task Performance

The findings from the current study suggest a relation between speech perception and short-term working memory as a function of age group, modality (auditory-visual vs., auditory-only) and acoustic condition (quiet vs. noise). The Baddeley and Hitch (1974) model may provide a framework for the findings. Results suggest short-term working memory is a crucial part of speech perception. When the listener’s speech perception abilities are impacted by noise, absence of visual cues, or continued skill development (as may be the case in children), one or more of the components of short-term working may be impacted. This may compromise the listener’s ability to encode, manipulate, rehearse, or maintain the stimuli, negatively affecting target word recall.

Age group differences were observed, supporting the notion that there are differences in short-term working memory abilities and use of visual cues at different points in chronological age (Gathercole, 1998; Ross et al., 2007). The group of children showed more overall target word recall errors compared to the young and older adults. The number of overall target word recall errors from the older adults was not significantly different compared to the young adults. This is inconsistent with what is typically reported in the literature. Once hearing loss had been controlled for, older adults show poorer performance compared to young adults on a variety of tasks (Cervera, Soler, Dasi, & Ruiz, 2009; Fogerty, Humes & Kewley-Port, 2010; Kidd & Humes, 2012).

Across groups, the number of lag errors for target word stimuli further back in the string (lags 2, 3 and 4) increased as a function of lag position, providing support for a short-term working memory effect. Increasing lag position may have placed a greater cognitive load on the
central executive’s attentional switching and focus functions. Fewer lag errors were seen in quiet compared to noise, and adult groups rated higher perceived frustration levels for conditions with noise. The presence of noise may have increased cognitive load, straining the episodic buffer’s ability to coordinate functions between the central executive, phonological loop, and visuospatial sketchpad. This may have resulted in a higher number of recall errors across lag positions.

The addition of visual cues may facilitate short-term working memory for some listeners, particularly in noise, as evidenced by fewer lag errors in conditions with the talker’s face present compared to those without the talker’s face. Early visual cues in the AVN condition may have acted as a spotlight on relevant stimuli when noise was present. The Attention in Time theory posits these early cues may have directed attention to relevant visual cues, potentially facilitating encoding in the visuospatial sketchpad and later resulting in improved word recall (Jones, Johnston, & Puente, 2006; Large & Jones, 1999; Nobre & Coull, 2010; Nobre, Correa, & Coull, 2007). Previous data showed older adults with lip-reading scores that fell below the median within the experimental sample had fewer recall errors on running memory task in noise when visual cues were present (Brault, et al., 2010). In the present study, less proficient lip-readers might be able to make use of coarse visual cues that may have been helpful in background noise.

**Additional Measures**

Several additional measures were administered in order to examine possible sources of variation within groups, or across age groups. Participants completed the digit span backwards task (Bromley, 1958) to examine short-term working memory abilities and the Stroop color-word task (Stroop, 1935) for inhibition. They also complete the Peabody Picture Vocabulary Test, Fourth Edition (Pearson, 2007), a target word receptive vocabulary task, and a target word
meaning task to examine vocabulary skills. The j factor score (Boothroyd, & Nittrouer, 1988) was calculated to examine the possible role of contextual cues in word recall. The incorrect responses were coded and categorized by error type and phoneme errors.

**Digit Span Backwards.** The digit span backwards had been administered in order to examine individual differences. Span scores from the children were lower compared to the adult groups; however, scores were appropriate for chronological age. These scores suggest that participants had age appropriate short-term working memory skills. Perhaps the running memory task in the current experiment utilized different or additional short-term working memory skills compared to the backwards digit span.

**Inhibition.** The Stroop task was administered to evaluate participants’ ability to inhibit irrelevant information. Inhibition of irrelevant stimuli is one cognitive function that is part of several executive functions (Blakemore & Choudhury, 2006). Older adults may have difficulty inhibiting irrelevant stimuli (Hasher & Zacks, 1988). In children, imaging studies using DTI have provided objective evidence that brain structures associated with inhibitory functions continue to develop up through late adolescence (Madsen, et. al., 2010).

Accuracy on the Stroop task was high, with the mean percent correct scores ranging between 95-100 percent correct, across groups. Reaction time was also examined. Children had slower reaction times (in milliseconds) across Stroop task conditions compared to the younger and older adults. Perhaps the slower reaction times from the children suggest they required additional time to inhibit the irrelevant stimuli, providing support to the evidence their inhibitory abilities continue to develop.

Significant correlations were found between Stroop reaction time and lag 3, lag 4, and pooled lag 3 and 4 errors. The trend across groups and experimental conditions revealed fewer
lag errors were associated with faster Stroop reaction time. Perhaps participants with faster Stroop reaction times were also able to quickly inhibit irrelevant word stimuli, or had more efficient short-term working memory skills, resulting in fewer lag errors. It is important to point out that there is no direct evidence to support this possibility, as participants were not limited in the amount of time given to respond on the experimental task or on the Stroop task. This was done intentionally in order to keep task instructions consistent across tasks.

**Vocabulary Measures.** Vocabulary measures were administered to determine if participants who had the target words in their receptive vocabulary and knew their meaning would have fewer lag errors on the running memory task compared to those who did not. These measures included the PPVT-4 (Pearson, 2007), the target word receptive vocabulary measure and the target word meaning measure.

Standardized scores on the PPVT revealed the young and older adults fell into the moderately high to extremely high score categories. The standardized scores from the group of children were significantly lower compared to the young and older adults; however their scores were in the average to moderately high categories and were appropriate for their grade level. The standardized scores on the PPVT show that all groups had scores appropriate for chronological age or grade level, and the adult groups had greater vocabulary knowledge compared to the children.

The target word receptive vocabulary measure was modeled after the PPVT; the experimenter chose photos to represent the target word and three foils. Participants were instructed to point to the photo that represented the target word. The target word meaning task consisted of a target word printed at the top of the page and three foils below. Participants were
instructed to choose the foil that had the same meaning as the target word. A limitation of the measures created by the experimenter was a high error rate on some test items, however when these items were removed from the analysis accuracy scores were high across groups. This suggests that the group of children did not have more errors on the experimental task compared to the adult groups because they had poor vocabulary skills, did not have the target words as part of their vocabulary, or did not know the meaning of the target words.

**J Factor Score.** The j factor score was calculated to determine if the addition of visual cues might allow participants to use context to facilitate target word recall. This was a new application of the j factor score, which has been used previously to examine the impact of context on syllable, word and sentence recognition in adults and children (Benkí, 2003; Boothroyd & Nittrouer, 1988; Caldwell & Nittrouer, 2013) The j factor score provides a measure of the tendency to perceive parts or chunks of an entire unit of stimulus context (Boothroyd, & Nittrouer, 1988) and parts of a stimulus (Benkí, 2003).

J factor scores in the AVN condition were plotted against lip-reading scores to examine the possibility that participants with good lip-reading proficiency might be better able to use visual cues as supportive context, facilitating target word recall. Children and young adults with high lip-reading scores demonstrated lower j factor scores in the AVN condition, suggesting these participants may have been able to use visual cues to provide context, facilitating accurate target word recall. Perhaps participants with greater lip-reading proficiency also had a richer representation of the visual cues of each of the phonemes in the visuospatial sketchpad (Baddeley, 2000; Baddeley & Hitch, 1974). This representation in the visuospatial sketchpad would also have been communicated to the episodic buffer, as information is shared between
subcomponents (Baddeley, 2000). The episodic buffer is also responsible for holding information in a multi-dimensional code from both the visuospatial sketchpad and auditory information from the phonological loop (Baddeley, 2000). In conditions where the auditory representation of the stimuli may be degraded (i.e., AON condition) the episodic buffer may have been able to rely more on the visual contextual cues from the visuospatial sketchpad to facilitate correct recall for high proficiency lip-readers.

**Error Type.** Each recall error was coded into an error type category. Appendix B provides a detailed description of how error types were categorized. The reliability of error type coding was high across the experimenter and trained undergraduate assistants. Across groups, the part word error type and phoneme substitution errors stand out when comparing the auditory-visual and auditory-only conditions and when comparing quiet and noise conditions. Participants in all three groups had more part word errors and phoneme substitutions in conditions without visual cues compared to those with visual cues. They also had more part word errors and phoneme substitutions in conditions with noise compared to quiet.

An incorrectly recalled word was categorized as a part word error if it shared at least one syllable with the target, or shared a majority of the phonemes in one of the two syllables. The phonemes in the incorrectly recalled word had to be in the same order as the target word, the response had to be a real word, and the response could not be semantically related to the target word. Based on these criteria, part word errors and phoneme substitutions suggest that some representation of the target word was present in short-term working memory. In the presence of low level background noise, or without visual cues, the increased cognitive load may have made
fewer resources available for short-term working processes (encoding, rehearsal, maintenance, manipulation) in one or more of the components of the Baddeley and Hitch model.

Phonotactic probability may provide a possible explanation for significantly higher number of phoneme substitution errors in the group of children compared to the young and older adults and the higher number of phoneme substitution errors in noise compared to quiet. Phonotactic probability refers to the likelihood of a sound sequence in a word (Vitevitch & Luce, 1998, 1999). Sound sequences with higher phonotactic probability are recognized faster and more accurately, and such sequences have been shown to facilitate spoken word recognition compared to less probable sound sequences (Vitevitch, 2003). If the child was unsure of the target word, perhaps he or she utilized phonotactic knowledge and substituted a sound sequence that was more familiar to him or her, resulting in more phoneme substitution errors. In the presence of noise, the increased cognitive load could have impacted the resources available for processing functions in one or more of the components of the Baddeley and Hitch model, and participants utilized phonotactic knowledge to represent the target word.

**Study Limitations**

It is important to note that the participants in the present study were self-selected; they volunteered to complete the experiment. This means the sample that completed the study might have been more random than if participants had not been volunteers. Since the experiment advertised a memory task, those who felt they had good memory skills might have chosen to participate.

Another potential limitation may be that the older adult population contains potential subgroups in terms of executive functioning. A recent investigation suggests the presence of
variation in executive functions (e.g., inhibition, working memory) among older adults, including those who have cognitive functioning that is normal, below normal, or elite (de Frias, Dixon, & Strauss, 2009). de Frias, Dixon & Strauss (2009) suggested that older adults with cognitively elite status have executive functioning resembling that of younger adults, or themselves when they were younger adults. Perhaps the older adults in the present study represent a portion of the older adult population with cognitively elite executive functioning, reflected in their comparable recall performance to young adults.

**Future Directions**

The current experiment utilized one male and one female talker. In the future, several different talkers could be used to present the stimuli. Different talkers would allow for the examination of the potential impact of acoustic cues such as voice-onset time and formant transitions on word recall. This may be especially valuable in examining potential age group differences, as children have been shown to be less sensitive to these types of cues compared to adults (Elliott, 1986; Elliott & Hammer, 1988; Elliott, Longinotti, Meyer, Raz, & Zucker, 1981; Sussman & Carney, 1989).

The type of running memory task used in the present experiment could be utilized to examine the utility of visual cues and short-term working memory for spoken speech perception in different types of noise. Multi-talker babble represents a situation in which the listener must separate the target talker from distractors. The addition of visual cues might facilitate short-term working memory in multi-talker babble situations where the target talker and distractors are similar. Previous evidence has shown greater dissimilarity facilitates the ability to separate the target from distractors (Moore & Gockel, 2002). Recent evidence from Ezzatain, Li, Pichora-
Fuller, and Schneider, (2012) suggests that providing listeners with cues, such as spatial separation or vocal fine structure cues facilitates the separation of the target from distractors. It is also possible that adding visual cues might also provide a way to provide salient cues about the target timing (Golumbic, Cogan, Schroeder, & Poeppel, 2013), in order to facilitate separating the target from distractors in multi-talker babble. This improved separation might enhance short-term working memory processes.

Multiple signal-to-noise ratios might also be used, as there is evidence to suggest that the addition of visual cues may be more valuable to enhancing auditory speech perception at higher signal-to-noise ratios (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Although the present study utilized a +5 dB signal-to-noise ratio, Ross et al. (2007) suggests the addition of visual cues may provide greater benefit the more challenging the SNRs. Exploring recall accuracy across a range of SNR’s would allow for the examination of how increasing SNR might impact cognitive function. Visual cues might support short-term working memory at higher SNRs by providing visual information for the visuospatial sketchpad when auditory stimuli in the phonological loop may be degraded.

Older adults with hearing loss could also take part in a running memory experiment in order to investigate working memory and spoken speech perception in those with decreased hearing sensitivity. Recent studies have focused on identifying ways to evaluate the speech perception abilities and cognitive function of older adults who have hearing loss, with the larger goal of improving clinical rehabilitation and amplification strategies (Besser, Zekveld, Kramer, Rönnberg, & Festen, 2012; Kramer, Zekveld, & Houtgast, 2009; Pichora-Fuller, 2008). Additional data from older adults with hearing loss on the type of task utilized in the present
study might allow for the development of new evaluation or intervention strategies that could be used to improve spoken speech perception.

In future experiments, children could be tested utilizing a running memory task, beginning around age four and ranging up through late adolescence. The wide age range would be useful because previous evidence suggest short-term working memory functions in children younger than age five are different compared to children at age seven (Gathercole, Pickering, Ambridge, & Wearing, 2004). Recent data suggest short-term working memory continues to develop through late adolescence, as well as auditory-visual speech perception abilities (Ross et al., 2011). Previous studies have indicated large individual variation is present in short-term working memory abilities in children (Baddeley, 2007). Additional evidence for the relation between short-term working memory and spoken speech perception in typically developing, normal hearing children across a range of chronological ages may be valuable in understanding both the potential development of cognitive functions, auditory-visual speech perception abilities, and possible sources of individual variation. These findings could then be useful in understanding potential deficits in short-term working memory and/or auditory-visual speech perception abilities in children with hearing loss. Short-term working memory functions have been indicated as an important component in positive outcomes for children with cochlear implants (Harris, Kronenberger, Gao, Hoen, Miyamoto, & Pisoni, 2013).

Overall, the current study provides evidence that factors such as age, presentation modality, and acoustic condition have an impact on short-term working memory. Findings suggest visual cues facilitate short-term working memory, especially in noise. Visual cues from the talker’s face may provide a richer representation of the stimuli in one or more of the
components of the Baddeley and Hitch multi-component model, (Baddeley, 2000; Baddeley and Hitch, 1974) facilitating short-term working memory function. Future studies may continue to investigate the utility of visual cues in facilitating cognitive functions for listeners with normal hearing and hearing loss.
REFERENCES


*Perception and Psychophysics, 5*, 363-373.


Service, E. (2000). Phonological complexity and word duration in immediate recall: 
   Different paradigms answer different questions. A comment on Cowan, Nugent, 

Shallice, T. (2004). The fractionation of supervisory control. In M. Gazzaniga (Ed) The 


   Trends in Amplification, 12(4), 283-299.


   Journal of Experimental Psychology Section A: Human Experimental Psychology, 
   42(2), 291-304.

   correlates. Journal of Cognitive Neuroscience. 23(9), 2248-2259.

   from audibility in older listeners with hearing loss: Effects of age, amplification, and 


APPENDIX A

Participant Questionnaires

Participant Number: ________________

1.) Have you ever had a heart attack, stroke or neurological problem?
   Circle one: Yes   No

2.) Do you have any problems with your vision? (Glasses or contacts are OK)
   Circle one: Yes   No

3.) Have you had chronic otitis media, exposure to loud sounds, or use a hearing aid(s)?
   Circle one: Yes   No

4.) Do you have a history of any speech, language or reading problems?
   Circle one: Yes   No

5.) Do you have at least 8 years of education (have completed up through 8th grade)?
   Circle one: Yes   No
6.) Does your child have any problems with his/her vision?
Circle one: Yes No

7.) Has your child had frequent ear infections, exposure to loud sounds, or use a hearing aid(s)?
Circle one: Yes No

8.) Does your child have any speech or language problems?
Circle one: Yes No

9.) Does your child have problems paying attention or staying on task?
Circle one: Yes No

10.) Do you have any concerns about your child’s speech, reading, vocabulary development or academic performance?
Circle one: Yes No

11.) What is your child’s age?
Age:

12.) What grade in school?
Grade in school
APPENDIX B

Error Code Definitions

Incorrectly recalled words were coded based on the following categories. Errors that were determined to be “earlier in the string”, “earlier in experiment”, “repeated another target” or “no response” did not receive further coding. All other errors were phonetically transcribed and were compared to the target word to determine if they shared common phonemes.

Earlier in String: The word recalled appeared earlier in the string of words.
Example: The four target words are “finance, mainly, result, portion” and the participant recalled “finance, mainly, result, trouble”. The word trouble is the fifth word back in the string and would be marked as an earlier in string error. The lag position of fifth word back would also be noted.

No Response: The participant does not give a verbal response, or says they are unable to recall the target word.

Earlier in Experiment: The word recalled incorrectly appeared earlier in the experiment. The incorrectly recalled word is a whole word.

Repeated Another Target: One of the 4 target words was incorrectly recalled twice.
Example: The four target words are “finance, portion, mainly, result” and the participant recalls “finance, portion, mainly, finance”

Semantically Related: The incorrectly recalled word is related in meaning to the target word.
Example: “house” for “palace”

Unrelated Word (phonetically and semantically unrelated): The word recalled is not one of the target words, or one of the words heard earlier in the string. It does not share phonemes or syllables in the same order the target word, it is not semantically related to the target and it is a real word.
Example: “feeling” for “wisdom”.

**Non-Sense word:** The word recalled is not a real word. It is also not a part word. The part word error category takes precedence over the nonsense word category.

Example: “repold” for “response” would be a nonsense word, while “a-mand” for “command” would be a part word error.

**Single phoneme:** The incorrectly recalled word is a single phoneme from the target word. It is in the same position and in the same syllable as the target word.

“/s/ sound” for “season”

**Part Word (phonetically related, but semantically unrelated):** The incorrectly recalled word shares at least one syllable with the target, or shares a majority of the phonemes in one of the two syllables. The phonemes in the incorrectly recalled word must be in the same order as the target word. The incorrectly recalled word must be a real word.

Example: “builder” for “building” or “working” for “feeling”, “liar” for “higher”, Example for majority: “hardware” for “harbor”, “future” for “buzzer”, “uh” for “agree”.

**Substitution:** In the word recalled, one or more phonemes have been substituted from the target word.

Example: “butter” for “building”

**Addition to the full target word:** In the word recalled, one or more phonemes have been added to the target word.

Example: “loner” for “owner”

**Deletion from the full or part of the target word:** In the word recalled, one or more phonemes have been deleted from the target word.

Example: “sud” for “sudden”.

158